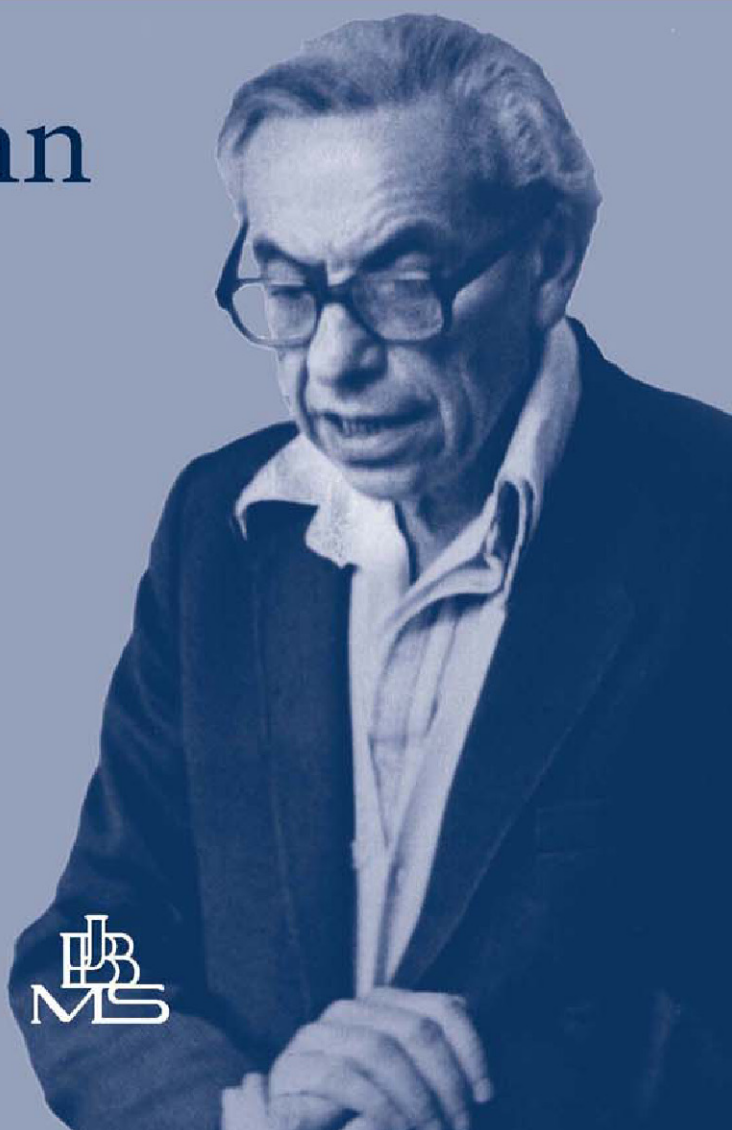


BOLYAI SOCIETY
MATHEMATICAL STUDIES

25

László Lovász · Imre Z. Ruzsa · Vera T. Sós
Editors

Erdős Centenn



 Springer



BOLYAI SOCIETY
MATHEMATICAL STUDIES

25

BOLYAI SOCIETY MATHEMATICAL STUDIES

Editor-in-Chief:
Gábor Fejes Tóth

Series Editor:
Dezső Miklós

Publication Board:

Gyula O. H. Katona · László Lovász · Péter Pál Pálffy
András Recski · András Stipsicz · Domokos Szász

1. **Combinatorics, Paul Erdős is Eighty, Vol. 1**
D. Miklós, V.T. Sós, T. Szőnyi (Eds.)
2. **Combinatorics, Paul Erdős is Eighty, Vol. 2**
D. Miklós, V.T. Sós, T. Szőnyi (Eds.)
3. **Extremal Problems for Finite Sets**
P. Frankl, Z. Füredi, G. O. H. Katona, D. Miklós (Eds.)
4. **Topology with Applications**
A. Császár (Ed.)
5. **Approximation Theory and Function Series**
P. Vértesi, L. Leindler, Sz. Révész, J. Szabados, V. Totik (Eds.)
6. **Intuitive Geometry**
I. Bárány, K. Böröczky (Eds.)
7. **Graph Theory and Combinatorial Biology**
L. Lovász, A. Gyárfás, G. Katona, A. Recski (Eds.)
8. **Low Dimensional Topology**
K. Böröczky, Jr., W. Neumann, A. Stipsicz (Eds.)
9. **Random Walks**
P. Révész, B. Tóth (Eds.)
10. **Contemporary Combinatorics**
B. Bollobás (Ed.)
11. **Paul Erdős and His Mathematics I+II**
G. Halász, L. Lovász, M. Simonovits, V. T. Sós (Eds.)
12. **Higher Dimensional Varieties and Rational Points**
K. Böröczky, Jr., J. Kollár, T. Szamuely (Eds.)
13. **Surgery on Contact 3-Manifolds and Stein Surfaces**
B. Ozbagci, A. I. Stipsicz
14. **A Panorama of Hungarian Mathematics in the Twentieth Century, Vol. 1**
J. Horváth (Ed.)
15. **More Sets, Graphs and Numbers**
E. Györi, G. O. H. Katona, L. Lovász (Eds.)
16. **Entropy, Search, Complexity**
I. Csizsár, G. O. H. Katona, G. Tardos (Eds.)
17. **Horizons of Combinatorics**
E. Györi, G. O. H. Katona, L. Lovász (Eds.)
18. **Handbook of Large-Scale Random Networks**
B. Bollobás, R. Kozma, D. Miklós (Eds.)
19. **Building Bridges**
M. Grötschel, G. O. H. Katona (Eds.)
20. **Fete of Combinatorics and Computer Science**
G. O. H. Katona, A. Schrijver, T. Szőnyi (Eds.)
21. **An Irregular Mind, Szemerédi is 70**
I. Bárány, J. Solymosi (Eds.)
22. **Cylindric-like Algebras and Algebraic Logic**
H. Andréka, M. Ferenczi, I. Németi (Eds.)
23. **Deformations of Surface Singularities**
A. Némethi, Á. Szilárd (Eds.)
24. **Geometry (Intuitive, Discrete and Convex), A Tribute to László Fejes Tóth**
I. Bárány, K. Böröczky, Jr., G. Fejes Tóth, J. Pach (Eds.)

László Lovász
Imre Z. Ruzsa
Vera T. Sós
(Eds.)

Erdős Centennial



Springer



JÁNOS BOLYAI MATHEMATICAL SOCIETY

László Lovász
Eötvös Lóránd University
Department of Computer Science
Pázmány P. sétány 1/c
Budapest 1117
Hungary
e-mail: lovasz@cs.elte.hu

Managing Editor:
Dömötör Pálvölgyi
Eötvös Lóránd University
Department of Computer Science
Pázmány P. sétány 1/c
Budapest 1117
Hungary
e-mail: dom@cs.elte.hu

Imre Z. Ruzsa
Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
Reáltanoda u. 13–15
Budapest 1053
Hungary
e-mail: imre.z.ruzsa@renyi.mta.hu

Vera T. Sós
Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
Reáltanoda u. 13–15
Budapest 1053
Hungary
e-mail: t.sos.vera@renyi.mta.hu

Mathematics Subject Classification (2010):
03EXX, 05-XX, 11BXX, 11KXX, 11MXX, 11PXX, 20BXX, 28-XX, 41-XX, 60BXX

Library of Congress Control Number: 2013941387

ISSN 1217-4696
ISBN 978-3-642-39285-6 Springer Berlin Heidelberg New York
ISBN 978-963-9453-18-0 János Bolyai Mathematical Society, Budapest

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable for prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springer.com

© 2013 János Bolyai Mathematical Society and Springer-Verlag

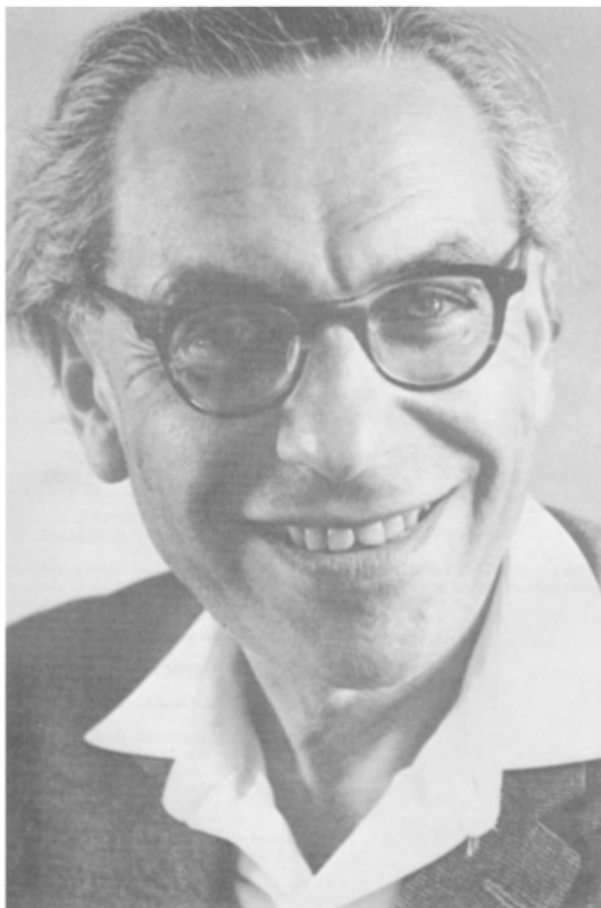
The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover photo of Paul Erdős is courtesy of J. Schönheim
Cover design: WMXDesign GmbH, Heidelberg
Printed on acid-free paper 44/3142/db - 5 4 3 2 1 0

CONTENTS

CONTENTS	5
PREFACE	9
ALON, N.: Paul Erdős and Probabilistic Reasoning	11
BENJAMINI, I.: Euclidean vs. Graph Metric	35
BOLLOBÁS, B. and RIORDAN, O.: The Phase Transition in the Erdős–Rényi Random Graph Process	59
BOURGAIN, J.: Around the Sum-product Phenomenon	111
BREUILLARD, E., GREEN, B. and TAO, T.: Small Doubling in Groups	129
DIAMOND, H. G.: Erdős and Multiplicative Number Theory	153
FÜREDI, Z. AND SIMONOVITS, M.: The History of Degenerate (Bipartite) Extremal Graph Problems	169
GOWERS, W. T.: Erdős and Arithmetic Progressions	265
GRAHAM, R. L.: Paul Erdős and Egyptian Fractions	289
GYÖRY, K.: Perfect Powers in Products with Consecutive Terms from Arithmetic Progressions, II	311
KOMJÁTH, P.: Erdős’s Work on Infinite Graphs	325
KUNEN, K.: The Impact of Paul Erdős on Set Theory	347
MAULDIN, R. D.: Some Problems and Ideas of Erdős in Analysis and Geometry	365
MONTGOMERY, H. L.: L^2 Majorant Principles	377
NEŠETŘIL, J.: A Combinatorial Classic – Sparse Graphs with High Chromatic Number	383
NGUYEN, H. H. and VU, V. H.: Small Ball Probability, Inverse Theorems, and Applications	409
PACH, J.: The Beginnings of Geometric Graph Theory	465
PINTZ, J.: Paul Erdős and the Difference of Primes	485
POLLACK, P. and POMERANCE, C.: Paul Erdős and the Rise of Statistical Thinking in Elementary Number Theory	515

RÖDL, V. and SCHACHT, M.: Extremal Results in Random Graphs	535
SCHINZEL, A.: Erdős's Work on the Sum of Divisors Function and on Euler's Function	585
SHALEV, A.: Some Results and Problems in the Theory of Word Maps	611
TENENBAUM, G.: Some of Erdős' Unconventional Problems in Number Theory, Thirty-four Years Later	651
TOTIK, V.: Erdős on Polynomials	683
VÉRTESI, P.: Paul Erdős and Interpolation: Problems, Results, New Developments	711



Paul Erdős

Paul Erdős 1913–1996

PREFACE

Paul Erdős was one of the most influential mathematicians of the twentieth century. His work in number theory, combinatorics, set theory, and other branches of mathematics has determined the development in large areas of these fields. His name is forever attached to combinatorial and additive number theory, combinatorial geometry, extremal graph and hypergraph theory, random graphs, and the probabilistic method. His contributions to set theory, the theory of primes, analysis, probability, and other classical areas in mathematics are also fundamental.

Paul Erdős passed away in 1996. Three years later, a conference was organized in Budapest to survey his work, his contributions to mathematics, and the far-reaching impact of his work on many branches of mathematics. A 2-volume collection of papers, “Paul Erdős and his Mathematics” (János Bolyai Mathematical Society and Springer-Verlag 2002), was also published, which contained papers about his life, surveys of areas which he initiated or contributed to, and personal reminiscences by his friends and collaborators.

We feel that in 2013, on the 100th anniversary of his birth, it was time to have another look on the long-term impact of his work. We are organizing another conference devoted to his mathematics. This volume (which is not the Proceedings of this conference, but of course having the similar goals) undertakes the almost impossible task to describe the ways in which problems raised by him and topics initiated by him (indeed, whole branches of mathematics) continue to flourish.

Written by outstanding researchers in these areas, the papers in this volume include extensive surveys of classical results as well as of new developments. It would be even more hopeless to be comprehensive than in 1999, but we hope that this volume, as well as the lectures at the conference, will give a glimpse into how his mind was working, and a feeling for his tremendous influence on modern mathematics.

The interested reader should also consult the home page of the conference (<http://www.renyi.hu/erdos100>), which contains more material, including the program and abstracts of posters submitted to the conference. We plan that recordings of plenary talks will also be made available. The Paul Erdős page (http://www.renyi.hu/~p_erdos) contains scanned copies of most Erdős papers, along with many photos and a lot of other material.

Our thanks are due to Dömötör Pálvölgyi for his very careful and efficient work as managing editor of this volume, to Dezső Miklós for organizing the production, and to Ildikó Miklós for the expert production of the \LaTeX files.

Budapest, May 2013

László Lovász
Imre Z. Ruzsa
Vera T. Sós

PAUL ERDŐS AND PROBABILISTIC REASONING

NOGA ALON*

One of the major contributions of Paul Erdős is the development of the Probabilistic Method and its applications in Combinatorics, Graph Theory, Additive Number Theory and Combinatorial Geometry. This short paper describes some of the beautiful applications of the method, focusing on the long-term impact of the work, questions and results of Erdős. This is mostly a survey, but it contains a few novel results as well.

1. THE PROBABILISTIC METHOD

The Probabilistic Method is one of the most significant contributions of Paul Erdős, and part of his greatness is the fact that applications of the probabilistic method and of random graphs have become so common that it is now possible to use those without explicitly mentioning him. The method is a powerful tool with numerous applications in Combinatorics, Graph theory, Additive Number Theory and Geometry and had an immense impact on the development of theoretical Computer Science as well. The results and tools are far too numerous to cover in a short survey, even if the focus is only on those influenced directly by the work and problems of Erdős, and thus this paper is mainly a selection of topics that illustrate the method, and is not meant to be a comprehensive treatment of the whole area. Several books that contain more material on the subject are [13], [18], [54], [60].

It is convenient to classify the applications of probabilistic techniques in Discrete Mathematics into three groups. The first one deals with the study of random combinatorial objects, like random graphs or random matrices. The results here are essentially results in Probability Theory,

*Research supported in part by an ERC Advanced grant, by a USA-Israeli BSF grant and by the Israeli I-Core program.

although many of them are motivated by problems in Combinatorics. The second group consists of probabilistic constructions. These are applications of probabilistic arguments in order to prove the existence of combinatorial structures which satisfy a list of prescribed properties. Existence proofs of this type often supply extremal examples to various questions in Discrete Mathematics. The third group, which contains some of the most striking examples, focuses on the application of probabilistic reasoning in the proofs of deterministic statements whose formulation does not give any indication that randomness may be helpful in their study.

Random graphs are covered in another chapter of this volume. The present chapter contains a brief description of several results in each of the other two groups, as well as a very brief discussion of some of the applications of the probabilistic method in theoretical Computer Science. The influence of the work and questions of Paul Erdős in all these has been crucial.

This is mostly a survey paper, but it contains several new results, presented in subsections 3.2 and 3.5, as well.

2. PROBABILISTIC CONSTRUCTIONS

The applications of probabilistic constructions have yielded numerous results in Combinatorics, Graph Theory, Combinatorial Geometry and Additive Number Theory. Below is a selection of several representative examples.

2.1. Ramsey Numbers

Let H_1, H_2, \dots, H_k be k finite, undirected, simple graphs. The (multicolor) *Ramsey number*

$$r(H_1, H_2, \dots, H_k)$$

is the minimum integer r such that in every edge coloring of the complete graph on r vertices by k colors, there is a monochromatic copy of H_i in color i for some $1 \leq i \leq k$. By a (special case of) a well known theorem of Ramsey (c.f., e.g., [49]), this number is finite for every sequence of graphs H_i .

The determination or estimation of these numbers is usually a very difficult problem. When each graph H_i is a complete graph with more than two vertices, the only values that are known precisely are those of $r(K_3, K_m)$ for $m \leq 9$, $r(K_4, K_4)$, $r(K_4, K_5)$ and $r(K_3, K_3, K_3)$. Even the determination of the asymptotic behavior of Ramsey numbers up to a constant factor is a hard problem, and despite a lot of efforts by various researchers (see, e.g., [49], [22] and their references), there are only a few infinite families of graphs for which this behavior is known.

In one of the first applications of the probabilistic method in Combinatorics, Erdős [26] proved that if $\binom{n}{k} 2^{1-\binom{k}{2}} < 1$ then $R(K_k, K_k) > n$, that is, there exists a 2-coloring of the edges of the complete graph on n vertices containing no monochromatic clique of size k . This implies that $R(K_k, K_k) > 2^{k/2}$ for all $k \geq 3$. The proof is extremely short: the probability that a random two-edge coloring of K_n contains a monochromatic copy of K_k is at most $\binom{n}{k} 2^{1-\binom{k}{2}} < 1$, and hence there is a coloring with the required property.

It is worth noting that although this argument seems almost trivial today, it was far from being obvious when published in 1947. In fact, several prominent researchers believed, before the publication of this short paper, that $R(K_k, K_k)$ may well be bounded by a polynomial in k . In particular, Paul Turán writes in [67] that he had conjectured for a while that $R(K_k, K_k)$ is roughly k^2 , and that Erdős's result showed that this quantity behaves very differently than expected.

A particularly interesting example of an infinite family for which the asymptotic behavior of the Ramsey number is known, is the following result of Kim and of Ajtai, Komlós and Szemerédi.

Theorem 2.1 ([56], [3]). *There are two absolute positive constants c_1, c_2 such that*

$$c_1 m^2 / \log m \leq r(K_3, K_m) \leq c_2 m^2 / \log m$$

for all $m > 1$.

The upper bound, proved in [3], is probabilistic, and applies a certain random greedy algorithm. There are several subsequent proofs, all are based on probabilistic arguments. The lower bound is proved by a “semi-random” construction and proceeds in stages. The detailed analysis is subtle, and is based on certain large deviation inequalities. An alternative analysis of this probabilistic construction, inspired by the differential equation method of Wormald [71], is given by Bohman in [17]. It is worth noting that the question of obtaining a super-linear lower bound for $r(K_3, K_m)$ is mentioned already in [26], and Erdős has established in [28], by an appropriate probabilistic construction, an $\Omega(m^2 / \log^2 m)$ lower bound. More on this appears in another chapter of this volume.

Even less is known about the asymptotic behavior of multicolor Ramsey numbers, that is, Ramsey numbers with at least 3 colors. The asymptotic behavior of $r(K_3, K_3, K_m)$, for example, has been very poorly understood for quite some time, and Erdős and Sós conjectured in 1979 (c.f., e.g., [22]) that

$$\lim_{m \rightarrow \infty} \frac{r(K_3, K_3, K_m)}{r(K_3, K_m)} = \infty.$$

This has been proved in [12], where it is shown that in fact $r(K_3, K_3, K_m)$ is equal, up to logarithmic factors, to m^3 . A more complicated, related result proved in [12], that supplies the asymptotic behavior of infinitely many families of Ramsey numbers up to a constant factor is the following.

Theorem 2.2. *For every $t > 1$ and $s \geq (t - 1)! + 1$ there are two positive constants c_1, c_2 such that for every $m > 1$*

$$c_1 \frac{m^t}{\log^t m} \leq r(K_{t,s}, K_{t,s}, K_{t,s}, K_m) \leq c_2 \frac{m^t}{\log^t m},$$

where $K_{t,s}$ is the complete bipartite graph with t vertices in one color class and s vertices in the other.

The proof of the lower bound forms yet another example of a probabilistic construction, where each of the first three color classes is a randomly shifted copy of an appropriate $K_{t,s}$ -free graph that contains a relatively small number of large independent sets, as shown by combining some spectral techniques with character sum estimates.

2.2. Combinatorial Geometry

There are several striking examples where a probabilistic construction supplies rather easily counter-examples to well studied conjectures in Combinatorial Geometry. The following result of Erdős and Füredi illustrates this point.

Theorem 2.3 ([34]). *For every $d \geq 1$ there is a set of at least $\lfloor \frac{1}{2} \left(\frac{2}{\sqrt{3}} \right)^d \rfloor$ points in the d -dimensional Euclidean space R^d , such that all angles determined by three points from the set are strictly less than $\pi/2$.*

The proof is obtained by considering a random set of binary vectors in R^d . We omit the details but mention that this disproves an old conjecture of Danzer and Grünbaum [23] which suggests that the maximum cardinality of such a set is at most $2d - 1$. The authors of [23] did prove, motivated by a question of Erdős and Klee, that the maximum cardinality of a set of points in R^d in which all angles are at most $\pi/2$ is 2^d .

A *range space* S is a pair (X, R) , where X is a (finite or infinite) set and R is a (finite or infinite) family of subsets of X . The members of X are called *points* and those of R are called *ranges*. If A is a subset of X then $P_R(A) = \{r \cap A : r \in R\}$ is the *projection* of R on A . In case this projection contains all subsets of A we say that A is *shattered*. The *Vapnik-Chervonenkis* dimension (or VC-dimension) of S , denoted by $VC(S)$, is the

maximum cardinality of a shattered subset of X . If there are arbitrarily large shattered subsets then $VC(S) = \infty$.

A subset $N \subset A$ is an ε -net for A if any range $r \in R$ satisfying $|r \cap A| \geq \varepsilon|A|$ contains at least one point of N .

A well known result of Haussler and Welzl [52], following earlier work of Vapnik and Chervonenkis [68], asserts that for any n and $\varepsilon > 0$, any set of size n in a range space of VC-dimension d contains an ε -net of size at most $O\left(\frac{d}{\varepsilon} \log(1/\varepsilon)\right)$.

The authors of [61] asked in 1990 whether or not in all natural geometric scenarios of bounded VC-dimension, there always exists an ε -net of size $O(1/\varepsilon)$. This problem received a considerable amount of attention over the years, until it has finally been answered negatively in [5] and in [62], by constructions that have essential probabilistic ingredients. The following, however, is still open.

Problem 2.4. Are there sets X_n of points in the plane and a sequence $\varepsilon_n > 0$ tending to zero so that the minimum size of an ε_n -net for X_n with respect to line ranges is $\Omega\left(\frac{1}{\varepsilon_n} \log\left(\frac{1}{\varepsilon_n}\right)\right)$?

2.3. Additive Number Theory

Erdős and Turán [41] asked if for any asymptotic basis of order 2 of the positive integers (that is, a set A of positive integers so that each sufficiently large integer has a representation as a sum of two elements of A), there must be, for any constant t , integers that have more than t such representations.

Erdős has used in [27] a probabilistic construction to prove the existence of a set A of integers such that every n is represented as $n = x + y$ with $x, y \in A$ at least once but at most $O(\ln n)$ times. This settles a problem posed by Sidon and shows that in the Erdős-Turán question mentioned above one cannot expect to necessarily have too many representations of an integer n , although the question, as posed, is still wide open.

A somewhat similar question is considered by Canfield and Wilf in [21] and by Ljujić and Nathanson in [59]. For two sets A and M of positive integers and for a positive integer n , let $p(n, A, M)$ denote the number of partitions of n with parts in A and multiplicities in M , that is, the number of representations of n in the form $n = \sum_{a \in A} m_a a$ where $m_a \in M \cup \{0\}$ for all a , and all numbers m_a but finitely many are 0. There are simple examples of M and A in which M is finite so that $p(n, A, M) = 1$ for all n , but it seems more difficult to find infinite sets A and M for which $p(n, A, M)$ has a polynomial growth in n . For the specific cases of $A = \{k!\}_{k=1}^{\infty}$, $A = \{k^k\}_{k=1}^{\infty}$ (and many other cases), the existence of such an infinite M is proved in

[6] using a probabilistic construction and answering questions raised in [21] and [59]. These constructions are tailored to fit the growth of the given sequence A , and are general enough to ensure that the same sequence M can work simultaneously for several sequences A . The analysis is based on some large deviation inequalities.

Erdős and Newman studied in [39] another problem dealing with bases for sets of integers. They studied bases for m -element subsets A of $\{1, \dots, n\}$, where a set B is a basis for A if $A \subset B + B = \{b_1 + b_2 : b_1, b_2 \in B\}$. Since $\{0\} \cup A$ is a basis for A , and there is a set X with at most $c\sqrt{n}$ elements such that $X + X \supset \{1, \dots, n\}$ it follows that for any m -element subset of $\{1, \dots, n\}$ there is always a basis of size $\min(c\sqrt{n}, m + 1)$. Erdős and Newman showed by a simple probabilistic construction that if m is somewhat smaller than \sqrt{n} , say $m = O(n^{1/2-\varepsilon})$, then almost no m -element set has a basis of size $o(m)$. Similarly, if m is at least $n^{1/2+\varepsilon}$ then almost all m -element sets require a basis of size at least $c\sqrt{n}$. For the borderline case when m is of the order \sqrt{n} their counting argument only yields existence of sets that need a basis of size $c\sqrt{n} \log \log n / \log n$, and they asked if every m -set of size $m = \sqrt{n}$ has a basis with $o(m)$ elements. This is established in [7], where it is shown that in fact any such set has a basis of size $O(\sqrt{n} \log \log n / \log n)$. The argument is probabilistic.

Estimating the size of the smallest possible basis for explicitly given sets is often far harder. Erdős and Newman showed that any basis for the set of squares $\{t^2 : t = 1, \dots, n\}$ (which is a subset of $\{1, 2, \dots, n^2\}$) is of size at least $n^{2/3-o(1)}$ for large values of n , which is an improvement over the trivial lower bound of $n^{1/2}$. They constructed a small basis for the squares, of size only $O\left(\frac{n}{\log^M n}\right)$ for any M . Wooley asked about powers other than the squares. Whereas it is likely that any basis for the set of d -th powers $\{t^d : t = 1, \dots, n\}$ is of size $\Omega(n^{1-\varepsilon})$ for every $\varepsilon > 0$ and $d \geq 2$, only a modest improvement of the $n^{2/3-o(1)}$ lower bound of Erdős and Newman for large values of d is proved in [7], where it is shown that the set $\{t^d : t = 1, \dots, n\}$ does not have a basis of size $O\left(n^{3/4 - \frac{1}{2\sqrt{d}} - \frac{1}{2(d-1)} - \varepsilon}\right)$ for any $\varepsilon > 0$.

3. DETERMINISTIC THEOREMS

3.1. Sum-free subsets

A subset A of an abelian group is called sum-free if there is no solution to the equation $x + y = z$ with $x, y, z \in A$. Erdős [31] showed that any set of n positive integers contains a sum-free subset of size at least $n/3$. The proof is a simple yet intriguing application of the probabilistic method, and proceeds as follows. Let A be a set of n positive integers, choose a real x uniformly between 0 and 1 and let $B = B_x$ be the set of all $a \in A$ so that $ax \bmod 1 \in (1/3, 2/3)$. It is not difficult to check that B is always sum-free, and that the expected value of the size $|B_x|$ of B is $n/3$. Therefore, there is a fixed x so that the size of B_x is at least $n/3$, providing the required result.

In [8] the authors showed that a similar proof gives a lower bound of $(n+1)/3$. Bourgain [20] has further improved this estimate to $(n+2)/3$. It seems possible that the constant $1/3$ cannot be replaced by a larger constant, but this is an open problem. The best known upper bound is $11/28$, proved by Lewko [58], improving earlier estimates of $3/7$ in [31] and $12/29$ in [8]. In subsection 3.2 we present a further (modest) improvement. It is worth noting that for general abelian groups there is a similar result proved in [8]: any set of n nonzero elements in any abelian group contains a sum-free subset with more than $2n/7$ elements. The constant $2/7$ is best possible.

3.2. The sum-free subset constant

For a set B of nonzero integers, let $s(B)$ denote the maximum cardinality of a sum-free subset of B . The infimum value of the ratio $\frac{s(B)}{|B|}$ as B ranges over all nonempty sets of nonzero integers is called the sum-free subset constant, and is denoted by δ . As mentioned in the previous subsection Erdős proved that $\delta \geq 1/3$ and observed that $\delta \leq 3/7$. The upper bound has been improved in [8] and further improved in [58]. All these upper bounds are established by exhibiting a set B and by computing $s(B)$. The next statement shows that for any given example B it is possible to construct another one which gives a (slightly) better upper bound for δ .

Proposition 3.1. *Let B be a finite set of b nonzero integers and define $s = s(B)$. Put*

$$p = [b(b-1) + 1](b-s+1), \quad q = [p!(e - e^{-1} + 3)/2] - p + 2$$

and

$$m = \left\lceil \frac{q}{b(b-1) + 1} \right\rceil b.$$

Then there is a set C of at most m elements so that

$$\frac{s(C)}{|C|} \leq \frac{s(B)}{|B|} - \frac{1}{|C|}.$$

The result of [58] is proved by exhibiting an explicit set B of 28 nonzero integers for which $s(B) = 11$. Therefore $\delta \leq 11/28$. By the proposition above this can be improved to $11/28 - \varepsilon$ for some ε which is roughly $10^{-50,000}$. It is possible to get a slightly bigger value of ε , but as this is certainly far from giving a tight bound, we make no serious attempt to optimize this value here. Note that the proposition above implies that δ is an infimum, and not a minimum, that is, there is no finite set B so that $\delta = \frac{s(B)}{|B|}$.

Proof. Put $|B| = b$, $s = s(B)$. Let n be a large integer, to be chosen later, and let G be the graph whose set of vertices is $\{1, 2, \dots, n\}$, where i and j are adjacent iff the two sets iB and jB intersect (and $i \neq j$). It is clear that the maximum degree of this graph is at most $b(b-1)$ and hence, by the Hajnal-Szemerédi Theorem [51], it has a proper coloring f with $k = b(b-1) + 1$ colors and nearly equal color classes. This coloring provides a partition of $[n] = \{1, 2, \dots, n\}$ into k sets I_j , so that each of the set $B_j = \cup_{i \in I_j} iB$ is a set of exactly $|I_j|b$ nonzero integers.

Claim: If n is sufficiently large then at least one of these sets B_j does not contain a sum-free subset containing s elements from each of the sets iB for all $i \in I_j$.

Indeed, assuming this is not the case, fix a sum-free subset A_j in each B_j so that $|A_j \cap iB| = s$ for all $i \in I_j$. Using the sets A_j , define a coloring g of I_j by $b-s+1$ colors as follows. Let $x_1 < x_2 < \dots < x_b$ be the members of B and suppose $i \in I_j$. By assumption A_j contains at least one of the elements ix_q for some $q \in \{1, 2, \dots, b-s+1\}$. Let q be the smallest index for which this holds and define $g(i) = q$. The ordered pair $(f(i), g(i))$ defines a coloring of the integers in $[n]$ by $k(b-s+1) = [b(b-1) + 1](b-s+1)$ colors.

Note that there is no monochromatic Schur triple in this coloring, that is, there are no $i, j, t \in [n]$ so that $i + j = t$ and $(f(i), g(i)) = (f(j), g(j)) = (f(t), g(t))$. This is because if there is such a triple then for $(f', g') = (f(i), g(i))$ we have $iB \cup jB \cup tB \subset B_{f'}$, and for $x_{g'} \in B$ $ix_{g'}, jx_{g'}, tx_{g'}$ all lie in $A_{f'}$. This contradicts the fact that $A_{f'}$ is sum-free, as $ix_{g'} + jx_{g'} = tx_{g'}$. Thus there are indeed no monochromatic Schur triples.

An old Theorem of Schur (c.f., e.g., [49]) asserts that if n is sufficiently large as a function of the number of colors used then there must be a

monochromatic Schur triple, contradiction. This contradiction proves the assertion of the claim.

Returning to the proof of the proposition, note that the number of colors in the construction above is $p = \lceil b(b-1) + 1 \rceil (b-s+1)$. By [70] if n is at least $q = \lceil p!(e - e^{-1} + 3)/2 \rceil - p + 2$ then there is a monochromatic Schur triple. This implies that if indeed n is at least that large, then at least one of the sets B_j cannot contain a sum-free subset that consists of s elements from each iB for $i \in I_j$. Hence $s(B_j) \leq |I_j|s - 1$ and as the size of each set I_j is at most $\lceil \frac{q}{b(b-1)+1} \rceil$ the set $C = B_j$ completes the proof of the proposition. ■

3.3. List coloring and Euclidean Ramsey Theory

The *list chromatic number* (or *choice number*) $\chi_\ell(G)$ of a graph $G = (V, E)$ is the minimum integer s such that for every assignment of a list L_v of s colors to each vertex v of G , there is a proper vertex coloring of G in which the color of each vertex is in its list. This notion was introduced independently by Vizing in [69] and by Erdős, Rubin and Taylor in [40]. In both papers the authors realized that this is a variant of usual coloring that exhibits several new interesting properties, and that in general $\chi_\ell(G)$, which is always at least as large as the chromatic number of G , may be arbitrarily large even for graphs G of chromatic number 2.

It is natural to extend the notion of list coloring to hypergraphs. The list chromatic number $\chi_\ell(H)$ of a hypergraph H is the minimum integer s such that for every assignment of a list of s colors to each vertex of H , there is a vertex coloring of H assigning to each vertex a color from its list, with no monochromatic edges.

An intriguing property of list coloring of graphs, which is not shared by ordinary vertex coloring, is the fact that the list chromatic number of any (simple) graph with a large average degree is large. Indeed, it is shown in [4] that the list chromatic number of any graph with average degree d is at least $(\frac{1}{2} - o(1)) \log_2 d$, where the $o(1)$ -term tends to zero as d tends to infinity. For $r \geq 3$, simple examples show that there is no nontrivial lower bound on the list chromatic number of an r -graph in terms of its average degree. However, such a result does hold for simple hypergraphs. Recall that a hypergraph is *simple* if every two of its distinct edges share at most one vertex. The following result is proved in [10].

Theorem 3.2. *For every fixed $r \geq 2$ and $s \geq 2$, there is a $d = d(r, s)$, such that the list chromatic number of any simple r -graph with n vertices and nd edges is greater than s .*

A similar result for the special case of d -regular 3-uniform simple hypergraphs has been obtained independently in [53]. A subsequent proof with a better upper estimate for $d(r, s)$ appears in a recent paper of Saxton and Thomason [66].

The proof of the theorem is probabilistic and proceeds by induction on r . For simplicity we only outline the idea for the case of graphs with a large minimum degree. Let $G = (V, E)$ be a graph with n vertices and minimum degree d . Choose a random set B of about n/\sqrt{d} vertices and assign a random list of size s out of a set S of $2s - 1$ colors to each vertex of B . A simple computation shows that if, say, $d > 10^s$, then with positive (and in fact high) probability many of the vertices v not in B have every subset of size s of S assigned to at least one of their B -neighbors. Fix such a choice of the set B and lists of colors to its vertices. Note now that for each fixed choice of a coloring f of the vertices of B from their lists, at least s distinct colors appear on the B -neighbors of any vertex v of the type mentioned above. If we now assign a random list to such a vertex v , then with probability at least $\binom{2s-1}{s}^{-1} > 4^{-s}$ it will be a forbidden list, that is, it will consist only of colors assigned by f to its neighbors, showing that the coloring f of the B vertices cannot be extended to a proper list coloring of the whole graph. There are only $s^{|B|}$ possible colorings of the vertices of B from their lists, and the probability that no vertex v gets a forbidden list is small enough to ensure that this will not happen for any of these colorings. This argument suffices to show that the list chromatic number of G exceeds s . The hypergraph case is more complicated, and we do not include it here.

The argument above suggests an interesting algorithmic question: given a graph $G = (V, E)$ with minimum degree $d > 10^s$, can we find, deterministically and efficiently, lists of size s for each $v \in V$ so that there is no proper coloring of G assigning to each vertex a color from its list? This problem is open, as is the simpler NP version of it, that is, that of finding sets S_v and providing a certificate that there is no proper coloring using the lists. Here the sets do not have to be found efficiently, and we only require that one will be able to check the certificate efficiently.

The last theorem has an interesting application in Euclidean Ramsey Theory – yet another subject initiated by Erdős and his collaborators. A well known problem of Hadwiger and Nelson is that of determining the minimum number of colors required to color the points of the Euclidean plane so that no two points at distance 1 have the same color. Hadwiger showed already in 1945 that 7 colors suffice, and Moser and Moser noted in 1961 that 3 colors do not suffice. These bounds have not been improved,

despite a considerable amount of effort by various researchers, see [55, pp. 150–152] and the references therein for more on the history of the problem.

A more general problem is considered in [35], [36], [37], where the main question is the investigation of finite point sets K in the Euclidean space for which any coloring of an Euclidean space of dimension d by r colors must contain a monochromatic copy of K . There are lots of intriguing conjectures that appear in these papers. One of them asserts that for any set K of 3 points which do not form an equilateral triangle the minimum number of colors required for coloring the plane with no monochromatic isometric copy of K is 3. The situation is very different for list coloring. A simple Corollary of the theorem above is the following.

Theorem 3.3 ([10]). *For any finite set X in the Euclidean plane and for any positive integer s , there is an assignment of a list of size s to every point of the plane, such that whenever we color the points of the plane from their lists, there is a monochromatic isometric copy of X .*

3.4. Turán numbers and Dependent random choice

For a graph H and an integer n , the Turán number $ex(n, H)$ is the maximum possible number of edges in a simple graph on n vertices that contains no copy of H . The asymptotic behavior of these numbers for graphs H of chromatic number at least 3 is well known, and is determined by the Erdős-Stone-Simonovits Theorem. For bipartite graphs H , however, the situation is considerably more complicated, and there are relatively few nontrivial such graphs H for which the order of magnitude of $ex(n, H)$ is known. A rather general result with a relatively simple proof, described in [11], asserts that for every fixed bipartite graph H in which the degrees of all vertices in one color class are at most r , there is a constant $c = c(H)$ so that $ex(n, H) \leq cn^{2-1/r}$. This is tight for all values of r , as it is known that for every r and $t > (r - 1)!$, there is a simple graph with n vertices and at least $c_{r,t}n^{2-1/r}$ edges, containing no copy of the complete bipartite graph $K_{r,t}$.

The basic tool in the proof is a simple and yet surprisingly powerful method, whose probabilistic proof may be called “dependent random choice”, as it involves a random selection of a set of vertices, where the choices are dependent in a way that increases the probability that r -tuples of the selected vertices will have many common neighbors. An early version of this lemma has first been proven in [50] and [57], and many variants and extension have been obtained afterwards. See [44] for a survey containing lots of applications in Extremal Graph Theory and in Additive Number Theory.

One of the basic versions of the lemma is the following.

Lemma 3.4 ([11]). *Let a, b, n, r be positive integers. Let $G = (V, E)$ be a graph on $|V| = n$ vertices with average degree $d = 2|E|/n$. If*

$$(1) \quad \frac{d^r}{n^{r-1}} - \binom{n}{r} \left(\frac{b-1}{n} \right)^r > a - 1,$$

then G contains a subset A_0 of at least a vertices so that every r vertices of A_0 have at least b common neighbors.

The proof proceeds by considering a (multi)-set T of r random vertices of G , chosen uniformly with repetitions. Let A be the set of all vertices of G which are neighbors of all members of T . The crucial fact is that the expected value of $|A|$ is large, by linearity of expectation and convexity, whereas the expected number of r -tuples of vertices of A with a small number of common neighbors is small, as it is not likely that all vertices of T fall into such a small set of common neighbors. The set A_0 can thus be obtained from A by deleting a vertex from each such undesirable r -tuple.

The lemma above easily implies the following result, that can also be derived from an earlier result of Füredi [47] proved by a different method, in response to a question of Erdős.

Theorem 3.5. *Let H be a bipartite graph with maximum degree r on one side. Then there exists a constant $c = c(H) > 0$ such that*

$$ex(n, H) < cn^{2-\frac{1}{r}}.$$

The method yields several related results, but does not suffice to settle the following problem, suggested by Erdős.

Problem 3.6 ([33]). *A graph is r -degenerate if every subgraph of it contains a vertex of degree at most r . Is it true that for every fixed r -degenerate bipartite graph H , $ex(n, H) \leq O(n^{2-1/r})$?*

As shown in [11], the method of dependent random choice with some twists does imply that for each such H on h vertices, $ex(n, H) \leq h^{1/2r} n^{2-\frac{1}{4r}}$.

3.5. Hypergraph coloring

Erdős realized already in the 60s that probabilistic methods are powerful in the study of hypergraph coloring problems. Several examples appear in [29], [30], [38]. A k -uniform hypergraph is two-colorable if it has a vertex coloring by two colors so that no edge is monochromatic. In [29], [30] Erdős applies probabilistic arguments to prove that the minimum possible number of edges in a k -hypergraph that is not two-colorable is at least 2^{k-1} and at most $O(k^2 2^k)$. The lower bound has been improved several times, and all the improved proofs apply the probabilistic method. The current record is $\Omega\left(\sqrt{\frac{k}{\log k}} 2^k\right)$, due to Radhakrishnan and Srinivasan [64]. See also [63] for a weaker $\Omega(k^{1/4} 2^k)$ bound, with a beautiful short (probabilistic) proof.

One of the main motivations for proving the Lovász Local Lemma in [38] has also been the study of the minimum possible number of edges of a *simple* k -uniform hypergraph which is not two-colorable.

A recent result of Blais, Weinsein and Yoshida [16] deals with a new intriguing variant of hypergraph coloring. In the rest of this section we describe this notion and present some new results about it.

A hypergraph \mathcal{F} is t -intersecting if the intersection of any two of its edges is of size at least t . A vertex coloring of \mathcal{F} is c -strong if any edge F contains vertices of at least $\min\{|F|, c\}$ colors. Let $\chi(t, c)$ denote the minimum f so that any t -intersecting hypergraph admits a c -strong coloring with at most f colors, (∞ if there is no such f).

This notion is defined in [16] where the authors observe that $\chi(t, c)$ is infinite for all $t \leq c - 2$, $\chi(c - 1, c) \geq 2c - 1$ and that $\chi(t, c) \geq 2c - 2$ for all $t \geq c \geq 2$, and prove that $\chi(c, c) < \sqrt{c} e^c$ and that for all $t \geq 2c$, $\chi(t, c) \leq 2c^2$.

They raise several questions regarding the determination of this function, and in particular note that their method does not provide any sub-quadratic (in c) bound for $\chi(t, c)$ for any t , and ask whether or not for each fixed c the limit of $\chi(t, c)$ as t tends to infinity is $2c - 2$.

The following theorem nearly settles this question.

Theorem 3.7. *For every fixed $c \geq 2$ there exists a $t_0 = t_0(c)$ ($\leq O(c^2)$) so that for all $t > t_0$, $\chi(t, c) \leq 2c - 1$.*

The proof follows the basic approach of [16], showing that a random coloring with $2c - 1$ colors provides a c -strong coloring with positive probability bounded away from zero. We note that the example of all subsets of cardinality at least $(n + t)/2$ of an n -element set, where $n \gg t^2$, shows that for a random coloring $2c - 2$ colors do not suffice, as with high probability the largest $c - 1$ color classes will contain more than $(n + t)/2$ elements. A

more careful analysis sketched at the end of this section shows that for random colorings with $2c - 1$ colors, the $O(c^2)$ estimate for the intersection t is optimal as well.

We need a result about the biased measure of t -intersecting hypergraphs. A sharp version of this result was first proved in [2], and can be deduced from the main result of [1]. See also [14], [24], [46] for subsequent related statements. Here we give a much simpler, self-contained proof of a somewhat weaker estimate that suffices for our purpose.

For a hypergraph \mathcal{F} and a real p , $0 \leq p \leq 1/2$, let $\mu_p(\mathcal{F})$ denote the p -measure of \mathcal{F} , that is, the probability that a random set of vertices of \mathcal{F} obtained by selecting each vertex, randomly and independently, with probability p , forms an edge in \mathcal{F} . Thus $\mu_p(\mathcal{F}) = \sum_{F \in \mathcal{F}} \mu_p(F)$, where $\mu_p(F) = p^{|F|}(1-p)^{n-|F|}$, and n is the number of vertices of \mathcal{F} . It is convenient to formulate the results in terms of escape probabilities of random walks. A p -biased random walk of length n is a sequence of independent, identically distributed random variables X_1, X_2, \dots, X_n where each X_i is $+1$ with probability p and -1 with probability $1-p$. Put $S_i = \sum_{j=1}^i X_j$, let $W(p, t, i)$ be the probability that $S_i \geq t$ and let $W(p, t)$ denote the probability that there exists some i so that $S_i \geq t$.

Associate each subset F of $[n] = \{1, 2, \dots, n\}$ with an assignment of values to the variables X_1, X_2, \dots, X_n by defining $X_i = 1$ if $i \in F$ and $X_i = -1$ otherwise. With this assignment, $\mu_p(F)$ is exactly the probability of the corresponding walk.

Let W_i denote the set of all walks for which $S_i \geq t$, and let F_i denote the corresponding family of subsets. It is easy to see that this family is t -intersecting. Indeed, if two sets in the family correspond to the walks (X_1, X_2, \dots, X_n) and (Y_1, Y_2, \dots, Y_n) , then $\sum_{j=1}^i (X_j + Y_j) \geq 2t$ and as each term $X_j + Y_j$ lies in $\{-2, 0, 2\}$, at least t of the terms are 2, providing the required intersection. Therefore, for every $i \leq n$ there is a t -intersecting family of subsets of $[n]$ of p -measure at least $W(p, t, i)$. It turns out that the maximum possible p -measure of such a family is exactly $\max_{i \leq n} W(p, t, i)$.

Lemma 3.8 ([2]). *For any t -intersecting hypergraph \mathcal{F} on n vertices and any $p < 1/2$, $\mu_p(\mathcal{F}) \leq \max_{i \leq n} W(p, t, i)$.*

Here we give a simple proof of the following weaker estimate

Lemma 3.9. *For any (finite) t -intersecting hypergraph \mathcal{F} and any $p < 1/2$, $\mu_p(\mathcal{F}) \leq W(p, t)$.*

Proof: We apply shifting, which is a common technique in the area, see, e.g., [45]. Let $[n]$ be the set of vertices of \mathcal{F} . For each $1 \leq i <$

$j \leq n$ define an operator S_{ij} on the edges of \mathcal{F} , where for each $F \in \mathcal{F}$, $S_{ij}(F) = F - \{j\} \cup \{i\}$ if $j \in F, i \notin F$ and $F - \{j\} \cup \{i\} \notin \mathcal{F}$, and $S_{ij}(F) = F$ otherwise. Put $S_{ij}(\mathcal{F}) = \{S_{ij}(F) : F \in \mathcal{F}\}$. It is easy and well known that if \mathcal{F} is t -intersecting so is $S_{ij}(\mathcal{F})$. It is also clear that $S_{ij}(\mathcal{F})$ has exactly the same p -measure as \mathcal{F} . Moreover, if $S_{ij}(\mathcal{F})$ differs from \mathcal{F} , then the sum of elements in all edges of $S_{ij}(\mathcal{F})$ is smaller than that of the elements in all edges of \mathcal{F} . We can thus keep applying the shift operators S_{ij} to our hypergraph until the process stabilizes, providing a left-shifted family of subsets, which, with a slight abuse of notation, we also denote by \mathcal{F} . By the comments above this is still t -intersecting and has the same measure as the original family. The important property of the shifted family is that if it contains an edge F , it also contains every set obtained from F by shifting elements to the left, that is, by replacing some elements of F by smaller elements not in F .

We claim that in the shifted family we cannot have a set corresponding to a walk whose partial sums are all at most $t - 1$. This is because if we have such a set, we can show that it intersects some shifted copy of itself by less than t elements, contradiction. Indeed, let F be such a set. Using F , define another set G as follows. Consider the elements of F one by one, in order, starting with the smallest. The first (smallest) $t - 1$ elements of F stay in G . Each subsequent element of F in its turn is replaced by the smallest element which is not in F and is also not one of the elements placed already in G . We claim that in this process, every element of F besides the first $t - 1$ is replaced by a smaller element (which is not in F). Indeed, otherwise the first time in which the process fails to replace a member of F by a smaller member is some element f_{t-1+i} in F , where the elements of F are listed in increasing order, so that there are only $i - 1$ non-elements of F smaller than it. But this means that the random walk corresponding to F has $t - 1 + i$ times $+1$ and only $i - 1$ times -1 up to this point, meaning its value at this point is t , contradicting the assumption. Therefore G is obtained from F by left shifts, and as \mathcal{F} is shifted, G belongs to \mathcal{F} as well. But by construction G intersects F in only $t - 1$ elements, contradicting the assumption that \mathcal{F} is t -intersecting.

The claim about the measure follows, completing the proof. ■

We need the following standard estimate for Binomial distributions. See, e.g., [13], Theorem A.1.4.

Lemma 3.10. *Let $Y_i, 1 \leq i \leq n$ be independent identically distributed random variables where each Y_i is $+1$ with probability p and -1 with probability $1 - p$, and put $Y = \sum_{i=1}^n Y_i$. Then the probability that $Y - E(Y) \geq b$ is at most $e^{-b^2/2n}$.*

Corollary 3.11. Suppose $c \geq 2$, and put $p = \frac{c-1}{2c-1}$. Then:

(i) For all t and i , $W(p, t, i) \leq e^{-t/c}$. In particular, if $t \geq 2c^2$ then $W(p, t, i) < e^{-2c}$.

(ii) For all $t \geq 8c^2$, $W(p, t) < e^{-2c}$.

Proof. Part (i) follows by substituting $n = i$, $E(Y) = -\frac{i}{2c-1}$ and $b = t + \frac{i}{2c-1}$ in Lemma 3.10. This gives

$$W(p, t, i) \leq e^{-b^2/(2i)} \leq e^{-4it/[2i(2c-1)]} = e^{-2t/(2c-1)} \leq e^{-t/c},$$

as needed. To prove part (ii) note that if for a random walk X_1, X_2, X_3, \dots no partial sum $S_{it} = \sum_{j \leq it} X_j$ satisfies

$$(2) \quad S_{it} \geq t/2$$

then all partial sums S_i stay below t . We can thus bound $W(p, t)$ by the sum of probabilities of the events in (2), which we denote by E_i . By Lemma 3.10 the probability of E_i is at most

$$e^{-(\frac{it}{2c-1} + \frac{t}{2})^2/(2it)} \leq e^{-\frac{(i+c)^2 t}{8c^2 i}}.$$

The right hand side is at most $e^{-t/(2c)}$ for all i , since $(i+c)^2 \geq 4ic$, and it is also at most $e^{-it/(8c^2)}$ for all i . Therefore, for $t \geq 8c^2$, the sum over all $i \geq 1$ is smaller than

$$(3) \quad \sum_{i=1}^{8c^2} e^{-t/(2c)} + \sum_{i>8c^2} e^{-it/(8c^2)} < 8c^2 e^{-t/(2c)} + e^{-t}$$

where the last term is an upper estimate for the infinite geometric series $\sum_{i>8c^2} e^{-it/(8c^2)}$. For $t \geq 8c^2$ (and $c \geq 2$) the quantity in (3) is smaller than e^{-2c} , completing the proof. ■

Proof of Theorem 3.7. Let \mathcal{F} be a t -intersecting hypergraph, and let $[n]$ be its set of vertices. Add to the hypergraph any subset of $[n]$ that contains a member of \mathcal{F} and note that the modified hypergraph is still t -intersecting and its p -measure $\mu_p(\mathcal{F})$ is precisely the probability that a random subset of $[n]$ obtained by picking each element independently with probability p contains an edge of the hypergraph. Put $p = \frac{c-1}{2c-1}$, and let ε be smaller than $\binom{2c-1}{c-1}^{-1}$. Choose t_0 so that $W(p, t) < \varepsilon$ for all $t > t_0$. Note that by Corollary 3.11, part (ii) $t_0 \leq O(c^2)$. Now color randomly by $2c-1$ colors. The probability there is a set that gets only $c-1$ colors is bounded by $\binom{2c-1}{c-1} \mu_p(\mathcal{F})$, implying the desired result. ■

Remarks:

- The proof above together with Lemma 3.8 and Corollary 3.11, part (i) shows that the statement of Theorem 3.7 holds with $t_0 = 2c^2$ (with room to spare). Lemma 3.9 and Corollary 3.11, part (ii) provide a simple, self-contained proof that works with a somewhat larger value of t_0 (which is still $O(c^2)$).
- The above argument, with an appropriate choice of parameters, supplies a tradeoff between the number of colors used and the required size of the intersection. In particular it implies, for example, that $\chi(2c, c) \leq O(c)$.
- As mentioned above, if we apply random colorings, both the term $2c - 1$ and the $O(c^2)$ upper estimate for t_0 in Theorem 3.7 are tight. The fact that $2c - 1$ is tight for any fixed t is very simple, as mentioned above. Here is a sketch of the argument that for $2c - 1$ colors the $O(c^2)$ estimate for t is tight. Without making any attempt to optimize the constants, consider the family of all subsets of cardinality at least $n/2 + c^2/10000$ in an n element set $[n]$, where $n = (2c - 1)^3/10000$ and c is a large integer. Consider a random coloring of $[n]$ by $2c - 1$ colors. For a fixed color i , the expected number of elements colored i is $n/(2c - 1) = (2c - 1)^2/10000$ and the variance is $n \frac{1}{2c-1} (1 - \frac{1}{2c-1})$ which is roughly $(2c - 1)^2/10000$. Thus, the standard deviation is roughly $(2c - 1)/100$. Expose the color classes in order, two at a time, $c - 1$ times, leaving the final color class to the end. It is not difficult to show that for any given history, assuming that at least some $n/2c$ elements are not yet in the color classes exposed (as is the case with high probability) when we expose the next pair of color classes the probability that the difference between their sizes is at least, say, $c/200$, exceeds $1/2$. Thus with high probability we will have at least $c/4$ pairs with difference at least $c/200$. If this is the case, then by picking the larger color class of every pair we will cover at least $c/4 \times c/200 = c^2/800$ more elements than by picking the smaller class in each pair, and as with high probability the last color class is not bigger than $2 \cdot (2c - 1)^2/10000 < 8c^2/10000$ these $c - 1$ large color classes will contain, with high probability, a full edge. This shows that t_0 has to be at least $\Omega(c^2)$.
- The study of the random variant of the problem of determining $\chi(t, c)$ seems interesting. This is the problem of determining or estimating the smallest possible $f = f(t, c)$ so that a random vertex coloring of any t -intersecting hypergraph by f colors is c -strong with probability at least, say, 0.1 .

Note that the two functions f and χ differ. Indeed, the function $\chi(t, 2)$ is known for all values of t , as described in [16]. Specifically,

$\chi(0, 2) = \infty$, $\chi(1, 2) = 3$ and $\chi(t, 2) = 2$ for all $t \geq 2$. In contrast, it is easy to see that $f(0, 2) = f(1, 2) = \infty$. This is because for every fixed number of colors r , a random r -coloring of the vertices of a star with $m > r$ edges will contain a monochromatic edge with probability that tends to 1 as m tends to infinity. (The same argument implies that $f(c - 1, c) = \infty$ for all $c > 2$.) The arguments in [16] and here also show that $f(t, 2) = 3$ for all $t \geq 2$.

The results here and the earlier ones in [16] show that the function f is somewhat better understood than χ . In particular, we have shown here that for every c and all $t > 2c^2$, $f(t, c) = 2c - 1$.

4. APPLICATIONS IN THEORETICAL COMPUTER SCIENCE

The results and questions of Erdős have not been motivated by applications in Theoretical Computer Science (TCS), and yet the impact of his work on the development of TCS has been substantial. This short section includes some brief comments on this aspect of his work, focusing on applications of probabilistic techniques.

The Probabilistic Method plays a crucial role in the development of randomized algorithms. The quest for explicit constructions advocated time and again by Erdős is one of the early drives for derandomization – the process of converting randomized algorithms into deterministic ones. A specific problem he kept repeating over the years is that of finding explicit constructions of Ramsey graphs – graphs on n vertices in which the largest clique and largest independent set are of size $O(\log n)$, as well as explicit examples providing lower bounds for off-diagonal Ramsey number, like $r(3, n)$ – see [32].

The most successful attempts to find good explicit constructions of Ramsey graphs led to improved constructions of dispersers which are useful for derandomization, see [15]. Moreover, these constructions rely heavily on sum-product theorems initiated in the work of Erdős and Szemerédi [43] (although these are finite field analogs of the Erdős–Szemerédi results).

The method of conditional expectations, which is one of the very basic techniques in derandomization, was initiated in the paper of Erdős and Selfridge that introduced the study of combinatorial games [42].

Another useful technique which we only mention in passing is the Erdős–Rado delta-system (sunflower) method, that appears in work on circuit complexity and on matrix multiplication. A large body of work in Computational Geometry is also motivated by the results and questions of Erdős.

Finally, the area of Graph Property Testing (c.f., e.g., [13], Chapter 17), which is closely related to questions in computational learning and approximation algorithms, has its roots in old questions and results of Erdős. We do not include here a discussion of the general area, and only mention that one of the basic questions studied in it deals with the local and global nature of graph coloring. The specific question here is the ability to distinguish between graphs on n vertices that are k -colorable and graphs from which one has to delete at least εn^2 edges to get a k -colorable graph, by sampling a random induced subgraph on a small number of vertices. The first papers dealing with this question are [19] by Erdős and his collaborators and [65]. Better quantitative results appear in [48], where the systematic study of Graph Property Testing has been initiated, and in [9]. As is the case with so many other topics, the initial questions and results here can be traced back to the work of Paul Erdős.

Note added in proof: Very recently, Eberhard, Green and Manners have proved in [25] that the sum-free subset constant discussed in subsection 3.2 is in fact $1/3$. The problem of deciding whether or not every set of n nonzero integers contains a sum-free subset of cardinality at least $n/3 + w(n)$, where $w(n)$ tends to infinity with n , remains open.

REFERENCES

- [1] R. F. Ahlswede and L. H. Khachatrian, The complete intersection theorem for systems of finite sets, *European J. Combin.* 18 (1997), 125–136.
- [2] R. F. Ahlswede and L. H. Khachatrian, The diametric theorem in Hamming spaces – optimal anticodes, *Adv. in Appl. Math.* 20 (1998), 429–449.
- [3] M. Ajtai, J. Komlós and E. Szemerédi, A note on Ramsey numbers, *J. Combinatorial Theory Ser. A* 29 (1980), 354–360.
- [4] N. Alon, Degrees and choice numbers, *Random Structures & Algorithms* 16 (2000), 364–368.
- [5] N. Alon, A non-linear lower bound for planar epsilon-nets, *Proc. of the 51th IEEE FOCS* (2010), 341–346. Also: *Discrete and Computational Geometry* 47 (2012), 235–244.
- [6] N. Alon, Restricted integer partition functions, *Integers* 13 (2013), A16, 9pp.
- [7] N. Alon, B. Bukh and B. Sudakov, Discrete Kakeya-type problems and small bases, *Israel J. Math.* 174 (2009), 285–301.
- [8] N. Alon and D. J. Kleitman, Sum-free subsets, in: “A Tribute to Paul Erdős” (A. Baker, B. Bollobás and A. Hajnal eds.), Cambridge University Press, Cambridge, England 1990, 13–26.
- [9] N. Alon and M. Krivelevich, Testing k -colorability, *SIAM J. Discrete Math.* 15 (2002), 211–227.

-
- [10] N. Alon and A. V. Kostochka, Hypergraph list coloring and Euclidean Ramsey Theory, *Random Structures and Algorithms* 39 (2011), 377–390.
- [11] N. Alon, M. Krivelevich and B. Sudakov, Turán numbers of bipartite graphs and related Ramsey-type questions, *Combinatorics, Probability and Computing* 12 (2003), 477–494.
- [12] N. Alon and V. Rödl, Asymptotically tight bounds for some multicolored Ramsey numbers, *Combinatorica* 25 (2005), 125–141.
- [13] N. Alon and J. H. Spencer, **The Probabilistic Method**, Third Edition, Wiley, New York, 2008.
- [14] C. Bey and K. Engel, Old and new results for the weighted t -intersection problem via AK-methods, in: *Numbers, information and complexity* (Bielefeld, 1998), 45–74, Kluwer Acad. Publ., Boston, MA, 2000.
- [15] B. Barak, A. Rao, R. Shaltiel and A. Wigderson, 2-source dispersers for sub-polynomial entropy and Ramsey graphs beating the Frankl-Wilson construction, *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, ACM, New York, 2006, 671–680.
- [16] E. Blais, A. Weinstein and Y. Yoshida, Semi-strong coloring of intersecting hypergraphs, to appear.
- [17] T. Bohman, The triangle-free process, *Adv. Math.* 221 (2009), no. 5, 1653–1677.
- [18] B. Bollobás, **Random Graphs**, Second Edition, Academic Press, London, 2001.
- [19] B. Bollobás, P. Erdős, M. Simonovits and E. Szemerédi, Extremal graphs without large forbidden subgraphs, *Ann. Discrete Math.*, 3 (1978), pp. 29–41.
- [20] J. Bourgain, Estimates related to sumfree subsets of sets of integers, *Israel J. Math.* 97 (1997), 71–92.
- [21] E. R. Canfield and H. S. Wilf, On the growth of restricted integer partition functions, arXiv: 1009.4404, 2010.
- [22] F. Chung and R. L. Graham, **Erdős on Graphs: His Legacy of Unsolved Problems**, A. K. Peters, Wellesley, MA, 1998.
- [23] L. Danzer and B. Grünbaum, Über zwei Probleme bezüglich konvexer Körper von P. Erdős und von V. L. Klee, *Math. Z.* 79 (1962), 95–99.
- [24] I. Dinur and S. Safra, On the hardness of approximating minimum vertex cover, *Ann. of Math.* 162 (2005), 439–485.
- [25] S. Eberhard, B. Green and F. Manners, Sets of integers with no large sum-free subset, arXiv:1301.4579, 2013.
- [26] P. Erdős, Some remarks on the theory of graphs, *Bulletin of the Amer. Math. Soc.* 53 (1947), 292–294.
- [27] P. Erdős, Problems and results in additive number theory, *Colloque sur la Théorie des Nombres*, Bruxelles, 1955, 127–137, George Thone, Liège; Masson and Cie, Paris, 1956.
- [28] P. Erdős, Graph theory and probability II, *Canad. J. Math.* 13 (1961), 346–352.
- [29] P. Erdős, On a combinatorial problem, *Nordisk. Mat. Tidskr.* 11 (1963), 220–223.

- [30] P. Erdős, On a combinatorial problem II, *Acta. Math. Acad. Sci. Hungar.* **15** (1964), 445–447.
- [31] P. Erdős, Extremal problems in number theory, *Proc. Sympos. Pure Math.*, Vol. VIII (1965), pp. 181–189.
- [32] P. Erdős, On the construction of certain graphs, *J. Combinatorial Theory* 1 (1966), 149–153.
- [33] P. Erdős, Some recent results on extremal problems in graph theory, in: *Theory of Graphs (Rome, 1966)* Gordon and Breach, New York, 1967, 117–123.
- [34] P. Erdős and Z. Füredi, The greatest angle among n points in the d -dimensional Euclidean space, *Annals of Discrete Math.* **17** (1983), 275–283.
- [35] P. Erdős, R. L. Graham, P. Montgomery, B. L. Rothschild, J. Spencer and E. G. Straus, Euclidean Ramsey theorems. I. *J. Combinatorial Theory Ser. A* **14** (1973), 341–363.
- [36] P. Erdős, R. L. Graham, P. Montgomery, B. L. Rothschild, J. Spencer and E. G. Straus, Euclidean Ramsey theorems. II. Infinite and finite sets (Colloq., Keszthely, 1973) Vol. I, pp. 529–557. *Colloq. Math. Soc. János Bolyai*, Vol. 10, North-Holland, Amsterdam, 1975.
- [37] P. Erdős, R. L. Graham, P. Montgomery, B. L. Rothschild, J. Spencer and E. G. Straus, Euclidean Ramsey theorems. III. Infinite and finite sets (Colloq., Keszthely, 1973, Vol. I, pp. 559–583. *Colloq. Math. Soc. János Bolyai*, Vol. 10, North-Holland, Amsterdam, 1975.
- [38] P. Erdős and L. Lovász, Problems and results on 3-chromatic hypergraphs and some related questions, in: “Infinite and Finite Sets” (A. Hajnal et. al. eds), *Colloq. Math. Soc. J. Bolyai* **11**, North Holland, Amsterdam, 1975, pp. 609–627.
- [39] P. Erdős and D.J. Newman, Bases for sets of integers, *J. Number Theory* **9** (1977), 420–425.
- [40] P. Erdős, A. L. Rubin and H. Taylor, Choosability in graphs, *Proc. West Coast Conf. on Combinatorics, Graph Theory and Computing, Congressus Numerantium XXVI*, 1979, 125–157.
- [41] P. Erdős and P. Turán, On a problem of Sidon in additive number theory and some related problems, *J. London Math. Soc.* **16** (1941), 212–215; *Addendum* (by P. Erdős), *ibid.* **19** (1944), 208.
- [42] P. Erdős and J. L. Selfridge, On a combinatorial game, *J. Combinatorial Theory Ser. A* **14** (1973), 298–301.
- [43] P. Erdős and E. Szemerédi, On sums and products of integers, in: *Studies in pure mathematics*, Birkhäuser, Basel (1983), 213–218.
- [44] J. Fox and B. Sudakov, Dependent Random Choice, *Random Structures and Algorithms* **38** (2011), 1–32.
- [45] P. Frankl, Extremal Set Systems, in: R. L. Graham, M. Grötschel and L. Lovász, Editors, **Handbook of Combinatorics**, North Holland, Amsterdam, 1995.
- [46] E. Friedgut, On the measure of intersecting families, uniqueness and stability, *Combinatorica* **28** (2008), 503–528.
- [47] Z. Füredi, On a Turán type problem of Erdős, *Combinatorica* **11** (1991), 75–79.

- [48] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connection to learning and approximation, *J. ACM*, 45 (1998), pp. 653–750.
- [49] R. L. Graham, B. L. Rothschild and J. H. Spencer, **Ramsey Theory**, Second Edition, Wiley, New York, 1990.
- [50] W. T. Gowers, A new proof of Szemerédi’s theorem for arithmetic progressions of length four, *Geom. Funct. Anal.* 8 (1998), 529–551.
- [51] A. Hajnal and E. Szemerédi, Proof of a conjecture of P. Erdős, *Combinatorial theory and its applications, II* (Proc. Colloq., Balatonfüred, 1969), pp. 601–623. North-Holland, Amsterdam, 1970.
- [52] D. Haussler and E. Welzl, ϵ -nets and simplex range queries, *Discrete and Computational Geometry 2* (1987), 127–151.
- [53] P. E. Haxell and J. Verstraete, List coloring hypergraphs, *Electron. J. Combin.* 17 (2010), no. 1, Research Paper 129, 12 pp.
- [54] S. Janson, T. Łuczak and A. Ruciński, **Random Graphs**, Wiley, New York, 2000.
- [55] T. Jensen and B. Toft, **Graph Coloring Problems**, John Wiley and Sons Inc., New York, 1995.
- [56] J. H. Kim, The Ramsey number $R(3, t)$ has order of magnitude $t^2/\log t$, *Random Structures and Algorithms* 7 (1995), 173–207.
- [57] A. Kostochka and V. Rödl, *On graphs with small Ramsey numbers*, *J. Graph Theory* 37 (2001), 198–204.
- [58] M. Lewko, An improved upper bound for the sum-free subset constant, *J. Integer Seq.* 13 (2010), no. 8, Article 10.8.3.
- [59] Z. Ljajić and M. Nathanson, On a partition problem of Canfield and Wilf, *Integers*, 12A (2012), A11, 8pp.
- [60] M. Molloy and B. Reed, **Graph Coloring and the Probabilistic Method**, Springer-Verlag, Berlin, 2001.
- [61] J. Matoušek, R. Seidel and E. Welzl, How to net a lot with little: Small ϵ -nets for disks and halfspaces, In *Proc. 6th Annu. ACM Sympos. Comput. Geom.*, pages 16–22, 1990.
- [62] J. Pach and G. Tardos, Tight lower bounds for the size of epsilon-nets, *Computational geometry (SCG’11)*, ACM, New York, 2011, 458–463.
- [63] A. Pluhár, Greedy colorings of uniform hypergraphs, *Random Structures Algorithms* 35 (2009), no. 2, 216–221.
- [64] J. Radhakrishnan and A. Srinivasan, Improved bounds and algorithms for hypergraph two-coloring, *Random Structures and Algorithms* 16 (2000), 4–32.
- [65] V. Rödl and R. Duke, On graphs with small subgraphs of large chromatic number, *Graphs Combin.*, 1 (1985), pp. 91–96.
- [66] D. Saxton and A. Thomason, Hypergraph containers, to appear.
- [67] P. Turán, On the theory of graphs, *Colloquium Math.* 3 (1954), 19–30.
- [68] V. N. Vapnik and A. Ya Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, *Theory Probab. Appl.*, 16 (1971), 264–280.

- [69] V. G. Vizing, Coloring the vertices of a graph in prescribed colors (in Russian), *Diskret. Analiz.* No. 29, *Metody Diskret. Anal. v. Teorii Kodov i Shem* 101 (1976), 3–10.
- [70] H. Wan, Upper bounds for Ramsey numbers $R(3, 3, \dots, 3)$ and Schur numbers, *J. Graph Theory* 26 (1997), no. 3, 119–122
- [71] N. Wormald, The differential equations method for random graph processes and greedy algorithms, in: M. Karonski, H. J. Prömel (Eds.), *Lectures on Approximation and Randomized Algorithms*, PWN, Warsaw, 1999, pp. 73–155.

Noga Alon

Sackler School of Mathematics

and

Blavatnik School of Computer Science,

Tel Aviv University,

Tel Aviv 69978,

Israel

e-mail: `nogaa@tau.ac.il`

EUCLIDEAN VS. GRAPH METRIC

ITAI BENJAMINI

1. INTRODUCTION

The theory of sparse graph limits concerns itself with versions of local convergence and global convergence, see e.g. [44]. Informally, in local convergence we look at a large neighborhood around a random uniformly chosen vertex in a graph and in global convergence we observe the whole graph from afar. In this note rather than surveying the general theory we will consider some concrete examples and problems of global and local convergence, with a geometric viewpoint. We will discuss how well large graphs approximate continuous spaces such as the Euclidean space. Or how properties of Euclidean space such as scale invariance and rotational invariance can appear in large graphs.

The first sections consider approximating the Euclidean and Finsler metrics by graphs. We study the emergence of rotational, scale and conformal invariance in large graph metrics. We then move on to comment on random graph metrics. Starting with graphs obtained by perturbing the Euclidean metric, and then moving on to random graphs that are restricted to have a planar topology. In particular, we will study graphs generated by random subdivisions. Local and global graph limits will be woven into the whole discussion.

2. NOTIONS OF DISTANCE BETWEEN METRIC SPACES

Given a graph $G = (V, E)$, the *graph distance* between any two vertices is the length of the shortest path between them. A graph G is called *vertex transitive* if for any $u, v \in V$ there exists a graph automorphism mapping u to v .

In the note we will consider three notions of distances between metric spaces.

The first is that of quasi-isometry and slack-isometry between spaces, the second the Gromov-Hausdorff distance which is suitable for comparison between bounded spaces and is therefore useful for studying scaling limits, and the third is regarding a local statistical similarity between spaces.

See Burago and Ivanov [18] for background on metric spaces, including the first two notions and Lovász [44] for local limits.

Definition 2.1. Two metric spaces G and H are said to be *quasi-isometric* if there exists a map $f : G \rightarrow H$ and two constants $1 \leq C < \infty$ and $0 \leq c < \infty$, such that

- $C^{-1}d_H(f(x), f(y)) - c \leq d_G(x, y) \leq Cd_H(f(x), f(y)) + c$ for every $x, y \in G$,
- For every $y \in H$ there is an $x \in G$ so that $d_H(f(x), y) < c$.

Two metric spaces are said to be *slack-isometric* iff they are quasi-isometric with multiplicative constant equal to 1. That is, if we can take $C = 1$ in the definition.

For global convergence we use: the *Gromov-Hausdorff distance* between two metric spaces is obtained by taking the infimum over all the Hausdorff distances between isometric embeddings of the two spaces in a common metric space.

One way to look at a large finite graph is to look at a large neighborhood around a random uniformly chosen vertex. Often such neighborhood statistics capture quantities of interest and their asymptotics. Thus, one is led to take limits of such statistics and thereby define a probability measure on infinite rooted graphs, where the neighborhood of the root has the statistics that arise as the limit statistics of the finite graphs. Such a limit of a sequence of finite graphs is local limit. All such limit measures have a property known as unimodularity; it is not known whether all unimodular measures are limits of finite graphs. This fundamental question was asked in [2]. Those that are such limits are called sofic. Intuitively, a probability measure on rooted graphs is unimodular if its root is chosen “uniformly” from among all its vertices. This, of course, only makes sense for finite graphs. It is formalized for networks on infinite graphs by requiring a certain conservation property known as the Mass-Transport Principle, see [13] [2] [8].

For local limit we follow [13]: a *limit* of finite graphs G_n is a random rooted infinite graph (G, ρ) with the property that neighborhoods of G_n around a random vertex converge in distribution to neighborhoods of G around ρ .

Formally, let (G, o) and $(G_1, o_1), (G_2, o_2), \dots$ be random connected rooted locally finite graphs. We say that (G, o) is the limit of (G_j, o_j) as $j \rightarrow \infty$ if for every $r > 0$ and for every finite rooted graph (H, o') , the probability that (H, o') is isomorphic to a ball of radius r in G_j centered at o_j converges to the probability that (H, o') is isomorphic to a ball of radius r in G centered at o .

Given a (possibly random) graph we will consider the distribution on rooted graphs obtained by rooting at a random uniform vertex.

Exercise: what is the limit of n -level full binary trees?

Hint: it is not the infinite full binary tree.

In [13] it was shown that local limits of bounded degree graphs are a.s. recurrent for the simple random walk. A graph admits the $f(n)$ separation property if for any vertex set S in the graph of size n , by removing not more than $f(n)$ vertices from S , the connected components of S has size at most $|S|/2$.

Limits of graphs having $f(n)$ -separation function, for some $f(n)$, suggests studying quantitative versions of Elek's hyperfiniteness [44]. See [14] for more on separation.

Continuity of graph parameters with respect to local convergence is of current interest, here is one example.

Define $SAW(n)$ as the uniform measure on all the self-avoiding paths of length n from a fixed root. By sub-multiplicativity $\mu = \lim |SAW(n)|^{1/n}$ exists and is called the *connective constant* of the graph.

Conjecture 2.2. μ is continuous with respect to local convergence of infinite vertex transitive graphs.

We will also need the following notion,

Definition 2.3. Let $G = (V, E)$ be a finite graph. Define the Cheeger constant of G to be

$$h(G) = \inf_{0 < |S| < \frac{|V|}{2}} \frac{|\partial S|}{|S|}.$$

If G is an infinite graph we set

$$h(G) = \inf_{0 < |S| < \infty} \frac{|\partial S|}{|S|}.$$

An infinite graph G with $h(G) > 0$ is called non-amenable. Otherwise it is called amenable.

3. ROTATIONAL INVARIANCE

How well can the graph metric on bounded degree graphs approximate the metric of homogeneous manifolds equipped with some invariant length metric.

3.1. Slack Euclidean?

Recall that the scaling limit of the \mathbb{Z}^2 grid is the l^1 metric on the plane.

The following question was raised by Gady Kozma in a discussion with Oded Schramm and myself.

Question 3.1. *Is there a bounded degree graph which is slack-isometric to the Euclidean plane?*

The Pinwheel tiling, which is a non-periodic tiling defined by Charles Radin [50], is a graph quasi-isometric to the Euclidean plane where the multiplicative constant goes to 1 uniformly in the distance.

By sampling a Poisson process in the Euclidean plane and drawing the corresponding Voronoi tiles we get the *Poisson-Voronoi tessellation* (see Wikipedia). The graph metric on the tiles is almost surely has an asymptotically Euclidean metric see e.g. Howard-Newman [29].

Question 3.2. *What is the asymptotic shape of a ball in a Poisson-Voronoi tessellation where the underlying space is the plane with an l^p metric?*

See the closely related [19].

3.2. Near critical percolation

Can the l^2 or other given Finsler metric “naturally” emerge as a limit of bounded degree graph metrics in the Gromov-Hausdorff distance?

Consider the natural embedding of the square grid in the plane.

Dilute the planar square grid by *removing* edges independently with probability $q < 1/2$. Since $1/2$ is the critical percolation probability (Kesten [38]) almost surely there is a unique connected dense infinite subgrid left.

Condition on the origin to be in the infinite connected component and look at large balls rescaled to have diameter 1.

For any fixed q the subadditive ergodic theorem was used in the context of first passage percolation to show that the rescaled large balls around the origin will a.s. converge in the *Gromov-Hausdorff* distance to a centrally symmetric convex body in the Euclidean plane.

Conjecture 3.3. *As $q \rightarrow 1/2$ the limiting shape Gromov-Hausdorff converges to an Euclidean ball.*

Seems hard, since metric properties do not follow from conformal geometry. Yet simple simulations seem convincing.

4. SCALE INVARIANCE

4.1. Rotational and scale invariant Euclidean structures

Is there a distribution on tilings of the Euclidean plane which is rotation and translation invariant, mixing (that is, what is observed in far apart fixed Euclidean balls decorrelates with the distance between the balls), and stationary scale invariant (that is, there is a stationary matching or clustering of neighboring tiles resulting in a rescaled sample)? The Pinwheel tiling [50] is such. What if we further require spatial Markovity. That is given a tile you can not tell the tiling of the complement e.g. at which points of its boundary 3 tiles meet? Consider space filling Schramm's $SLE(8)$ curve and remove from it an independent Poisson process in the plane, the curve is then cut into pieces of finite area. As suggested by Wendelin Werner, variants on this observation might provide the exotic tilings we are after.

Aldous [1] initiated a study of random road networks whose distributions are *exactly* invariant under Euclidean scaling. He introduced a natural axiomatization of a class of structures he called *scale-invariant random spatial networks*, whose primitives are routes between each pair of points in the plane and constructed a model, based on minimum-time routes in a binary hierarchy of roads with different speed limits, satisfying the axioms.

We mention briefly an open problem of remotely similar spirit. Can you foliate \mathbb{R}^d with Brownian paths?

5. ONE LARGE SCALE CONTROL, SYMMETRIC GRAPHS

Let (G_n) be an unbounded sequence of finite, connected, vertex transitive graphs such that $|G_n| = o(\text{diam}(G_n)^d)$ for some $d > 0$. In [10] the following theorem is shown.

Theorem 5.1. *After taking a subsequence and rescaling by the diameter, the sequence (G_n) converges in the Gromov-Hausdorff distance to a torus of dimension $< d$, equipped with some invariant Finsler metric.*

In particular, if the sequence admits a doubling property at a large yet sub diameter scale, then the limit will be a torus equipped with some invariant length metric. Otherwise it will not converge to a finite dimensional manifold. When the degrees are uniformly bounded the limiting metric is a polygonal Finsler metric.

The proof relies on a recent quantitative version of Gromov's theorem on groups with polynomial growth obtained by Breuillard, Green and Tao [17] and a scaling limit theorem for nilpotent groups by Pansu [48]. See also Glander [32]. Establishing quantitative versions will have applications to random walks and percolation on vertex transitive graphs. For example in the spirit of Varopoulos' theorem that the only recurrent finitely generated groups have at most quadratic growth [52]:

Let G be a finite, d -regular connected vertex transitive graph. View G as an electrical network in which each edge is a one Ohm conductor.

Conjecture 5.2 (with Gady Kozma). *For any two vertices*

$$\text{electric resistance}(v, u) < C_d + \frac{\text{diam}^2(G) \log |G|}{|G|}.$$

In addition for a sequence of vertex transitive graphs, if the diameter is $o(|G_n|)$ then the electric resistance between any two vertices is $o(\text{diam}(G_n))$.

Since finite vertex transitive graphs, when they converge to a manifold, converge to a torus, it follows that the infimum, over all such, of the Gromov-Hausdorff distance to S^n is attained. Which one is the closest?

Question 5.3. *Is the skeleton of the truncated icosahedron (soccer ball) the closest to S^2 ?*

“Proof”: Otherwise we would have a different design for soccer balls. See also *Géode (géométrie)* in French Wikipedia.

5.1. Expander at all scales?

A sequence of graphs $\{G_n\}$ is of an expander if there is $h > 0$, for all n , $h(G_n) > h$.

Question 5.4. *Is there a family $\{G_n\}$ of finite d -regular graphs, $|G_n| \rightarrow \infty$, so that all the induced balls in all the G_n 's are expanders?*

That is, there is $h > 0$, for all $r > 0$ and any v in any of the graphs G_n 's the ball $B(v, r)$ is h - expander, expander with a uniform edge expansion

constant h . Note e.g. that if G_n is a sequence of expanders with girth growing to infinity, then if r is smaller than the girth then the balls of radius r are trees and thus they are not uniform expanders as r grows.

We *conjecture* that there is no such family. For vertex transitive graphs a positive answer to the following conjecture regarding percolation on expanders will show that no such family exists. The proof will proceed by constructing a limiting nonamenable vertex transitive graph with a unique infinite cluster whenever percolation occurs, we omit the outline.

Question 5.5. *Let G be a bounded degree expander, further assume that there is a fixed vertex $v \in G$, so that after performing $p = 1/2$ percolation on G ,*

$P_{1/2}$ (the connected component of v has diameter $> \text{diameter}(G)/2$) $> 1/2$,

Is there a giant component w.h.p? G is not assumed to be transitive.

The following two questions are regarding the rigidity of the global structure given local information.

Question 5.6. *Given a fixed rooted ball $B(o, r)$, assume there is a finite graph such that all its r -balls are isomorphic to $B(o, r)$, e.g. $B(o, r)$ is a ball in a finite vertex transitive graph, what is the minimal diameter of a graph with all of its r -balls isomorphic to $B(o, r)$? Any bounds on this minimal diameter, assuming the degree of o is d ? Any example where it grows faster than linear in r , when d is fixed?*

Note that some r -ball in the grandparent graph, or any infinite *non-unimodular* vertex transitive graphs, does not appear as a ball in a finite vertex transitive graph. As by [13] local limit of finite graphs is unimodular. When the rooted ball is a tree, this is the girth problem. One can consider a weaker version e.g. when we require only that most balls are isomorphic to $B(o, r)$. Not assuming a bound on the degree, consider the 3-ball in the hypercube, is there a graph with a smaller diameter than the hypercube so that all its 3-balls are that of the hypercube?

Question 5.7 (with Romain Tessera). *Let X is the Euclidean or hyperbolic plane, together with a triangulation, whose triangles are at most of diameter r . Suppose for each pair of Euclidean (or hyperbolic) balls of radius r , B_1, B_2 centered on vertices of this triangulation, there is a Euclidean (or hyperbolic) isometry mapping B_1 to B_2 respecting the triangulation (in the obvious way).*

Does it imply that the triangulation is periodic?

5.1.1. Roughly transitive graphs. A metric space X is (C, c) -roughly transitive if for every pair of points $x, y \in X$ there is a (C, c) -quasi-isometry sending x to y .

If G_n is only roughly transitive and $|G_n| = o(\text{diam}(G_n)^{1+\delta})$ for $\delta > 0$ sufficiently small, we are able to prove, this time by elementary means, that (G_n) converges to a circle.

Question 5.8. *Is there an infinite (C, c) -roughly transitive graph, with C, c finite, which is not quasi-isometric to a homogeneous space?*

Here a homogeneous space is a metric space with a transitive isometry group. The same question can be asked in the wider category of Coarse embeddings.

See [6] and references there for the study of quasi-isometry between random spaces.

6. PACKING

Packing one graph in another space can be viewed as *large scale-rough conformal geometry*. Large scale conformal geometry is developed in a work by Pierre Pansu [49]. We present a sample.

Question 6.1. *Which graphs can be realized as the nerve graph of a sphere packing in Euclidean d -dimensional space?*

Here vertices correspond to spheres with disjoint interiors and edges to pairs of touching spheres.

The rich two dimensional theory started with Koebe, who proved that every planar graph admits a circle packing.

In higher dimensions, Thurston observed that packability implies an upper bound of order $|G|^{(d-1)/d}$ on the size of minimal separators, see e.g. [46]. There is an emerging theory with many still open directions. Local graph limits were useful in the proof of the last two theorems below. Denote by T_3 the 3-regular tree.

Theorem 6.2 (with Oded Schramm). *The grid \mathbb{Z}^4 , $T_3 \times \mathbb{Z}$ and lattices in hyperbolic 4-space do not admit sphere packing in Euclidean \mathbb{R}^3 .*

Let (G_n) be a sequence of finite, $(k > 2)$ -regular graphs with girth growing to infinity.

Theorem 6.3. *For every d there exists an $N(d)$ such that G_n does not admit a regular sphere packing in Euclidean d -dimensional space, for any $n > N(d)$.*

The following is an extension to higher dimension of a theorem of Bowditch [16] following a suggestion by Gromov.

Theorem 6.4. *Let G be an infinite locally finite connected graph which admits a regular packing in \mathbb{R}^d . Then we have the following alternative: either G has a positive Cheeger constant, or there are arbitrarily large subsets S of G such that $|\partial S| < |S|^{\frac{d-1}{d} + o(1)}$.*

By *regularly* we mean uniform upper bound on the ratio of the radii of neighboring spheres. The proof of the last two theorems in [7] uses sparse graphs limits: by [13] local limits of bounded degree finite planar graphs are a.s. recurrent for the simple random walk, in [7] the proof was adapted to show that local limit of finite graphs that are regularly packed in \mathbb{R}^d , are d -parabolic. Which is the key to the results above.

Question 6.5 (with Oded Schramm). *Show that any packing of \mathbb{Z}^3 in \mathbb{R}^3 has at most one accumulation point in $\mathbb{R}^3 \cup \{\infty\}$.*

7. PERTURBING THE EUCLIDEAN METRIC

Some families of metric spaces are naturally parameterized by the reals. The critical spaces are usually more exotic. We will present a few examples. These spaces sometimes admit combinations of properties which are impossible in the vertex transitive world. We start with the classical model of first passage percolation for perspective.

7.1. First passage percolation

One natural way to randomly perturb the Euclidean planar metric is that of first passage percolation (FPP), see [39] and [33] for background. That is, consider the square grid lattice, denoted \mathbb{Z}^2 , and to each edge assign an i.i.d. random positive length. There are other ways to randomly perturb the Euclidean metric and many features are not expected to be model dependent. Large balls converge after rescaling to a convex centrally symmetric shape. Richardson (1973) proved the first shape theorem, when the length has exponential distribution and the graph is the \mathbb{Z}^d lattice. Simulations indicate

that the limiting shape is not the Euclidean ball. Kesten (unpublished) showed that the shape is not the Euclidean ball in high enough dimension.

The boundary fluctuations are conjectured to have a Tracy-Widom distribution. The variance of the distance from the origin to $(n, 0)$ is conjectured to be of order $n^{2/3}$. So far only an upper bound of $\frac{n}{\log n}$ was established, see [11]. Optimal bounds on the length of efficient algorithms for finding the shortest path or to estimate its length are still unknown.

The structure of geodesic rays and two-sided infinite geodesics in first passage percolation is still far from being understood. Furstenberg asked in the 80's (attending a talk by Kesten) to show that almost surely there are no two sided infinite geodesics for natural FPP's, e.g. exponential length on edges.

Hägström and Pemantle introduced [26] competitions based on FPP, see [23] for a survey. Here is a related problem. Start two independent simple random walks on \mathbb{Z}^2 walking with the same clock, with the one additional condition, that the walkers are not allowed to step on vertices already visited by the other walk, and otherwise choose uniformly among allowed vertices. Show that almost surely, one walker will be trapped in a finite domain. Prove that this is not the case in higher dimensions.

7.2. Perturbations, beyond first passage percolation

We now describe several random metrics, the first two of which can be viewed as perturbations of the grid like FPP, but with slightly stronger perturbation “causing the underlying grid metric to almost disappear”.

7.2.1. LRP. Start with the one dimensional finite grid $\mathbb{Z}/n\mathbb{Z}$ with the nearest neighbor edges, add additional edges to it as follows. Between, i and j add an edge with probability $\beta|i - j|^{-s}$, independently for any pair. The main problem in *long range percolation* is, how does the distance between 0 and $n/2$ typically grows in this random graph?

The off critical cases: when $s > 2$ the distance is of order n , for $1 < s < 2$ the distance is polylog n (see [15] for the exact result, background and history). For $s = 1$ Coppersmith, Gamarnik and Sviridenko showed that the distance is $\frac{\log n}{\log \log n}$ and if $s < 1$ the distance is uniformly bounded.

The critical case: when $s = 2$ the distance is of the form $\theta(n^{f(\beta)})$, where f is strictly between 0 and 1 (Sly and Ding [24]). Continuity, monotonicity, or even a guess of f are still open. We believe that there is a scaling limit for the $s = 2$ long range percolation random graphs.

These natural random graphs admit a combination of properties which is impossible for vertex transitive graphs. E.g. when $1 < s < 2$ the mixing time of the simple random walk is a.s. n^{s-1} . That is, small diameter does not exclude small bottlenecks as in vertex transitive graphs [5].

7.2.2. CCCP. Examine bond percolation on \mathbb{Z}^d . Each edge is open with probability p independently. Clusters are connected components of open edges. For any $d > 1$, there is $0 < p_c < 1$, such that if $p < p_c$ all the clusters are finite a.s. and the diameter of the clusters has exponential tail. If $p > p_c$ there is a unique infinite cluster. While for the critical probability p_c it is conjectured that there is no infinite cluster and that the diameter of clusters has polynomial tail. This is true in dimensions 2 and d large.

The unique infinite cluster, for $p > p_c$ is a random perturbation of the grid. E.g. asymptotics of the heat kernel are the same, how can we get “interesting” critical geometry?

Conditioning on the critical percolation to have an infinite cluster results in a “thin” graph with infinitely many cut points.

Here is a suggestion: contract each cluster into a single vertex. The result is a random graph G of high degree (each vertex $v \in G$ is a cluster \mathcal{C} in \mathbb{Z}^d and its degree is the number of closed edges coming out of \mathcal{C}). When the percolation is *subcritical* one expects to see a perturbation of the lattice, analogous to first passage percolation. When the percolation is *critical* the random geometric structure obtained is rather different.

We refer to the above random graph G as CCCP (Contracting Clusters of Critical Percolation). For example (with Ori Gurel-Gurevich and Gady Kozma) we have: when $d = 2$, the CCCP has exponential volume growth a.s. When $d > 6$ a.s. the CCCP has double-exponential volume growth.

8. RANDOM PLANAR METRIC

Above we reviewed random perturbations of the Euclidean plane. How to define and model a genuine random planar metric?

8.1. Local convergence

Plane topology. Angel and Schramm [3, 4] constructed the uniform infinite planar triangulation (UIPT), a rooted infinite random triangulation which is the limit (in the sense of [13]) of finite random triangulations (the uniform measure on all nonisomorphic triangulations of the sphere of size n), a model that was studied extensively by many (see e.g. [41]). The UIPQ is a similar

construction with quadrangulation. The UIPT/Q looks very different from random perturbations of the plane as in the Poisson-Voronoi triangulation and has a rather surprising geometry at first encounter, e.g. volume growth of balls in the UIPT is asymptotically r^4 . The UIPQ is recurrent [34] and subdiffusive [9] for the simple random walk and in particular hyperfinite. A collection of graphs is hyperfinite if for every $\varepsilon > 0$ there is some finite k such that each graph G in the collection can be broken into connected components of size at most k such that each has a boundary of size at most ε of its size. What about a hyperbolic nonhyperfinite counterpart?

Hyperbolic analog? Guth, Parlier and Young [35] studied pants decomposition of random closed surfaces obtained by randomly gluing N Euclidean triangles (with unit side length) together. They gave bounds on the size of pants decomposition of random compact surfaces with no genus restriction as a function of N . Their work indicates that the injectivity radius around a typical point is growing to infinity. Gamburd and Makover [30] showed that as N grows the genus will converge to $N/4$ and by Euler's characteristic the average degree will grow to infinity. What about a local limit of random finite triangulation/quadrangulation with genus growing linearly in the number of quadrangulation.

In the quadrangulation bijective techniques help a lot see [51]. In particular, Chassaing and Durhuus constructed the UIPQ from a random infinite labeled tree, followed by another construction in [21] from a labeled *critical* geometric Galton-Watson tree conditioned to survive. With Nicolas Curien we propose a model of infinite random quadrangulation constructed similarly from a labeled *supercritical* Galton-Watson tree. We *conjecture* that such a *stochastic hyperbolic infinite quadrangulation (shiq)* describes the limit of random finite quadrangulations with genus growing linearly in the number of quadrangulation. The Shiq is not hyperfinite and the simple random walk on the Shiq has positive speed almost surely.

Kaibel and Ziegler [37] survey a model of random lattice triangulations. They proved the existence of local limit and studying its properties, such as volume growth, seems interesting.

8.2. Global convergence

Scaling limits of random triangulations were also studied, see Le Gall [43] and Miermont [45] advancing over [20], who proved that the random triangulations scaled Gromov-Hausdorff converge to a random compact metric space of dimension 4. This limiting surface called the Brownian map can be seen as the two-dimensional sphere equipped with a random metric which induces the usual topology but makes it a fractal space of Hausdorff di-

mension 4. It is of interest to obtain quantitative estimates on the rate of convergence as in the Hungarian coupling of random walks and Brownian motion [40]. Also this theory is believed to be connected to 2D quantum gravity and conformal invariance via the following construction:

Conformal invariance. Let T_n be is a uniform triangulation of the sphere with n faces. It is possible to get a “canonical” drawing of T_n on the sphere by conformal tools. E.g. if T_n has no loops or multiple edges, we can use the well-known circle packing theorem (see Wikipedia, [27]):

Theorem 8.1. *If T is a finite triangulation without loops or multiple edges then there exists a circle packing $P = (P_c)_{c \in C}$ in the sphere \mathbb{S}_2 such that the contact graph of P is T . This packing is unique up to Möbius transformations.*

The circle packing enables us to take a “nice” representation of a triangulation, nevertheless the non-uniqueness is somehow disturbing because to fix a representation we can, for example, fix the images of three uniformly chosen vertices of T_n . Once this is done, we form the atomic measure μ_{T_n} formed by the Dirac’s at centers of the circles of the packing of T_n renormalized to have mass one. This constitutes a canonical discrete conformal random probability on the sphere. By standard arguments there exist weak limits μ_∞ of μ_{T_n} . Here are some tougher questions:

Questions

1. (Schramm [Talk about QG]) Determine coarse properties (invariant under $\text{SO}_3(\mathbb{R})$) of μ_∞ , e.g. what is the dimension of the support? Start with showing singularity.
2. Uniqueness (in law) of μ_∞ ? In particular can we describe μ_∞ in terms of Gaussian Free Field (GFF)?
Is it $\exp((8/3)^{1/2}GFF)$, does KPZ hold? See [25].
3. The random measure μ_∞ can come together with d_∞ a random distance on \mathbb{S}_2 (in the spirit of [42]). Can you describe links between μ_∞ and d_∞ ? Does one characterize the other? Is it a path metric space? See [31].

8.4. Recursive subdivision

Important properties of the UIPT holds for a larger family of planar graphs. Start with a finite directed graph and two marked vertices, one with one edge going out and one with one edge coming in and no other edges. Recursively replace each edge with a copy of the graph with the marked vertices mapped

to the two vertices defining the edge. Extension of this scheme by recursively replacing fixed subgraphs results in infinite graphs admitting the *doubling property*: There is $C < \infty$, such that for any r , the size of any ball of radius $2r$ in G is bounded by C times the size of a ball of radius r . For example the graph of the Sierpinski gasket satisfies this property.

By recursive subdivision one can construct planar graphs that have polynomial growth with arbitrarily large exponent. Still all these graphs are small in the following two senses. First, local limits of sequences of bounded degree planar graphs obtained by taking consecutive subdivisions are recurrent [13]. Second, in [12] the following is proved.

Theorem 8.2. *Let G be a planar graph such that the volume function of G satisfies the doubling property. Then for every vertex v of G and radius r , there is a connected subset Ω such that $B(v, r) \subset \Omega$, $\Omega \subset B(v, 6r)$ and the size of the boundary of Ω is at most of order r .*

Try to imagine the geometry of a planar recursive subdivision graph, when the volume growth is faster than quadratic. The facts above suggest heuristically that volume is generated by large fractal mushrooms like folds, and that the complements of balls have many connected components.

In particular we *conjecture* that the simple random walk spends a long time in such traps and hence is subdiffusive (that is, $\text{dist}(o, X_n) \asymp n^\alpha$ where X_n denotes the simple random walk starting at o and $\alpha < 1/2$). Is the critical probability of site percolation on any planar triangulation of uniform growth faster than quadratic $1/2$?

Here is a sketch of the proof of theorem 8.2. Let v be any vertex of Γ . Consider the balls $B(v, n)$, $B(v, 3n)$. Let N be an n -net of the boundary $\partial B(v, 3n)$. For each vertex w of N consider $B(w, n/2)$. Note that all such balls are disjoint since N is an n -net. Also all these balls are contained in $B(v, 4n)$. So, by the doubling property, we can have only boundedly many such balls, that is $|N| \leq \beta$, where β does not depend on n . Consider now the balls $B(w, 2n)$ for all $w \in N$. $\partial B(v, 3n)$ is contained in the union of these balls. Construct a closed curve that ‘blocks’ v from infinity as follows: if $w_1, w_2 \in N$ are such that $d(w_1, w_2) \leq 2n$ then we join them by a geodesic. So replace $\partial B(v, 3n)$ by the ‘polygonal line’ that we define using vertices in N . This ‘polygonal line’ blocks v from infinity and has length at most $2n\beta$. There are some technical issues to take care of, for example $\partial B(v, 3n)$ might not be connected (and could even have ‘large gaps’) and the geodesic segments have to be chosen carefully.

9. RANDOM SUBDIVISION

There is growing interest in establishing a rigorous theory of two dimensional continuum quantum gravity. Heuristically, quantum gravity is a metric chosen on the sphere uniformly among all possible metrics. Although there are successful discrete mathematical quantum gravity models, we do not yet have a satisfactory continuum definition of a planar random length metric space (rather than random measure). One possible toy model is to start with a unit square, divide it into four squares and then recursively at each stage pick a square uniformly at random from the current squares (ignoring their sizes) and divide it to four squares and so on. Look at the minimal number of squares needed in order to connect the bottom left and top right corner with a connected set of squares. We *conjecture* that there is a deterministic scaling function, such that after dividing the random minimal number of squares needed after n subdivisions by it, the result is a non degenerate random variable. Establishing the conjecture will provide a random planar length metric space. Does the shortest geodesic stabilizes as we further divide?

Since the conjecture seems hard, we start by studying the simplest recursive constructions after trees. As you will see below even here we mostly have questions and conjectures. The section is based on an ongoing project with Nicolas Curien.

9.1. (Fixed) Hierarchical graphs

Let us introduce the graphs we will work with. We start with a pattern, that is a finite connected graph G_1 with two distinguished point “source” and “sink” and such that the edges are oriented from source to sink. Inductively, the graph G_n is constructed from G_{n-1} by replacing each of its (oriented) edge by a copy of G_1 (source and sink respectively on the origin and target of the edge), see Fig. below.

9.2. Distance

Fix a pattern G_1 and consider the sequence of hierarchical graphs G_1, G_2, \dots constructed as above. We endow these graphs with a random distance (or first passage percolation) model on them: assign a positive weight (e.g. uniform over $[0, 1]$) independently for each edge of G_n . Recall that G_n has two distinguished points “source” and “sink” and put

$$D_n := \text{Weight of a minimal path linking source and sink in } G_n.$$

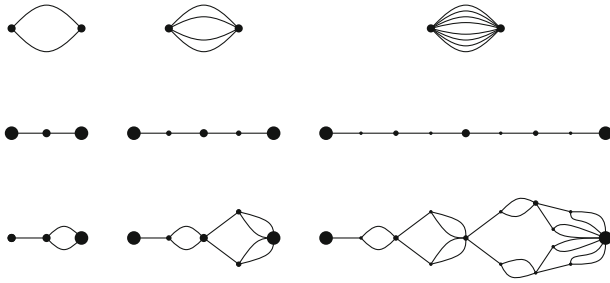


Fig. 1. A few examples of hierarchical graphs

Obviously the D_n 's satisfy a recursive distributional equation that is closely related to the initial pattern, e.g. for the three examples presented above we have for all $n \geq 2$

$$(1) \quad \begin{aligned} D_n &\stackrel{(d)}{=} \min(D_{n-1}, D'_{n-1}) \\ D_n &\stackrel{(d)}{=} D_{n-1} + D'_{n-1} \\ D_n &\stackrel{(d)}{=} D_{n-1} + \min(D'_{n-1}, D''_{n-1}), \end{aligned}$$

where $D_{n-1}, D'_{n-1}, D''_{n-1}$ are independent copies of D_{n-1} . The first two equations are straightforward to analyze but the last one is thorny because the recursive distributional equation combines $+$ (adding an edge in series) and \min (presence of cycles). We focus on the last case. Let us consider a (well-known) simplified model for the sake of comparison:

COMPARISON WITH BRANCHING RANDOM WALK. Consider T_n the full binary tree starting with an edge up to level n where each edge has been given an independent weight as above. In this case, the weight of the shortest path M_n up to level n satisfies

$$M_n = \xi + \min(M_{n-1}, M'_{n-1})$$

where ξ denotes the law of the weights on the edges. In this model (first passage percolation on a tree) we know that $M_n \approx \gamma_{brw} n$ with γ_{brw} explicit in terms of ξ as well as the lower order terms. This is due to the fact that the geometry of the tree does not constrain the model too much and in that case M_n is nearly obtained by considering all paths as independent. Also, a fairly simple argument due to Dekking and Host [23] shows that M_n is strongly concentrated (order $O(1)$) around its mean. Let us sketch it. Provided that ξ is bounded we can write

$$M_n \leq C + \min(M_{n-1}, M'_{n-1}).$$

Assume now that M_{n-1} is *not* concentrated around its mean, the key is to notice that in this case we have

$$E[\min(M_{n-1}, M'_{n-1})] \text{ sensibly less than } E[M_{n-1}].$$

Taking expectation we deduce that $E[M_n]$ is noticeably less than $E[M_{n-1}] + C$ however this cannot be the case since $E[M_n] \geq E[M_{n-1}]$.

Coming back to (1). We will compare D_n with M_{2^n} (the 2^n comes from the fact that the height of the graph G_n is 2^n compared to the height n of T_n). Clearly we have $D_n \leq 2^n$ and one can also show by induction that $D_n \geq M_{2^n}$, indeed notice that

$$M_{2^n} \leq M_{2^{n-1}} + \min(M'_{2^{n-1}}, M''_{2^{n-1}}),$$

and then use (1). Hence we have $\gamma_{brw} 2^n \leq E[D_n] \leq 2^n$ and a simple monotonicity argument shows that if ξ is non-degenerate then $\gamma_{rec} := \lim 2^{-n} E[D_n]$ exists and is in $[\gamma_{brw}, 1)$. In view of these remarks we have the following.

Question 9.1. *Compute γ_{rec} in terms of ξ in particular show (if true) $\gamma_{rec} > \gamma_{brw}$.*

We think that the convergence in mean of D_n implies (thanks to (1)) its convergence in probability. However subtle questions about D_n remain open.

Question 9.2. *What is the concentration of D_n around its mean? Lower order terms? More generally, ask the same questions as for the minimal position in a branching random walk.*

For background on branching random walks see e.g. [53].

9.3. External DLA

In the hierarchical graph G_n we launch particles one by one from the sink. The particle performs SRW and settles as soon as it hits a vertex adjacent to the source or previously settled particle. This is the standard model of External Diffusion Limited Aggregation on G_n . This process ends when a particle settles at the sink.

What is the proportion of G_n that is covered?

We denote by P_n the number of particles launched before the end of the process. Using the recursive structure of the graph G_n we can also write

a recursive distributional equation for P_n e.g. in the third case of Fig. 1 neglecting a few terms we have

$$(2) \quad P_n = P_{n-1} + 2 \min(P'_{n-1}, P''_{n-1}).$$

Compare with (1) (the presence of the factor 2 stems from the fact that the particles starting from the sink in G_n are (roughly speaking) split into two equal groups of particles in the two branches of the initial G_1). Note that the number of edges in G_n is 3^n so knowing whether G_n is almost full of particles is the same as knowing whether $E[P_n]$ is sensibly less than $3E[P_{n-1}]$ or not. Notice that if P_{n-1} is not concentrated then $2 \min(P'_{n-1}, P''_{n-1})$ is say less than $(2 - \varepsilon)E[P_{n-1}]$ thus $E[P_n] < (3 - \varepsilon)D_n$. But if P_{n-1} is concentrated we cannot say anything.

Question 9.3. *What is $\lim n^{-1} \log(E[P_n])$?*

9.3.1. The win-win situation Knowing whether the graph is full or not can be answered for a special type of recursive graph where “a shadowing effect”[©] takes place. Indeed, consider the sequence of graphs G_n with initial pattern $\bullet \diamond \bullet$, its fourth iteration is the figure below. In this case

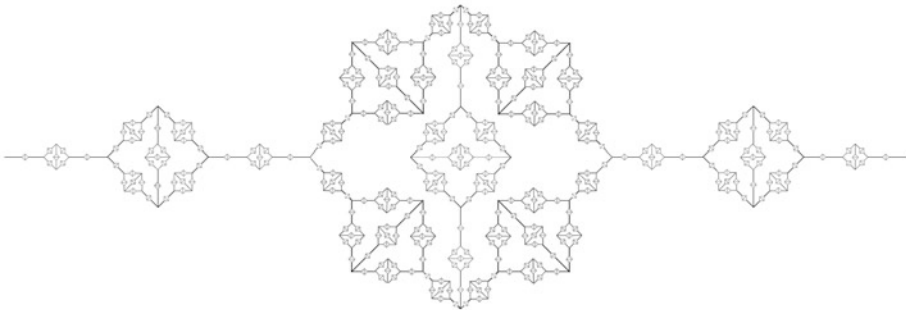


Fig. 2. Naomi's fractal

we can still write recursive distributional equations for the P_n but the heuristic argument goes as follows. Notice first that the volume of the graph grows like 7^n so we have to compare $E[P_n]$ with $7E[P_{n-1}]$. If P_{n-1} is not concentrated then $E[P_n] < (7 - \varepsilon)E[P_{n-1}]$ as above and we are done. So assume that P_{n-1} is concentrated. In the filling process of G_n the (offspring of the) first branch in G_1 linked to the source will be filled first which takes a time P_{n-1} and then the two branches adjacent to this one start to be filled. The key point is to notice that since P_{n-1} is concentrated, these two branches will be totally full at roughly the same time. In which case the

branch of the “middle” will receive no particles due to a shadowing effect of the last two branches. Finally in this case we expect $E[P_n] \approx 6E[P_{n-1}]$. We thus see that in all situations $E[P_n] < (7 - \varepsilon)E[P_{n-1}]$ and the following result essentially follows.

Proposition 9.4 (with Nicolas Curien). *We have $\limsup n^{-1} \log(E[P_n]) < 7$ and hence the graph G_n is not totally filled during the EDLA process, more precisely the aggregate covers a fractal portion of it.*

It will be nice to show the same for other fractals, starting with Sierpinski gasket.

9.4. Random hierarchical graphs

In this section, the graph we build are themselves random but still based on a hierarchical procedure. Let us describe one possible model. We start with a pattern G_1 . To get G_n from G_{n-1} we first pick one edge of G_{n-1} uniformly at random and replace it by a copy of the initial pattern G_1 . See Fig. ref below for an example. Using connection with branching processes, Thomas

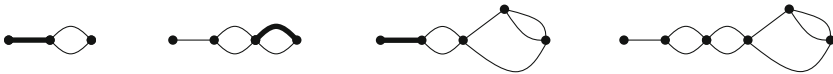


Fig. 3. Construction with the third pattern of Fig. 1

Duquesne (private communication) has been able to compute exactly the expectations of the number of oriented paths going from left to right in G_n . We denote by \mathcal{D}_n the distance between the two extremal points in G_n . Trivially $\mathcal{D}_n \leq n + 1$. A fairly simple sub-additivity argument shows that in fact

$$\frac{\log(E[\mathcal{D}_n])}{\log(n)} \xrightarrow{n \rightarrow \infty} \gamma \in [0, 1].$$

Ad-hoc calculations show that $\gamma \in (\varepsilon, 1/2 - \varepsilon)$. But the true value of γ remains mysterious. This model is intimately connected to a urn model: The volume of the graphs offspring of the three original edges form a standard Polya urn¹. So the limiting proportions of edges in these graphs $(\alpha_1, \alpha_2, \alpha_3)$ is distributed as a Dirichlet distribution of parameters $(1/2, 1/2, 1/2)$. Thus, loosely speaking, the recursive distributional equations satisfied by the \mathcal{D}_n 's are the following

$$\mathcal{D}_n \stackrel{(d)}{=} \mathcal{D}_{\alpha_1 n} + \min(\mathcal{D}'_{\alpha_2 n}, \mathcal{D}''_{\alpha_3 n}),$$

¹three balls of three colors initially, when a ball is picked it is replaced in the urn together with 2 balls of the same color

where (D_n) , (D'_n) and (D''_n) are copies of the original process and independent of the $(\alpha_1, \alpha_2, \alpha_3)$. In this model, the non-concentration of \mathcal{D}_n is granted so the interesting questions are the following.

Question 9.5. *What is the value of γ ? Can we rescale \mathcal{D}_n to have convergence in distribution? (this is equivalent to the Gromov-Hausdorff convergence of the rescaled graphs).*

Finally, we mention a last model in the same spirit. This is the series-parallel random graph introduced by Hambly and Jordan [36]. Fix a parameter $p \in [0, 1]$. The construction goes as follows. We start with a single edge. Then inductively at each stage, all the edges of the graph are replaced by two edges in series with probability p or two edges in parallel with probability $1 - p$. If Δ_n is the distance between the two extremal points in this graph then the recursive distributional equations are now

$$\Delta_n \stackrel{(d)}{=} \begin{cases} \text{with proba } p, & \Delta_{n-1} + \Delta'_{n-1} \\ \text{with proba } 1 - p, & \min(\Delta_{n-1}, \Delta'_{n-1}). \end{cases}$$

It is easy to see that when $p < 1/2$ then Δ_n remains bounded. However, when $p > 1/2$ this distance grows exponentially with n and by a subadditivity argument we get

$$E[\Delta_n] \approx e^{n\delta(p)+o(n)}.$$

Question 9.6. *What is the shape of $p \in [1/2, 1] \mapsto \delta(p)$. In particular, do we have $\delta(1/2) = 0$?*

Acknowledgements: Thanks to Nicolas Curien for substantial help with the writing and Naomi Benjamini for the drawing.

REFERENCES

- [1] D. Aldous, Scale-invariant random spatial networks, arXiv:1204.0817
- [2] D. Aldous and R. Lyons, Processes on unimodular random networks, *Electron. J. Probab.* paper 54, pages 1454-1508 (2007).
- [3] O. Angel, Growth and percolation on the uniform infinite planar triangulation, *Geometric And Functional Analysis* **13** 935–974 (2003).
- [4] O. Angel and O. Schramm, Uniform infinite planar triangulations, *Comm. Math. Phys.* **241**, 191–213 (2003).

- [5] L. Babai and M. Szegedy, Local expansion of symmetrical graphs, *Combinatorics, Probability and Computing* (1992), 1: 1–11.
- [6] R. Basu and A. Sly, Lipschitz embeddings of random sequences, arXiv:1204.2931
- [7] I. Benjamini and N. Curien, On limits of graphs sphere packed in Euclidean space and applications. *European J. Comb.* 32 975–984, (2011).
- [8] I. Benjamini and N. Curien, Ergodic theory on stationary random graphs, *Electron. J. Probab.* **17** 20 pp. (2012).
- [9] I. Benjamini and N. Curien, Simple random walk on the uniform infinite planar quadrangulation: Subdiffusivity via pioneer points, (2012) GAFA to appear.
- [10] I. Benjamini, H. Finucane and R. Tessera, On the scaling limit of finite vertex transitive graphs with large diameter, arXiv:1203.5624
- [11] I. Benjamini, G. Kalai and O. Schramm, First passage percolation has sublinear distance variance, *Ann. of Prob.* **31** 1970–1978 (2003).
- [12] I. Benjamini and P. Papasoglu, Growth and isoperimetric profile of planar graphs, *Proc. Amer. Math. Soc.* 139 (2011), no. 11, 4105–4111.
- [13] I. Benjamini and O. Schramm, Recurrence of Distributional Limits of Finite Planar Graphs, *Electron. J. Probab.* **6** 13 pp. (2001).
- [14] I. Benjamini, O. Schramm and A. Timar, On the separation profile of infinite graphs, *Group, Geometry and Dynamics.* **6** pp. 639–658 (2012).
- [15] M. Biskup, On the scaling of the chemical distance in long-range percolation models, *Ann. Prob.* **32** 2938–2977 (2004).
- [16] B. Bowditch, A short proof that a subquadratic isoperimetric inequality implies a linear one, *Michigan Math. J.* 42 (1995) 103–107.
- [17] E. Breuillard, B. Green and T. Tao, The structure of approximate groups, arXiv:1110.5008
- [18] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry, Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, (2001).
- [19] D. Burago and S. Ivanov, Uniform approximation of metrics by graphs, arXiv:1210.2435
- [20] P. Chassaing and G. Schaeffer, Random planar lattices and integrated superBrownian excursion, *Prob. Theor. and Rel. Fields* **128**, 161–212 (2004).
- [21] N. Curien, L. Ménard and G. Miermont, A view from infinity of the uniform infinite planar quadrangulation. <http://arxiv.org/abs/1201.1052>
- [22] M. Deijfen, O. Häggström, The pleasures and pains of studying the two-type Richardson model, *Analysis and Stochastics of Growth Processes and interface models*, 39–54 (2008).
- [23] F. Dekking and B. Host, Limit distributions for minimal displacement of branching random walks, *Probab. Theory Relat. Fields* 90, 403–426 (1991).
- [24] J. Ding and A. Sly, Distances in critical long range percolation, arXiv:1303.3995
- [25] B. Duplantier and S. Sheffield, Liouville quantum gravity and KPZ, *Inventiones Mathematicae* 185, 333–393 (2011).

- [26] O. Häggström and R. Pemantle, First passage percolation and a model for competing spatial growth, *Jour. Appl. Proba.* **35** 683–692 (1998).
- [27] Z-X. He and O. Schramm, Hyperbolic and parabolic packings, *Jour. Discrete and Computational Geometry* **14** 123–149, (1995).
- [28] C. Hoffman, Geodesics in first passage percolation, *Ann. Appl. Prob.* **18**, 1944–1969 (2008).
- [29] D. Howard and C. Newman, Euclidean models of first-passage percolation, *Probab. Theory Related Fields* **108**, 153–170 (1997).
- [30] A. Gamburd and E. Makover, On the genus of a random Riemann surface, *Contemp. Math.* **311** (2002), 133–140.
- [31] C. Garban, R. Rhodes and V. Vargas, Liouville Brownian motion, <http://arxiv.org/abs/1301.2876>
- [32] T. Gelander, A metric version of the Jordan–Turing theorem, arXiv:1205.6553
- [33] G. Grimmett and H. Kesten, Percolation since Saint-Flour, arXiv:1207.0373
- [34] O. Gurel-Gurevich and A. Nachmias, Recurrence of planar graph limits, *Annals of Math.* To appear (2012).
- [35] L. Guth, H. Parlier and R. Young, Pants decompositions of random surfaces, arXiv:1011.0612
- [36] B. Hambly and J. Jordan, A random hierarchical lattice: the series-parallel graph and its properties, *Adv. Appl. Prob.* **36**, 824–838 (2004).
- [37] V. Kaibel and G. Ziegler, Counting lattice triangulations, *London Math. Society Lecture Notes Series*, **307**, 277–307 (2003).
- [38] H. Kesten, The critical probability of bond percolation on the square lattice equals $1/2$, *Comm. Math. Phys.* **74** (1980), no. 1, 41–59.
- [39] H. Kesten, Aspects of first passage percolation, *Lecture Notes in Math*, Springer (1986).
- [40] J. Komlos, P. Major and G. Tusnady, An approximation of partial sums of independent RV'-s, and the sample DF. I, *Prob. Theor. and Rel. Fields.* **32**, (1975), 111–131.
- [41] M. Krikun, A uniformly distributed infinite planar triangulation and a related branching process, *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, **307** (Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 10):141–174, 282–283 (2004).
- [42] J. F. Le Gall, The topological structure of scaling limits of large planar maps, *Invent. Math.*, **169**(3):621–670 (2007).
- [43] J. F. Le Gall, Uniqueness and universality of the Brownian map, arXiv:1105.4842
- [44] L. Lovász, Large networks, graph homomorphisms and graph limits, *Amer. Math. Soc.* In preparation (2012).
- [45] G. Miermont, The Brownian map is the scaling limit of uniform random plane quadrangulations, arXiv:1104.1606
- [46] G. Miller, S. Teng, W. Thurston and S. Vavasis, Geometric separators for finite-element meshes, *SIAM J. Sci. Comput.* **19** (1998), no. 2, 364–386 (electronic).

- [47] V. Nekrashevych and G. Pete, Scale-invariant groups, *Groups, Geometry and Dynamics* 5 (2011), 139–167.
- [48] P. Pansu, Croissance des boules et des géodésiques fermées dans les nilvariétés, *Ergodic Theory Dyn. Syst.* 3, 415–445 (1983).
- [49] P. Pansu, Large scale conformal maps, Preprint (2012).
- [50] C. Radin, The Pinwheel tilings of the plane, *Annals of Mathematics* 139 (3): pp.661–702 (1994).
- [51] G. Schaeffer, Conjugaison d’arbres et cartes combinatoires aléatoires, phd thesis, (1998).
- [52] N. Varopoulos, L. Saloff-Coste and T. Coulhon, *Analysis and geometry on groups*, Cambridge University Press (1993).
- [53] O. Zeitouni, Branching random walks and Gaussian fields Notes for Lectures, <http://www.wisdom.weizmann.ac.il/~zeitouni/pdf/notesBRW.pdf>

Itai Benjamini

Weizmann Institute of Science,
234 Herzl St.,
Rehovot 76100,
Israel

e-mail: `itai.benjamini@gmail.com`

THE PHASE TRANSITION IN THE ERDŐS–RÉNYI RANDOM GRAPH PROCESS

BÉLA BOLLOBÁS* and OLIVER RIORDAN

We shall review the foundation of the theory of random graphs by Paul Erdős and Alfréd Rényi, and sketch some of the later developments concerning the giant component, including some very recent results.

1. INTRODUCTION

The theory of random graphs was founded in the late 1950s and early 1960s by the serendipitous partnership of Paul Erdős and Alfréd Rényi. Although they both worked in combinatorics *and* in probability, Erdős was the quintessential combinatorialist and Rényi the quintessential probabilist: working together, their formidable partnership was ideal for laying the foundations of a cohesive theory of random graphs. In this paper we shall review some of their ground-breaking results together with recent developments concerning the phase transition in graphs and in hypergraphs.

Our paper is organized as follows. In the next three sections we shall present some of the highlights of the work of Erdős and Rényi on the foundation of the theory of random graphs, emphasizing their ground-breaking results on the phase transition, the sudden emergence of the ‘giant component’ as our random graph acquires more and more edges. Section 5 is about the re-awakening of the interest in this phase transition, and the first results on its finer nature, correcting some misconceptions, together with a number of related results. In Section 6 we present some of the recent results proved on the phase transition in the standard random graph process. We do not and cannot aim for a comprehensive account since in the last twenty

*Research supported in part by NSF grant DMS-0906634 and EU MULTIPLEX grant 317532.

years there has been tremendous activity in the area. We intend to present some of the most important results, but our selection is bound to be greatly influenced by personal preferences. For more detailed accounts of the work of Erdős on probability theory and random graphs, see, e.g., [30] and [33].

In Section 7 we shall turn to models carrying more structure than the Erdős–Rényi graphs. These models include the configuration model for the space of graphs with a given degree sequence, some preferential attachment models like the LCD model (the mathematically precise form of the BA model), the BJR model, encompassing a huge array of earlier models, and the analogue of this model with clustering.

Our presentation is self-contained: all we shall assume is that the reader is familiar with the basic concepts of graph theory and probability. The notation we shall use is standard (see, e.g., [31]) although, when quoting from the papers of Erdős and Rényi, we use their somewhat unusual notation. The results of Erdős and Rényi described in the first part of this paper have of course been presented in many places, for example the books [32, 96] and the survey [108].

2. ERDŐS AND RÉNYI: THE BEGINNING

“Let us consider a “random graph” $\Gamma_{n,N}$ having n possible (labelled) vertices and N edges; in other words, let us choose at random (with equal probabilities) one of the $\binom{n}{2}$ possible graphs which can be formed from the n (labelled) vertices P_1, P_2, \dots, P_n by selecting N edges from the $\binom{n}{2}$ possible edges $\overline{P_i P_j}$ ($1 \leq i < j \leq n$).”

With this sentence, the very first sentence of [74], Erdős and Rényi launched the theory of random graphs. They start *in medias res*, as much as that is possible in mathematics, mentioning some earlier related results only later. As a homage to them we shall reproduce the most important results of this paper.

In [74], Erdős and Rényi were interested in the probability that $\Gamma_{n,N}$ is connected and in the structure of a ‘typical’ graph $\Gamma_{n,N}$, when N is in the vicinity of $N_0 = N_0(n)$ such that Γ_{n,N_0} is connected with probability bounded away from both 0 and 1. Before we state their results, we spell out the definition of $\Gamma_{n,N}$. Let $\mathcal{G}_{n,N}$ be the set of all graphs with vertex set $[n] = \{1, \dots, n\}$ and N edges, so that the cardinality of this set is

$$|\mathcal{G}_{n,N}| = \binom{\binom{n}{2}}{N}.$$

The random graph $\Gamma_{n,N}$ is obtained by picking an element of $\mathcal{G}_{n,N}$ uniformly at random. Thus if \mathcal{P} is a property of graphs then the probability that $\Gamma_{n,N}$ has \mathcal{P} is

$$\mathbb{P}_{n,N}(\mathcal{P}) = |\{G : G \in \mathcal{P} \cap \mathcal{G}_{n,N}\}| / \binom{\binom{n}{2}}{N}.$$

The main results of [74] are all based on the following fundamental lemma. Let us say that a graph has *property* \mathcal{A} if it has at most one component with more than one vertex, and let us write $\mathbb{P}_{n,N}(\mathcal{A})$ for the probability that $\Gamma_{n,N}$ has property \mathcal{A} . Furthermore, for a constant c , set

$$N_c = N_c(n) = n(\log n)/2 + cn = (n/2)(\log n + 2c).$$

Strictly speaking, N_c should be defined as $\lfloor n(\log n)/2 + cn \rfloor$ or, more generally, as an integer $n(\log n)/2 + c_n n$, where $c_n \rightarrow c$; our desire to reduce clutter by ignoring the rounding is unlikely to lead to any difficulties.

Lemma 1. *With N_c as above, for every $c \in \mathbb{R}$ we have*

$$(1) \quad \lim_{n \rightarrow \infty} \mathbb{P}_{n,N_c}(\mathcal{A}) = 1.$$

Proof. Let us start by noting that the largest component of Γ_{n,N_c} is not too small – not with large probability, but always. Write $L_1 = L_1(\Gamma_{n,N_c})$ for the maximal order of a component of Γ_{n,N_c} ; thus, if Γ_{n,N_c} has r components, with ℓ_1, \dots, ℓ_r vertices, then $L_1 = \max \ell_i$. Since

$$\sum_{i=1}^r \ell_i = n \quad \text{and} \quad \sum_{i=1}^r \binom{\ell_i}{2} \geq N_c,$$

the convexity of $\binom{x}{2}$ tells us that

$$(n/L_1) \binom{L_1}{2} \geq N_c,$$

so, with $\lambda = \log n + 2c$, we have

$$L_1(\Gamma_{n,N_c}) > \frac{2N_c}{n} = \lambda \sim \log n$$

for every graph $\Gamma_{n,N_c} \in \mathcal{G}_{n,N_c}$.

Given a real number μ , we say that $\Gamma_{n,N}$ has *property* \mathcal{B}_μ if $L_1(\Gamma_{n,N}) \geq n - \mu$; also, we write $\overline{\mathcal{B}}_\mu$ for the negation of this property. The heart of the proof is the following claim.

Claim 1. Let $\mu = \log \log n$. Then

$$\lim_{n \rightarrow \infty} \mathbb{P}_{n, N_c}(\mathcal{B}_\mu) = 1.$$

Proof of Claim 1. Write $\mathcal{G}(\overline{\mathcal{B}}_\mu; n, N_c)$ for the set of graphs in \mathcal{G}_{n, N_c} that do not have property \mathcal{B}_μ , so that our task is to prove that

$$\mathbb{P}_{n, N_c}(\overline{\mathcal{B}}_\mu) = |\mathcal{G}(\overline{\mathcal{B}}_\mu; n, N_c)| / \binom{n}{N_c} = o(1).$$

With $\mathcal{G}(L_1 = s; n, N_c) = \{G \in \mathcal{G}_{n, N_c} : L_1(G) = s\}$, we have

$$\mathcal{G}(\overline{\mathcal{B}}_\mu; n, N_c) = \bigcup_{\lambda \leq s \leq n - \mu} \mathcal{G}(L_1 = s; n, N_c).$$

If S is the vertex set of a component of a graph G then no edge of G joins a vertex in S to a vertex not in S . Hence,

$$|\mathcal{G}(L_1 = s; n, N_c)| \leq \binom{n}{s} \binom{\binom{n}{2} - s(n-s)}{N_c},$$

and so, setting

$$p_s = \binom{n}{s} \binom{\binom{n}{2} - s(n-s)}{N_c} / \binom{n}{N_c}$$

and noting that $p_{n-s} = p_s$, we have

$$(2) \quad \mathbb{P}_{n, N_c}(\overline{\mathcal{B}}_\mu) \leq \sum_{\lambda \leq s \leq n - \mu} p_s \leq 2 \sum_{\mu \leq s \leq n/2} p_s.$$

At this stage, Erdős and Rényi [74] say that ‘by using elementary estimations’

$$p_s \leq \frac{e^{(3-2c)s}}{s!}$$

for $s \leq n/2$. This bound is not so obvious (indeed, it is false for s close to $n/2$), so we shall proceed slowly, cutting the range of s into two parts. In fact, we shall prove a weaker inequality, which is still more than enough to imply Claim 1. Note that

$$q_s = \binom{\binom{n}{2} - s(n-s)}{N_c} / \binom{n}{N_c} \leq \left(1 - \frac{2s(n-s)}{n^2}\right)^{N_c}$$

$$\leq \exp \left(-N_c \left[\frac{2s(n-s)}{n^2} + \frac{4s^2(n-s)^2}{2n^4} \right] \right).$$

First, for $s \leq (\log n)^{-1/2}n$, assuming n is large enough whenever needed, with $t = s/n$ we have

$$\begin{aligned} (3) \quad q_s &\leq \exp \left(-N_c [2t - 2t^2 + 2t^2 - 4t^3 + 2t^4] \right) \\ &\leq \exp \left(-N_c [2t - 4t/(\log n)] \right) \\ &\leq \exp(-s \log n - 2cs + 3s) = n^{-s} e^{(3-2c)s}, \end{aligned}$$

so

$$(4) \quad p_s \leq \frac{e^{(3-2c)s}}{s!}.$$

Second, for $(\log n)^{-1/2}n \leq s \leq n/2$, we have $s(n-s) \geq (n^2/2)(\log n)^{-1/2}$, so

$$q_s \leq \exp \left(-N_c \frac{2s(n-s)}{n^2} \right) \leq \exp \left(-\frac{N_c}{(\log n)^{1/2}} \right) \leq \exp(-n(\log n)^{1/3}),$$

implying

$$(5) \quad p_s \leq e^{-n(\log n)^{1/4}}.$$

Using the bounds (4) and (5) in (2), we find that

$$\mathbb{P}_{n,N_c}(\overline{\mathcal{B}}_\mu) = o(1),$$

which is precisely Claim 1. ■

Returning to the proof of Lemma 1, note that if Γ_{n,N_c} has \mathcal{B}_μ but not \mathcal{A} then it has a component of order s with $2 \leq s \leq \mu = \log n$. Hence all we have to show is that the probability of this is $o(1)$. Now, since a component of Γ_{n,N_c} with $s \geq 2$ vertices has r edges with $1 \leq r \leq \binom{s}{2}$, and the other $N_c - r$ edges of Γ_{n,N_c} join vertices not in the component, the probability that Γ_{n,N_c} has a component of order s with $2 \leq s \leq \mu$ is at most

$$r_s = \binom{n}{s} \sum_{r=1}^{\binom{s}{2}} \binom{\binom{s}{2}}{r} \binom{\binom{n-s}{2}}{N_c - r} / \binom{\binom{n}{2}}{N_c}$$

$$\begin{aligned}
&\leq 2 \binom{n}{s} \binom{\binom{s}{2}}{1} \binom{\binom{n}{2} - s(n-s)}{N_c - 1} / \binom{\binom{n}{2}}{N_c} \\
&\leq \frac{3s^2 N_c}{n^2} \binom{n}{s} \binom{\binom{n}{2} - s(n-s)}{N_c} / \binom{\binom{n}{2}}{N_c} \\
&\leq \frac{2s^2 \log n}{n} \cdot \frac{e^{(3-2c)s}}{s!},
\end{aligned}$$

provided n is large enough, where in the last step we made use of (4). Hence

$$\sum_{r=2}^{\mu} r_s \leq \frac{2 \log n}{n} \sum_{s=2}^{\mu} \frac{s^2 e^{(3-2c)s}}{s!} = O((\log n)/n) = o(1),$$

completing the proof of the lemma. ■

It is interesting to note that in their proof of Lemma 1, Erdős and Rényi start with a lower bound on the order of the largest component – a lower bound that holds *always*, rather than *whp* (with high probability, i.e., with probability tending to 1). In fact, this part of the proof is not needed, since later it is shown that *whp* Γ_{n, N_c} has no nontrivial component of order at most $\mu = \log \log n$. In particular, *whp* Γ_{n, N_c} has a component of order greater than μ , which is all that is needed in the rest (the main part) of the proof.

Also, as we have followed the presentation of Erdős and Rényi, we did not state Lemma 1 in the following slightly more general form, whose proof goes through without *any* changes.

Lemma 1'. *Let $N(n) \geq \frac{n}{2} \log n - (\log \log \log n)n$. Then*

$$\lim_{n \rightarrow \infty} \mathbb{P}_{n, N(n)}(\mathcal{A}) = 1. \quad \blacksquare$$

With Lemma 1, Erdős and Rényi reduced the study of connectedness to the study of the number of isolated vertices, enabling them to deduce several fundamental results about the structure of $\Gamma_{n, N}$ for $N = \frac{n}{2} \log n + O(n)$. First we state these theorems, and then comment on their proofs. We shall continue using the notation $N_c = \frac{n}{2} \log n + cn$, with c constant.

Theorem 2. *The probability that Γ_{n, N_c} is connected tends to $e^{-e^{-2c}}$. ■*

Theorem 3. *Let k be a fixed non-negative integer. The probability that the largest component of Γ_{n,N_c} has order $n - k$ tends to*

$$\exp(-2kc - e^{-2c}) / k!.$$

This is also the limit of the probability that Γ_{n,N_c} has $k + 1$ components.

■

Theorem 4. *Let the edges of a random graph on $[n]$ be obtained as follows. Select the edges successively from among all the edges not yet selected, with all selections equiprobable. Continue this process until the graph formed by the edges selected becomes connected. Let ν_n denote the (random) number of edges of the resulting connected random graph. Then we have*

$$\mathbb{P}\left(\nu_n = \left\lfloor \frac{1}{2}n \log n \right\rfloor + \ell\right) \sim \frac{2}{n} \exp\left(-\frac{2\ell}{n} - e^{-\frac{2\ell}{n}}\right)$$

for $\ell = O(n)$ and, for x constant,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\nu_n - \frac{1}{2}n \log n}{n} < x\right) = e^{-e^{-2x}}. \quad \blacksquare$$

By Lemma 1, most of these results follow if we show that the number of isolated vertices of Γ_{n,N_c} satisfies the assertions, totally ignoring the rest of the graph. In fact, what these results claim is that the distribution of the number X_{n,N_c} of isolated vertices of Γ_{n,N_c} is asymptotically Poisson, with mean $\lambda_c = e^{-2c}$. The ‘natural’ way of proving this is to show that, for every fixed r , the r th factorial moment $\mathbb{E}_r(X_{n,N_c}) = \mathbb{E}(X_{n,N_c}(X_{n,N_c} - 1) \cdots (X_{n,N_c} - r + 1))$ of X_{n,N_c} tends to $\lambda_c^r = e^{-2rc}$. Somewhat surprisingly, this is *not* the route taken by Erdős and Rényi: they make use of Lemma 1 to pin down the distribution of X_{n,N_c} . (In their next paper [75], Erdős and Rényi do use factorial moments, see also [32].) This is how they proceed. By the inclusion–exclusion principle, the number of graphs in \mathcal{G}_{n,N_c} without isolated vertices is

$$\sum_{k=0}^n (-1)^k \binom{n}{k} \binom{\binom{n-k}{2}}{N_c},$$

with partial sums giving alternately upper and lower bounds, so

$$\mathbb{P}(\Gamma_{n,N_c} \text{ has no isolated vertices}) = \sum_{k=0}^n (-1)^k \binom{n}{k} \binom{\binom{n-k}{2}}{N_c} / \binom{\binom{n}{2}}{N_c}.$$

Since, for every fixed k ,

$$\lim_{n \rightarrow \infty} \binom{n}{k} \binom{\binom{n-k}{2}}{N_c} / \binom{\binom{n}{2}}{N_c} = \lambda_c^k / k!,$$

and $\sum_{k=0}^{\infty} (-1)^k \lambda_c^k / k! = e^{-\lambda_c} = e^{-e^{-2c}}$, we find that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\Gamma_{n, N_c} \text{ has no isolated vertices}) = e^{-\lambda_c}.$$

Recalling Lemma 1, this implies Theorem 2.

To prove Theorem 3, Erdős and Rényi note that, by Lemma 1 and Theorem 2,

$$\begin{aligned} & \mathbb{P}(\Gamma_{n, N_c} \text{ has precisely } k \text{ isolated vertices}) \\ & \sim \mathbb{P}(\Gamma_{n, N_c} \text{ has } k+1 \text{ components}) \\ & \sim \mathbb{P}(\Gamma_{n, N_c} \text{ has } k+1 \text{ components, } k \text{ of which are isolated vertices}) \\ & = \mathbb{P}(\Gamma_{n-k, N_c} \text{ is connected}) \binom{n}{k} \binom{\binom{n-k}{2}}{N_c} / \binom{\binom{n}{2}}{N_c} \sim e^{-\lambda_c} \lambda_c^k / k!, \end{aligned}$$

so Theorem 3 holds.

Finally, set $t = \lfloor \frac{1}{2}n \log n \rfloor + \ell$. How does $\nu_n = t$ come about? One way is for the first $t-1$ edges we have selected to define a graph with a single isolated vertex, with the t th edge incident with this vertex. In the light of Lemma 1, it is not surprising that this is the most likely way. (This is effectively asserted by Erdős and Rényi without proof; it is not obvious, but does turn out to be true.) Since $t-1 = N_c$ with $c \sim \ell/n$, and $t = o(n^2)$, we find that $\mathbb{P}(\nu_n = t)$ is asymptotically equal to

$$\mathbb{P}(\Gamma_{n, N_c} \text{ has exactly one isolated vertex}) \frac{n-1}{\binom{n}{2} - t + 1} \sim \frac{2}{n} \lambda_c e^{-\lambda_c},$$

which is the first claim of Theorem 4. The second claim is simply a restatement of Theorem 2.

In Theorem 4, Erdős and Rényi came close to proving a hitting time result: let us see how to make this precise. In what follows, we shall use the notation commonly accepted today; in particular, we shall write G instead of Γ for our graph. Nevertheless, when talking about the original Erdős–Rényi results, we keep their quaint notation $\Gamma_{n, N}$. A *graph process* $\tilde{G}_n = (G_{n, t})_{t=0}^{\binom{n}{2}}$

is a nested sequence of graphs on $[n]$, with $G_{n,t}$ having t edges. Equivalently, it is an enumeration $e_1, e_2, \dots, e_{\binom{n}{2}}$ of the edges of the complete graph on $[n]$, with $\{e_1, \dots, e_t\}$ taken as the edge set of $G_{n,t}$. Let $\tilde{\mathcal{G}}_n$ be the set of all $\binom{n}{2}!$ graph processes on $[n]$, and define a *random graph process* as an element \tilde{G}_n of $\tilde{\mathcal{G}}_n$ chosen uniformly at random. Given a property \mathcal{Q} of graphs on $[n]$, the *hitting time* of \mathcal{Q} is $\tau_{\mathcal{Q}} = \tau_{\mathcal{Q}}(\tilde{G}_n) = \min\{t : G_{n,t} \text{ has } \mathcal{Q}\}$, so that $\tau_{\mathcal{Q}}$ (if always defined, as it will be here) is a random variable whose values belong to $\{0, 1, \dots, \binom{n}{2}\}$. Here is a special case of a result from [53] (see also [32], p. 166) corresponding to Theorem 4.

Theorem 5. *Let $\tau_{\text{conn}}(\tilde{G}_n)$ be the hitting time of connectedness, and $\tau_{\delta \geq 1}(\tilde{G}_n)$ the hitting time of minimal degree at least 1, so that $\tau_{\delta \geq 1}(\tilde{G}_n) \leq \tau_{\text{conn}}(\tilde{G}_n)$ for every graph process \tilde{G}_n . Then equality holds whp.*

Proof. Let us define three properties of a graph process, two of which depend on a constant $c \leq -1$, say. In our notation we shall suppress the dependence on n . First, $\tilde{G}_n = (G_{n,t})$ has property \mathcal{C} if G_{n,N_ℓ} is connected, where $\ell = \log \log \log n$, say. Second, \tilde{G}_n has property \mathcal{A}_c if for some k , $1 \leq k \leq \lambda_c^2 = e^{-4c}$, the graph G_{n,N_c} consists of $k+1$ components, k of which are isolated vertices. Third, write $W = W(\tilde{G}_n)$ for the set of isolated vertices of G_{n,N_c} , and define \mathcal{B}_c to be the property that no edge of G_{n,N_ℓ} joins two vertices of W . Clearly, if $\tilde{G}_n \in \mathcal{A}_c \cap \mathcal{B}_c \cap \mathcal{C}$ then $\tau_{\text{conn}}(\tilde{G}_n) = \tau_{\delta \geq 1}(\tilde{G}_n)$, so to prove our theorem it suffices to show that $\lim_{c \rightarrow -\infty} \lim_{n \rightarrow \infty} \mathbb{P}_n(\mathcal{A}_c \cap \mathcal{B}_c \cap \mathcal{C}) = 1$.

Now, by Theorem 3, $\lim_{n \rightarrow \infty} \mathbb{P}_n(\mathcal{C}) = 1$ and $\lim_{c \rightarrow -\infty} \lim_{n \rightarrow \infty} \mathbb{P}_n(\mathcal{A}_c) = 1$. Also, $\mathbb{P}_n(\mathcal{B}_c \mid \mathcal{A}_c)$ is at least the probability that none of $N_\ell - N_c \leq 2\ell n$ edges chosen uniformly at random from a set of at least $\binom{n}{2}/2$ edges joins some two vertices of a fixed set of at most λ_c^2 vertices. Hence,

$$\mathbb{P}_n(\mathcal{B}_c \mid \mathcal{A}_c) \geq \left(1 - 2 \binom{\lambda_c^2}{2} / \binom{n}{2}\right)^{2\ell n} = 1 + o(1),$$

and so $\mathbb{P}_n(\mathcal{A}_c \cap \mathcal{B}_c) = \mathbb{P}_n(\mathcal{A}_c) + o(1)$, implying

$$\lim_{c \rightarrow -\infty} \lim_{n \rightarrow \infty} \mathbb{P}_n(\mathcal{A}_c \cap \mathcal{B}_c \cap \mathcal{C}) = 1,$$

as required. ■

As remarked in [74], Erdős and Rényi were not the first to study questions related to the probability $P(n, N)$ that $G_{n, N}$ is connected: among others, Riddell and Uhlenbeck [140] and Gilbert [89] (and, simultaneously with [74], Austin, Fagen, Penney and Riordan [11]) had worked earlier on such questions using generating functions. However, the results obtained did not help much to deduce asymptotic results similar to those obtained by Erdős and Rényi. The genius of Erdős and Rényi was precisely that they used the methods of probability theory rather than exact enumeration to prove asymptotic results. Actually, before the publication of [74], Erdős himself had considered $P(n, N)$: with Hassler Whitney he proved (but did not publish) that if $c < 1/2$ then $P(n, N) = o(1)$ for $N \leq cn \log n$, and if $c > 1/2$ then $P(n, N) = 1 + o(1)$ for $N \geq cn \log n$.

3. THE EVOLUTION OF RANDOM GRAPHS

The first Erdős–Rényi paper on random graphs, [74], ended with the promise of better things to come.

“The following more general questions can be asked: Consider the random graph $\Gamma_{n, N(n)}$ with n possible vertices and $N(n)$ edges. What is the distribution of the number of vertices of the greatest connected component of $\Gamma_{n, N(n)}$ and the distribution of the number of its components? What is the typical structure of $\Gamma_{n, N(n)}$ (in the sense in which, according to Lemma 1, the typical structure of Γ_{n, N_c} is that it belongs to type A)? We have solved these problems in the present paper only in the case $N(n) = \frac{1}{2}n \log n + cn$. We shall return to the general case in another paper [75].”

This ‘another paper’ is the most important paper in the theory of random graphs. Entitled “On the evolution of random graphs,” this paper was submitted on December 28, 1959, and was dedicated to Paul Erdős’s great friend, Paul Turán, on his 50th birthday. (A little later, an extended abstract of it [77] was published in Japan.) Given the importance of this paper, it is surprising that it was published (in English) in a Hungarian publication, the journal of the Mathematical Institute headed by Rényi himself. (A regrettable by-product of this was that the paper was not refereed carefully.) It could well be that at the time neither Erdős nor Rényi expected [75] to be a ground-breaking paper, so it is not surprising that, reviewing it in *Mathematical Reviews*, John Riordan also failed to recognize its importance.

In [75], Erdős and Rényi made two very important discoveries. First, that the random graph process $\tilde{G}_n = (G_{n,t})$ undergoes a dramatic change as t passes through $n/2$: there is a *phase transition* at $t = n/2$. Second, that many ‘natural’ monotone increasing properties of graphs (like being connected, containing a matching, having a small diameter, etc) arise *suddenly*: there is a *threshold function*, which is often sharp. We shall say a few words about both discoveries, although in this paper we shall concentrate on the first.

Rather than talking about a ‘phase transition’ (as Stepanov [159] did already in 1970, and most authors do in the last thirty years) Erdős and Rényi talked about the emergence of the ‘*giant component*’. In [77], this is how they summarized their results.

“If n is a fixed large positive integer and N is increasing from 1 to $\binom{n}{2}$, the evolution of $\Gamma_{n,N}$ passes through five clearly distinguishable phases. These phases correspond to ranges of growth of the number N of edges, these ranges being defined in terms of the number of n of vertices.

Phase 1 corresponds to the range $N(n) = o(n)$. For this phase it is characteristic that $\Gamma_{n,N(n)}$ consists almost surely (i.e. with probability tending to 1 for $n \rightarrow +\infty$) exclusively of components which are trees. Trees of order k appear only when $N(n)$ reaches the order of magnitude $n^{(k-2)/(k-1)}$ ($k = 3, 4, \dots$). If $N(n) \sim \rho n^{(k-2)/(k-1)}$ with $\rho > 0$, then the probability distribution of the number of components of $\Gamma_{n,N(n)}$ which are trees of order k tends for $n \rightarrow +\infty$ to the Poisson distribution with mean value $\lambda = \frac{(2\rho)^{k-1} k^{k-2}}{k!}$. If $N(n) n^{-(k-2)/(k-1)} \rightarrow +\infty$ then the distribution of the number of components which are trees of order k is approximately normal with mean $M_n = n \frac{k^{k-2}}{k!} \left(\frac{2N(n)}{n} \right)^{k-1} e^{-\frac{2kN(n)}{n}}$ and with variance also equal to M_n . This result holds also in the next two ranges, in fact it holds under the single condition that $M_n \rightarrow +\infty$ for $n \rightarrow +\infty$.

Phase 2 corresponds to the range $N(n) \sim cn$ with $0 < c < 1/2$. In this case $\Gamma_{n,N(n)}$ already contains cycles of any fixed order with probability tending to a positive limit: the distribution of the number of cycles of order k in $\Gamma_{n,N(n)}$ is approximately a Poisson-distribution with mean value $\frac{(2c)^k}{2k}$. In this range almost surely all components of $\Gamma_{n,N(n)}$ are either trees or components consisting of an equal number of edges and vertices, i.e. components containing exactly one cycle. ... In this phase though not all, but still almost all (i.e. $n - o(n)$) vertices belong to com-

ponents which are trees. The mean number of components is $n - N(n) + O(1)$, i.e. in this range by adding a new edge the number of components decreases by 1, except for a finite number of steps.

Phase 3 corresponds to the range $N(n) \sim cn$ with $c \geq 1/2$. When $N(n)$ passes the threshold $n/2$, the structure of $\Gamma_{n,N(n)}$ changes abruptly. As a matter of fact this sudden change of the structure of $\Gamma_{n,N(n)}$ is the most surprising fact discovered by the investigation of the evolution of random graphs. While for $N(n) \sim cn$ with $c < 1/2$ the greatest component of $\Gamma_{n,N(n)}$ is a tree and has (with probability tending to 1 for $n \rightarrow +\infty$) approximately $\frac{1}{\alpha} \left(\log n - \frac{5}{2} \log \log n \right)$ vertices, where $\alpha = 2c - 1 - \log 2c$, for $N(n) \sim n/2$ the greatest component has (with probability tending to 1 for $n \rightarrow +\infty$) approximately $n^{2/3}$ vertices and has a rather complex structure. Moreover for $N(n) \sim cn$ with $c > 1/2$ the greatest component of $\Gamma_{n,N(n)}$ has (with probability tending to 1 for $n \rightarrow +\infty$) approximately $G(c)n$ vertices, where

$$(6) \quad G(c) = 1 - \frac{1}{2c} \sum_{k=1}^{+\infty} \frac{k^{k-1}}{k!} (2ce^{-2c})^k$$

(clearly $G(1/2) = 0$ and $\lim_{c \rightarrow +\infty} G(c) = 1$).

Except this “giant” component, the other components are all relatively small, most of them being trees, the total number of vertices belonging to components, which are trees being almost surely $n(1 - G(c)) + o(n)$ for $c \geq 1/2 \dots$

The evolution of $\Gamma_{n,N(n)}$ in Phase 3 may be characterized by that the small components (most of which are trees) melt, each after another, into the giant component, the smaller components having the larger chance of “survival”; the survival time of a tree of order k which is present in $\Gamma_{n,N(n)}$ with $N(n) \sim cn$, $c > 1/2$ is approximately exponentially distributed with mean value $n/2k$.

Phase 4 corresponds to the range $N(n) \sim cn \log n$ with $c \leq 1/2$. In this phase the graph almost surely becomes connected. If

$$N(n) = \frac{n}{2k} \log n + \frac{k-1}{2k} n \log \log n + yn + o(n)$$

then there are with probability tending to 1 for $n \rightarrow +\infty$ only trees of order $\leq k$ outside the giant component, the distribution

of the number of trees of order k having in the limit again a Poisson distribution with mean value $\frac{e^{-2ky}}{k \cdot k!} \dots$

Phase 5 consists of the range $N(n) \sim (n \log n)\omega(n)$ where $\omega(n) \rightarrow +\infty$. In this range the whole graph is not only almost surely connected, but the orders [degrees] of all points [vertices] are almost surely asymptotically equal. Thus the graph becomes in this phase “asymptotically regular”.

Let us clear up one of the mysteries bound to puzzle a reader unfamiliar with these results: the form of the function $G(c)$ in (6). The explanation is that $1 - G(c)$ is about the proportion of vertices on tree components. Indeed, for $N(n) \sim cn$, the expected number of vertices of $\Gamma_{n,N(n)}$ on tree components is

$$\sum_{k=1}^n \binom{n}{k} k^{k-2} \binom{n-k}{N-k+1} / \binom{n}{N} \sim \frac{1}{2c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (2ce^{-2c})^k.$$

In the original ‘evolution’ paper [75], Erdős and Rényi gave an even more succinct description of their main results.

“Thus the situation can be summarized as follows: the largest component of $\Gamma_{n,N(n)}$ is of order $\log n$ for $\frac{N(n)}{n} \sim c < 1/2$, of order $n^{2/3}$ for $\frac{N(n)}{n} \sim \frac{1}{2}$ and of order n for $\frac{N(n)}{n} \sim c > 1/2$. This double “jump” of the size of the largest component when $\frac{N(n)}{n}$ passes through the value $1/2$ is one of the most striking facts concerning random graphs.”

We shall postpone our comments about later developments concerning the phase transition to later sections.

Turning to threshold functions, the quintessential example of a sharp threshold function is precisely $\frac{n}{2} \log n$, the threshold function for connect- edness. The other main example is the threshold for the appearance of a suitable ‘small’ subgraph. The average degree of a graph H with k ver- tices and ℓ edges is $a(H) = 2\ell/k$. Erdős and Rényi called H balanced if $a(H') \leq a(H)$ for every subgraph H' of H . Clearly, trees, cycles, unicyclic graphs, complete graphs and complete r -partite graphs are all balanced. As shown by Erdős and Rényi, balanced graphs appear rather suddenly in the random graph process.

Theorem 6. *Let H be a balanced graph of average degree $a = a(H) > 0$. Then $n^{2-2/a}$ is a threshold function for the property of containing (a subgraph isomorphic to) H in the sense that if $\omega(n) \rightarrow \infty$ then for*

$N(n) \leq n^{2-2/a}/\omega(n)$ the probability that $\Gamma_{n,N(n)}$ contains H tends to 0 as $n \rightarrow \infty$, and for $N(n) \geq \omega(n)n^{2-2/a}$ it tends to 1. ■

Over twenty years later, much sharper results were proved by Bollobás [27], Karoński and Ruciński [106, 107], Bollobás and Thomason [52], Ruciński and Vince [151, 152, 153, 154], and others. Also, Barbour [16] was the first to apply the Stein–Chen method for Poisson approximation to the number of complete subgraphs of a random graph; further results in that vein were proved by Barbour, Karoński and Ruciński [18], Arratia and Lander [10], Janson, Łuczak and Ruciński [95], Barbour, Janson, Karoński and Ruciński [17], and others.

A *monotone increasing property of graphs* is a collection \mathcal{Q} of graphs such that if a graph H is in \mathcal{Q} , every graph obtained from H by the addition of edges is also in \mathcal{Q} . Bollobás and Thomason [53] observed that the Kruskal–Katona theorem [113, 110] implies that *every* monotone increasing property has a threshold function. Much more importantly, Friedgut and Kalai [86] and Friedgut and Bourgain [85] proved several deep theorems about the sharp thresholds of *symmetric* monotone increasing properties, in particular, graph properties invariant under graph isomorphism. The starting point of these results is the fundamental KKL Theorem of Kahn, Kalai and Linial [101], and its extension by Bourgain, Kahn, Kalai, Katznelson and Linial [56].

To conclude this section, let us quote a prophetic passage from the main ‘evolution’ paper [75] of Erdős and Rényi, expressing the hope that the results will be extended to more complicated structures.

“We succeeded in revealing the emergence of certain structural properties of $\Gamma_{n,N}$. However a great deal remains to be done in this field. We shall call the attention of the reader to certain unsolved problems. It seems to us further that it would be worth while to consider besides graphs also more complex structures from the same point of view, i.e. to investigate the laws governing their evolution in a similar spirit. This may be interesting not only from a purely mathematical point of view. In fact, the evolution of graphs may be considered as a rather simplified model of the evolution of certain communication nets (railway, road or electric network systems, etc.) of a country or some other unit. (Of course, if one aims at describing such a real situation, one should replace the hypothesis of equiprobability of all connections by some more realistic hypothesis.) It seems plausible that by considering the random growth of more complicated structures (e.g. structures consisting of different sorts

of “points” and connections of different types) one could obtain fairly reasonable models of more complex real growth processes (e.g. the growth of a complex communication net consisting of different types of connections, and even of organic structures of living matter, etc.).”

As if in reply to the thoughts of Erdős and Rényi, for well over ten years now, numerous large-scale real-world networks have been modelled by a variety of spaces of random graphs (see, e.g., Watts and Strogatz [166], Barabási and Albert [14], Bollobás and Riordan [43, 44], Bollobás, Janson and Riordan [38, 39, 41]); these models are often very difficult to analyse. We return to this topic in Section 7.

The ideas of Erdős and Rényi about phase transitions continue to influence mathematical research: many of the papers published today (see, e.g., [3, 72, 73, 66]) in a variety of fields owe much to their pioneering work.

4. FURTHER RESULTS

Erdős and Rényi returned to connectivity questions in [76], proving, among other results, the following extension of Theorem 2.

Theorem 7. *Let k be a natural number, α a real, and*

$$N_{k,\alpha}(n) = \frac{n}{2} (\log n + (k-1) \log \log n + \alpha + o(1)).$$

Then whp the connectivity of $\Gamma_{n,N_{k,\alpha}}$ is either $k-1$ or k . Furthermore, the probability that the connectivity is k tends to $\exp(-e^{-\alpha}/(k-1)!)$. This is also the limit of the probability that the minimal degree is k . ■

Once again, Erdős and Rényi were very close to proving the following sharp hitting time result telling us that the primary obstruction of k -connectedness is the existence of a vertex of degree at most $k-1$.

Theorem 8. *Let $\tau_{k\text{-conn}}(\tilde{G}_n)$ be the hitting time of being k -connected and $\tau_{\delta \geq k}(\tilde{G}_n)$ that of having minimal degree at least k , so that $\tau_{\delta \geq k}(\tilde{G}_n) \leq \tau_{k\text{-conn}}(\tilde{G}_n)$ for every graph process \tilde{G}_n . Then equality holds whp. ■*

Call a graph *symmetric* if it has a non-trivial automorphism. Furthermore, for a graph G , set

$$A(G) = \min\{|E(G) \Delta E(G')| : G' \text{ is symmetric}\},$$

and

$$A(n) = \max\{A(G) : |G| = n\}.$$

It is easily checked that $A(2) = \dots = A(5) = 0$ and $A(6) = 1$; also, it is not hard to show that $A(G) \leq (n-1)/2$ for $|G| = n$. The main result of [78] is that for $N = \binom{n}{2}/2$

$$A(\Gamma_{n,N}) = \frac{n}{2} + o(n)$$

holds whp, so $\lim_{n \rightarrow \infty} A(n)/n = 1/2$.

The later papers [79, 80, 81] of the Erdős–Rényi series on random graphs concern k -factors of graphs and bipartite graphs. In particular, in [80] they prove that if n is even and $N(n) = \frac{n}{2} \log n + \omega(n)n$ then whp $\Gamma_{n,N(n)}$ has a 1-factor. The hitting time sharpening of this result also holds: $\tau_{1\text{-fact}}(\tilde{G}_n) = \tau_{\delta \geq 1}(\tilde{G}_n)$ holds whp, where $\tau_{1\text{-fact}}(\tilde{G}_n)$ is the hitting time of containing a 1-factor.

Not surprisingly, the results above have inspired much further research: here we shall hardly scrape the surface of the body of these results. Starting with the last result above: we know that the main obstruction to a complete matching (in a graph with an even number of vertices) is an isolated vertex. What happens if we *condition* on having no isolated vertices? Will fewer edges make the existence of a 1-factor likely? This question was answered by Bollobás and Frieze [37]: they proved that if $N(n) = \frac{1}{4}n \log n + \frac{1}{2}n \log \log n + c_n n$, then the probability that $\Gamma_{n,N(n)}$, conditioned on having minimal degree at least 1, has a complete matching tends to $\exp(-\frac{1}{8} \exp(-4c))$ if $c_n \rightarrow c$ as $n \rightarrow \infty$. Also, in a random n -by- n bipartite graph the primary obstruction to a complete matching is an isolated vertex, and the secondary is a pair of vertices of degree 1 joined to the same vertex.

Also, more recently, Frieze and Pittel [88], and Frieze [87] investigated what happens if we condition on having minimal degree at least 2, rather than 1. In this case the number of edges needed is drastically reduced: for $c \geq 2$, the probability that a random n -by- n bipartite graph with cn edges conditioned on having minimal degree at least 2 has a complete matching tends to 1 as $n \rightarrow \infty$.

Turning to asymmetric graphs, Babai, Erdős and Selkow [12] gave a naive algorithm to test all but $o(2^{\binom{n}{2}})$ graphs on $[n]$ for isomorphism against any other graph. The idea is very simple: in all but $o(2^{\binom{n}{2}})$ graphs the highest $n^{0.15}$ degrees are distinct (in fact, they are at least $n^{0.03}$ apart), and can be used to ‘anchor’ all other vertices.

Graphs with two isolated vertices are symmetric, and so are graphs with a vertex having two neighbours of degree 1. If we rule these out, our graphs are much less likely to be symmetric. Thus, answering a question of Wormald, it was proved by Kim, Sudakov and Vu [111] that for any $3 \leq d = d(n) \leq n - 4$ a random d -regular graph on $[n]$ is asymmetric whp.

5. PHASE TRANSITIONS – THE RESTART

The influence of the ‘evolution’ paper [75] on mathematics has a rather strange history. For the next two decades after its appearance, essentially no papers made any use of its main result, although the mantra of the ‘double jump’ was repeated many times. This lack of continuation was due partly to the fact that the evolution paper was ahead of its time, but also to the unfortunate misstatement of the main result: if $N(n) \sim cn$ then for $c < 1/2$ the maximal component has order $O(\log n)$, for $c > 1/2$ its order is linear in n , and for $c = 1/2$ its order is $\Theta(n^{2/3})$, with all assertions holding whp. There is nothing wrong with the first two statements, but the assertion concerning $N(n) \sim n/2$ is far from correct. If all statements had been correct, then this would have been the end of the story: essentially nothing else could have been said about the maximum of the orders of the components.

In the first subsection below, we describe the true nature of the phase transition; then we turn to planarity and another result from the evolution paper that had to be put right. The third and last subsection is about the core, which does appear suddenly, with a big jump.

5.1. No Double Jump, But a Smooth Transition

In 1984, Bollobás [29] noticed that there is no double jump; in fact, if $N(n) - n/2$ grows a little faster than $n^{2/3}$ then whp $G_{n,N(n)}$ contains a ‘giant’ component, with all other components at most half as large. Let us state this result in a somewhat simpler form. As usual, we shall write $L_1(G) \geq L_2(G) \geq \dots$ for the orders of the components of a graph G .

Theorem 9. *Whp $\tilde{G}_n = (G_{n,t})$ is such that if $t = n/2 + s$ with $(\log n)n^{2/3} \leq s = o(n)$ then*

$$L_1(G_{n,t}) \sim 4s \quad \text{and} \quad L_2(G_{n,t}) \leq (\log n)n^2/s^2. \quad \blacksquare$$

One can show that if s above is $O(n^{2/3})$ then for no function $f(n, s)$ does $L_1(G_{n,t}) \sim f(n, s)$ hold whp. Also, if $s = o(n^{2/3})$ then for $t_0 = n/2$

and $t_1 = n/2 + s$ the distributions of the random variables $L_1(G_{n,t_0})$ and $L_1(G_{n,t_1})$ are very close.

Needless to say, Theorem 9 rules out the existence of a double jump since knowing that $N(n) \sim n/2$ does not come close to telling us how large $L_1(G_{n,N})$ is likely to be. For example, if $N = n/2 + n^{3/4} \sim n/2$ then whp we have $L_1(G_{n,N}) \sim 4n^{3/4}$ and $L_2(G_{n,N}) \leq (\log n)n^{1/2}$, but if $N = n/2 + n/\log n$ then whp $L_1(G_{n,N}) \sim 4n/\log n$ and $L_2(G_{n,N}) \leq (\log n)^3$.

Although the proof of Theorem 9 made use of the graph process \tilde{G}_n by looking at the graphs $G_{n,t}$ at various times (values of t), the individual graphs $G_{n,t}$ were considered to be static random objects, as in [75]. In particular, the $\log n$ factor in the lower bound on s was for ease of calculations. In 1990, Łuczak [118] used better estimates and more delicate arguments to replace this $\log n$ factor by any function tending (crawling?) to infinity. Again we state a simplified form of the result.

Theorem 10. *Let $\omega(n) \rightarrow \infty$ and let $s = s(n)$ satisfy $\omega(n)n^{2/3} \leq s = o(n)$. Then for $N = n/2 - s$, whp*

$$L_i(G_{n,N}) \sim \frac{n^2}{2s^2} \log(s^3/n^2)$$

for any fixed i , while for $N = n/2 + s$, whp $L_1(G_{n,N}) \sim 4s$ and $L_2(G_{n,N}) \leq (\log n)n^2/s^2$. ■

The results of Bollobás and Łuczak show that the evolution of the random graph process near $N = n/2$ can be thought of in three phases: the *subcritical* phase where $N \leq n/2 - \omega(n)n^{2/3}$ in which there are many ‘largest’ components of almost equal size, which are whp trees, the *critical* phase $N = n/2 + O(n^{2/3})$ in which the largest few component sizes have, after rescaling, a non-trivial distribution (see Section 6.1), and the *supercritical* phase $N \geq n/2 + \omega(n)n^{2/3}$ in which the ‘giant’ component has ‘emerged’ as the largest component, has an asymptotically determined size, and will not be overtaken again. More details of this picture were established in 1994 in a monumental paper by Janson, Knuth, Łuczak and Pittel [93].

In later sections we shall present much sharper results that have been proved by a variety of different methods.

5.2. Planarity

Let us note two other interesting corrections to [75], this time due to Łuczak and Wierman [124] and Łuczak, Pittel and Wierman [123].

Write $X_{n,N}(d)$ for the number of cycles in $G_{n,N}$ with precisely d diagonals. A theorem in [75] states that for $N = n/2 + (c + o(1))\sqrt{n}$, with c constant, the random variable $X_{n,N}(d)$ has asymptotically Poisson distribution with a certain mean $\lambda_c(d)$, $0 < \lambda_c(d) < \infty$. Also, $\lambda_c(d) \rightarrow 0$ as $c \rightarrow -\infty$ and $\lambda_c(d) \rightarrow \infty$ as $c \rightarrow \infty$. From this it is essentially immediate that $\liminf_{n \rightarrow \infty} \mathbb{P}(G_{n,n/2} \text{ is non-planar}) > 0$. Also, if $N(n) = n/2 + \omega(n)\sqrt{n}$ with $\omega(n) \rightarrow \infty$ then whp $G_{n,N(n)}$ is non-planar.

Although the arguments showing these assertions looked convincing, the truth is rather different. In particular, Łuczak and Wierman [124] proved the following.

Theorem 11. *Whp $G_{n,n/2}$ contains no cycle with a diagonal, and so has chromatic number 3. ■*

The range of the transition from planarity to non-planarity was pinned down by Łuczak, Pittel and Wierman [123].

Theorem 12. *Set $N(n) = n/2 + \omega(n)n^{2/3}$. Then $G_{n,N(n)}$ is non-planar whp iff $\omega(n) \rightarrow \infty$. ■*

These results needed considerably more complicated arguments than Erdős and Rényi envisaged.

It is worth emphasizing that the various blemishes of the ‘evolution’ paper in no way diminish its importance. The fact that these imperfections went unnoticed for decades proves that the paper was way ahead of its time.

5.3. The Core – A Genuine Jump

Let us make some remarks on another kind of phase transition, the sudden emergence of the core. The concept of the core (really, k -core) of a graph was introduced by Bollobás [28] in 1984. For $k \geq 1$ the k -core of a graph is the unique maximal subgraph of minimal degree at least k , i.e., the union of all subgraphs of minimal degree k , provided there is such a subgraph. The conditions for the existence of the 1-core and 2-core are trivial: the 1-core exists if there are any edges, and the 2-core exists if there are any cycles. The fun starts at $k = 3$, so for the rest of this section we shall assume that $k \geq 3$.

The number of vertices in the k -core is a trivial upper bound on the number of vertices in a k -connected subgraph; also, if there is no k -core then the graph is certainly k -colourable.

Looking at it crudely, the emergence of the k -core is rather similar to the emergence of the giant component: for every $k \geq 3$ there is a constant $c_k > 0$ such that if $N(n) \sim cn/2$ for a constant c then

$$\lim_{n \rightarrow \infty} \mathbb{P}(G_{n,N(n)} \text{ has a } k\text{-core}) = \begin{cases} 0 & \text{if } c < c_k, \\ 1 & \text{if } c > c_k. \end{cases}$$

Concerning these constants, Bollobás and Thomason (see [62]) proved the simple inequality

$$c_k - \log(c_k + 1) \geq c_{k-1}.$$

Chvátal [62] used elaborate arguments to prove that $c_3 > 2.88$ and so whp $G_{n,1.44n}$ is 3-chromatic; combining this with the inequality above, he obtained $c_4 > 4.61$, $c_5 > 6.64$, etc. In 1991, Łuczak [119] determined the asymptotic value of c_k as $k \rightarrow \infty$, when he proved that for any constant $\gamma > 1/2$ we have $c_k = k + O(k^\gamma)$.

The existence of the constants c_k is far from surprising; what is fascinating about the k -core is that, as shown by Łuczak [119], it takes off with a bang.

Theorem 13. *Let $k \geq 3$ be fixed. For any $N(n) = \Theta(n)$, whp the k -core of $G_{n,N(n)}$ either contains no vertices, or at least $0.0002n$ vertices. ■*

Write $\tau_{k\text{-core}}(\tilde{G}_n)$ for the hitting time of the existence of a non-empty k -core in the random graph process \tilde{G}_n , and $s_{k\text{-core}}(\tilde{G}_n)$ for the number of vertices in the k -core at this hitting time. Since the size of the k -core can only increase as edges are added, Łuczak's result tells us that whp $\tilde{G}_n = (G_{n,t})_0^{\binom{n}{2}}$ is such that for $t = \tau_{k\text{-core}}(\tilde{G}_n)$ the graph $G_{n,t-1}$ does not have a k -core, but $G_{n,t}$ not only has a k -core, but its k -core is huge, of linear size. In fact, for $t = \tau_{k\text{-core}}(\tilde{G}_n)$ the jump due to the t th edge is whp from 0 to $(1 + o(1))s_k n$, for some constant $s_k > 0$. In 1996, Pittel, Spencer and Wormald [137] not only proved the existence of these constants c_k and s_k , but in a stunning theorem even determined them explicitly. As usual, we write $\text{Po}(\lambda)$ for a Poisson random variable with mean λ , so that $\mathbb{P}(\text{Po}(\lambda) = k) = e^{-\lambda} \lambda^k / k!$.

Theorem 14. *Let $k \geq 3$. For $\lambda > 0$, set $\pi_k(\lambda) = \mathbb{P}(\text{Po}(\lambda) \geq k - 1)$ and $c_k = \min_{\lambda > 0} \lambda / \pi_k(\lambda)$, and let s_k be the unique value of λ at which this minimum c_k is attained. Then c_k and s_k are as described above. Furthermore, $c_k = k + \sqrt{k \log k} + O(\log k)$. ■*

Numerical calculations show that $c_3 \approx 3.35$, $s_3 \approx 0.27$, $c_4 \approx 5.14$, $s_4 \approx 0.43$, $c_5 \approx 6.81$ and $s_5 \approx 0.55$.

The proof of Theorem 14 given by Pittel, Spencer and Wormald is a *tour de force*. There is an obvious algorithm for finding the k -core (if there is one): simply delete vertices of degree less than k one-by-one. In [137] this algorithm is analysed by an ingenious and difficult argument to prove Theorem 14. This approach left something of a mystery, however; as noted in [137] there is a natural branching process heuristic leading to the values they found for c_k and s_k , so can one reprove the result using branching processes (i.e., analysing the k -core ‘from the inside’ rather than by deletion)? Such a proof of a weaker (but more general) version of Theorem 14 was found much later by Riordan [142].

Although one of the initial uses of the core was to bound the chromatic number, it was never expected to give the exact threshold for k -colourability. In 1991, Łuczak [120] proved the beautiful theorem that, for every constant $c > 0$, whp the chromatic number of $G_{n, \lfloor cn \rfloor}$ is one of two consecutive numbers. (In fact, Łuczak proved considerably more; later his results were extended by Alon and Krivelevich [9] to the statement that if $\varepsilon > 0$ is fixed and $N(n) < n^{3/2-\varepsilon}$ then whp the chromatic number of $G_{n, N(n)}$ is one of two consecutive numbers.) As to what these two consecutive numbers are, this remained a mystery until Achlioptas and Naor [2] proved the following admirable result.

Theorem 15. *Given $c > 0$, let k be defined by $(k-1)\log(k-1) \leq c < k \log k$. Then whp the chromatic number of $G_{n, \lfloor cn \rfloor}$ is k or $k+1$. Also, if $(k-1/2)\log k \leq c < k \log k$ then whp the chromatic number of $G_{n, \lfloor cn \rfloor}$ is $k+1$. ■*

To conclude this section, let us note some results concerning subgraphs similar to cores: k -regular subgraphs. Alon, Friedland and Kalai [8] proved that every graph of maximum degree $2k-1$ and average degree greater than $2k-2$ has a k -regular subgraph. In 2006, this result was used by Bollobás, Kim and Verstraëte [42] to prove that if $c > 2k$ then whp $G_{n, \lfloor cn \rfloor}$ contains a k -regular subgraph. Prałat, Verstraëte and Wormald [139] proved that for k large, the threshold for the appearance of a k -regular subgraph of $G_{n, N(n)}$ is at most the threshold for the appearance of the $(k+2)$ -core, which is given in Theorem 14. Even more recently, Chan and Molloy [60] proved an essentially best possible result of this type, replacing $k+2$ by $k+1$.

Taking the two questions above together (the appearance of the k -core and the appearance of a k -regular subgraph) one is lead to the following question: when the k -core appears, does it contain an ℓ -regular spanning subgraph for ℓ close to k ? Even more, when the k -core appears, does

it contain ‘many’ edge-disjoint Hamilton cycles? In a paper to appear, Krivelevich, Lubetzky and Sudakov [112] proved the following beautiful result.

Theorem 16. *Let $k \geq 15$. Then whp \tilde{G}_n is such that for $t_k = \tau_{k\text{-core}}(\tilde{G}_n)$ the k -core of the graph $G_{n,t}$ is Hamiltonian for every $t \geq t_k$. Furthermore, for large k , the k -core of $G_{n,t}$ contains $\lfloor (k-3)/2 \rfloor$ edge-disjoint Hamilton cycles for every $t \geq t_k$. ■*

This is a great theorem indeed: to prove it, Krivelevich, Lubetzky and Sudakov condition the random graph process \tilde{G}_n on its future – a most unusual procedure. The result is almost best possible; at best it may be true that $\lfloor (k-3)/2 \rfloor$ and k large can be replaced by $\lfloor (k-1)/2 \rfloor$ and $k \geq 4$.

6. RECENT RESULTS ABOUT THE PHASE TRANSITION

After a slow start, the study of the phase transition initiated by Erdős and Rényi has blossomed into a major field, with many hundreds of papers written on this and related topics. Rather than attempt to survey this vast body of work, we shall focus on a small number of threads within it. Even within these there will be space to mention only a few of the results. The selection of topics and results naturally follows our own particular interests (nowadays no-one can be familiar with all of this work), and is not intended to be definitive in any way.

Broadly speaking, work continuing that in [75] falls into two main types: further, more detailed study of the phase transition in $G_{n,N}$ itself, and generalizations to other models. We describe some results of the first type in the rest of this section, turning to other models in the next.

First, let us comment on a minor technical point. Nowadays, in many contexts, instead of the Erdős–Rényi model $G_{n,N}$ (also called the ‘size model’) it is more common to study the ‘binomial’ model $G_{n,p}$, a random graph on $[n] = \{1, \dots, n\}$ in which each possible edge is present independently with probability p . This model was introduced by Gilbert [90] at around the same time that Erdős and Rényi introduced their model but (paradoxically, given the definitions of the models) he studied it in a much less probabilistic way. For many purposes the models are essentially equivalent if the parameters are chosen suitably, e.g., with $p = N/\binom{n}{2}$, but $G_{n,p}$ is often easier to work with. In the following sections we describe all results for $G_{n,p}$ rather than $G_{n,N}$. (Formally there is a clash in the notation here, but since $0 < p < 1$ while $N \geq 1$ there is no danger of confusion.)

6.1. Ever more precise results about $G_{n,p}$

Concerning the order L_1 of the largest component in $G_{n,p}$, the results of Bollobás [29] and Łuczak [118] are in one sense the last word. If, following Erdős and Rényi, we consider the most important features of $G_{n,p}$ to be its ‘typical’ properties, i.e., the properties that hold with probability tending to 1, then for any $p = p(n) = \Theta(1/n)$ these results give the complete answer as far as bounds on L_1 are concerned – they give necessary and sufficient conditions on a deterministic function $f(n)$ for $L_1(G_{n,p}) \leq f(n)$ or $L_1(G_{n,p}) \geq f(n)$ to hold whp. However, it is natural to go further: can we find the limiting value of $\mathbb{P}(L_1(G_{n,p}) \leq f(n))$ when this is bounded away from 0 and 1? Equivalently, can we describe the limiting (appropriately rescaled) distribution of the deviation of the size of the giant component from its typical value? Results of this type are known as ‘limit theorems’. Going even further, can we find the asymptotic value of the probability that $L_1(G_{n,p})$ is *exactly* equal to $f(n)$, for $f(n)$ within the ‘typical’ range? Such results are known as ‘local limit theorems’.

Throughout this section we discuss $G_{n,p}$ with $p = O(1/n)$. We usually write $p = \lambda/n$ where $\lambda = \lambda(n) = O(1)$. In the light of the results described in Section 5.1, we refer to the cases $(\lambda - 1)^3 n \rightarrow -\infty$, $(\lambda - 1)^3 n = O(1)$ and $(\lambda - 1)^3 n \rightarrow \infty$ as the *subcritical*, *critical* and *supercritical* regimes. Sometimes, we write $\varepsilon = \varepsilon(n)$ for $\lambda - 1$. We use standard notation for probabilistic asymptotics, as in [96], for example; in particular, $o_p(f(n))$ denotes a random quantity that, after division by the deterministic function $f(n)$, converges to 0 in probability. Thus $X_n = g(n) + o_p(f(n))$ means exactly that for any fixed $\varepsilon > 0$, whp $|X_n - g(n)| \leq \varepsilon f(n)$. We write $O_p(f(n))$ for a quantity that, when divided by $f(n)$, is bounded in probability. Thus $X_n = O_p(f(n))$ means that for any $\varepsilon > 0$ there is a C such that for all (large enough) n we have $\mathbb{P}(|X_n| \leq C f(n)) \geq 1 - \varepsilon$.

In the subcritical case (where the giant component is almost always a tree), Łuczak [118] proved a limit theorem showing that, appropriately rescaled, L_1 has the extreme value distribution associated to a Poisson process with exponential density.

Theorem 17. *Let $\lambda = 1 - \varepsilon$, where $\varepsilon = \varepsilon(n)$ satisfies $\varepsilon^3 n \rightarrow \infty$ and $\varepsilon = o(1/\log n)$. Then*

$$\begin{aligned} \mathbb{P} \left[L_1(G_{n,\lambda/n}) < 2\varepsilon^{-2} \left(\log(\varepsilon^3 n) - \frac{5}{2} \log \log(\varepsilon^3 n) + x \right) \right] \\ \rightarrow \exp \left(-\frac{1}{4\sqrt{\pi}} e^{-x} \right) \end{aligned}$$

as $n \rightarrow \infty$. ■

(In fact, in [118] this is stated with $\varepsilon = o(1)$ in the place of $\varepsilon = o(1/\log n)$, but under this more general assumption the formula is not quite correct; see [45].)

The critical case is much more complicated and more interesting. Define a stochastic process $W^\alpha(s)$ (a random function on $[0, \infty)$) by

$$W^\alpha(s) = W(s) + \alpha s - s^2/2,$$

where $W(s)$ is a standard Brownian motion. An *excursion* of this process is a maximal interval on which W^α exceeds its previous minimum value; let $(|\gamma_i|)_{i \geq 1}$ denote the lengths of the excursions sorted into decreasing order. Using random walk arguments based in part on ideas of Martin-Löf [125] and Karp [109], Aldous [7] proved the following result.

Theorem 18. *Let $p = \lambda/n$ where $\lambda = \lambda(n)$ satisfies*

$$(\lambda - 1)^3 n \rightarrow \alpha^3$$

for some $\alpha \in \mathbb{R}$. Then, for any fixed r , writing L_r for the number of vertices in the r th largest component of $G_{n,p}$, the sequence $(n^{-2/3} L_i)_{i=1}^r$ converges in distribution to $(|\gamma_i|)_{i=1}^r$. ■

In fact, Aldous proved more: convergence of the entire sequence of rescaled component sizes in l^2 , and convergence of the distribution of the nullities (number of ‘extra’ edges compared to a tree of the same order) of the components to an appropriate random process.

Turning to the supercritical case, for $\lambda > 1$ constant Stepanov [159] proved a limit theorem already in 1970; this was reproved by Pittel [135] by careful arguments based on tree counting.

Theorem 19. *Let $\lambda > 1$ be constant, let ρ_λ be the positive solution to $\rho_\lambda = 1 - e^{-\lambda \rho_\lambda}$, let $\lambda_* < 1$ satisfy $\lambda_* e^{-\lambda_*} = \lambda e^{-\lambda}$, and set*

$$\sigma_\lambda^2 = \frac{\rho_\lambda(1 - \rho_\lambda)}{(1 - \lambda_*)^2}.$$

If L_1 is the maximal order of a component of $G_{n,\lambda/n}$, then

$$\frac{L_1 - \rho_\lambda n}{\sigma_\lambda \sqrt{n}} \xrightarrow{d} \text{No}(0, 1),$$

where \xrightarrow{d} denotes convergence in distribution, and $\text{No}(0, 1)$ is the standard normal distribution. ■

The quantity ρ_λ defined in this result is (as it must be) none other than $G(\lambda)$ as defined in (6). Pittel and Wormald [138] extended Theorem 19 to the much more delicate case where the average degree tends down to 1.

Theorem 20. *Let $\varepsilon = \varepsilon(n)$ satisfy $\varepsilon \rightarrow 0$ but $\varepsilon^3 n \rightarrow \infty$. Then the order L_1 of the largest component of $G_{n,(1+\varepsilon)/n}$ satisfies*

$$\frac{L_1 - \rho_{1+\varepsilon} n}{\varepsilon^{-1/2} \sqrt{2n}} \xrightarrow{d} \text{No}(0, 1). \quad \blacksquare$$

In fact, they studied not only the order of the giant component, but also the number of edges, and the size of its 2-core (see Section 5.3). They came very close to, but did not quite, prove a local limit theorem. The local limit theorem for the order of the giant component was established by Łuczak and Łuczak in [117], as part of a result about the more general ‘random-cluster model’.

Recently, Nachmias and Peres [129] combined the random walk ideas of Martin-Löf [125], Karp [109] and Aldous [7] with martingale techniques to give a very simple proof of a weak form of Theorem 20; this was extended to the full statement above (but not the stronger results in [138]) in [47].

The quantity $\rho_\lambda = G(\lambda)$ appearing in the results above for the supercritical case has several interpretations, one of which is described in Section 3. Perhaps the most informative is in terms of the *Galton–Watson branching process* with Poisson offspring distribution $\text{Po}(\lambda)$. This is the random rooted tree in which (loosely speaking) each vertex has a $\text{Po}(\lambda)$ number of children, independently of the other vertices. Indeed, ρ_λ is the *survival probability* of this process, i.e., the probability that this random tree is infinite. It is rather easy to see that for λ constant, $G_{n,\lambda/n}$, explored outwards from a given or random vertex, is locally very similar to this process, which suggests that a fraction ρ_λ of the vertices will be in large components. Of course, considerable work is needed to show that not only is this true, but almost all such vertices are in a single ‘giant’ component. Still, this viewpoint makes it very easy to guess what the size of the giant component should be in a large number of random graph (or hypergraph) models; then one must actually *prove* this. In this interpretation λ_* is the parameter of the *dual branching process*, obtained by conditioning on extinction. It is easy to check that this is also a Poisson Galton–Watson process.

Although we shall consider generalizations of $G_{n,p}$ in the next section, there is one that is so close in behaviour to the original model that the results belong very much together, namely the *random k -uniform hypergraph* $H_k(n, p)$. This is the hypergraph with vertex set $[n]$ in which each of the

$\binom{n}{k}$ possible edges (k -element subsets of $[n]$) is present with probability p , independently of the others. There are several possible generalizations of the notion of component to hypergraphs, but the most natural is given by the transitive closure of the relation ‘lie in a common edge’ on the vertex set. In this context the natural scaling corresponding to $p = \lambda/n$ for $k = 2$ (the graph case) is to write

$$(7) \quad p = \lambda(k-2)!/n^{k-1},$$

where $\lambda = \lambda(n) = O(1)$. As shown by Schmidt-Pruzan and Shamir [155], the phase transition is then at $\lambda = 1$. Surprisingly, even the asymptotic size of the giant component was found only much later, by Coja-Oghlan, Moore and Sanwalani [63] in 2007 (though perhaps a form of this result was ‘folklore’ before then). Let

$$\rho_{k,\lambda} = 1 - (1 - \rho_\lambda)^{1/(k-1)}$$

where, as before, ρ_λ is $G(\lambda)$ as defined in (6). Note that $\rho_{k,\lambda}$ can be seen as the survival probability of a certain (compound Poisson) Galton–Watson branching process naturally associated to $H_k(n, p)$.

Theorem 21. *If $k \geq 3$ and $\lambda > 0$ are constant and $p = p(n)$ is defined by (7), then*

$$L_1(H_k(n, p)) = \rho_{k,\lambda}n + o_p(n). \quad \blacksquare$$

Turning to more precise results, Karoński and Łuczak [105] proved a local limit theorem in the barely supercritical phase, when $(\lambda - 1)^3n$ tends to infinity but more slowly than $\log n / \log \log n$. For the strongly supercritical case, where $\lambda > 1$ is fixed, the local limit theorem was proved by Behrisch, Coja-Oghlan and Kang [19]. A limit theorem covering the entire supercritical range was proved only very recently in [48]. For $k \geq 2$ and $\lambda > 1$ let

$$\sigma_{k,\lambda}^2 = \frac{\lambda(1 - \rho_{k,\lambda})^2 - \lambda_*(1 - \rho_{k,\lambda}) + \rho_{k,\lambda}(1 - \rho_{k,\lambda})}{(1 - \lambda_*)^2},$$

where, as before, $\lambda_* < 1$ satisfies $\lambda_*e^{-\lambda_*} = \lambda e^{-\lambda}$.

Theorem 22. *Let $k \geq 3$ be fixed, and define $p = p(n)$ by (7) where $\lambda = \lambda(n)$ is bounded and $(\lambda - 1)^3n \rightarrow \infty$. Then*

$$\frac{L_1(H_k(n, p)) - \rho_{k,\lambda}n}{\sigma_{k,\lambda}\sqrt{n}} \xrightarrow{d} \text{No}(0, 1),$$

where \xrightarrow{d} denotes convergence in distribution and $\text{No}(0, 1)$ is a standard normal random variable. \blacksquare

In fact, it turns out that it is possible to start from this result and use a smoothing argument (and some detailed tree counting results) to deduce a local limit theorem; see [50].

Aldous' result for the critical case, Theorem 18, is also generalized to k -uniform hypergraphs in [48].

6.2. Structural results

So far we have mainly discussed the order of the giant component, although many of the papers mentioned above discuss at least a little more, for example the number of edges. What can we say about other properties, for example the typical or maximum graph distance between vertices?

Recall that the *2-core* of a connected graph G is the maximal subgraph $G^{(2)}$ with minimum degree at least 2, which may be empty (if G is a tree). Any connected graph that is not a tree consists of its 2-core and its *mantle*, a set of trees each sharing one vertex with the 2-core. Looking further inside, the *kernel* of G is the multigraph obtained from the 2-core by contracting vertices of degree 2, so the 2-core is formed from the kernel by replacing the edges by internally vertex-disjoint paths. This way of viewing a graph was introduced in [28], and has been used many times since then, for example in [138]. For example, Łuczak [121] studied the degree distribution of the kernel of the young giant component (where $p = (1 + \varepsilon)/n$ with $\varepsilon^3 n \rightarrow \infty$ but $\varepsilon = o(1)$), using the fact that almost all vertices have degree 3 to show that (conditional on the whp Hamiltonicity of a random 3-regular graph, proved just after by Robinson and Wormald [150]) whp the largest component of $G_{n,(1+\varepsilon)/n}$ contains a cycle of length at least $(\frac{4}{3} + o(1))\varepsilon^2 n$.

Jumping to more recent results, Ding, Kim, Lubetzky and Peres [68] have given a detailed, and very usable, description of the distribution of the young giant component of $G_{n,p}$. To state this, let $\text{No}(\mu, \sigma^2)$ denote the normal distribution with mean μ and variance σ^2 , and $\text{Geom}(\varepsilon)$ the geometric distribution with mean $1/\varepsilon$. The *model giant component* $\tilde{\mathcal{L}}_1$ introduced in [68] is defined in three steps: First, let $Z \sim \text{No}(\frac{2}{3}\varepsilon^3 n, \varepsilon^3 n)$, and let K be a random 3-regular multigraph on $N = 2\lfloor Z \rfloor$ vertices. Next, replace each edge of K by a path, where the path lengths are independent and have the distribution $\text{Geom}(\varepsilon)$. Finally, attach a Poisson $\text{Po}(1 - \varepsilon)$ Galton–Watson branching process to each vertex. In the special case where $\varepsilon = o(n^{-1/4})$, the main result of [68] can be stated as follows. When describing the structure, not just the order, of the largest component of a graph G , we write \mathcal{L}_1 for this component (chosen according to any rule in the case of a tie).

Theorem 23. Let $\mathcal{L}_1 = \mathcal{L}_1(G_{n,p})$ be the largest component of the random graph $G_{n,p}$ for $p = (1 + \varepsilon)/n$, where $\varepsilon^3 n \rightarrow \infty$ and $\varepsilon = o(n^{-1/4})$. Then \mathcal{L}_1 is contiguous to the model $\tilde{\mathcal{L}}_1$, in the sense that, for any graph property \mathcal{Q} , $\lim_{n \rightarrow \infty} \mathbb{P}(\tilde{\mathcal{L}}_1 \text{ has } \mathcal{Q}) = 0$ implies that $\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{L}_1 \text{ has } \mathcal{Q}) = 0$. ■

The general result of Ding, Kim, Lubetzky and Peres [68] covers the entire range where $\varepsilon^3 n \rightarrow \infty$ and $\varepsilon \rightarrow 0$; the statement is more involved, since the kernel typically contains many vertices of degree greater than 3.

Theorem 23 and its more general companion are key ingredients in one of the main approaches to more detailed analysis of the young giant component. For example, using earlier results on first-passage percolation due to Bhamidi, Hooghiemstra and van der Hofstad [22], Ding, Kim, Lubetzky and Peres [68] obtained the following result.

Corollary 24. Let $\varepsilon^3 n \rightarrow \infty$ and $\varepsilon \rightarrow 0$, and let $\mathcal{L}_1^{(2)}$ be the 2-core of the largest component of the random graph $G_{n,p}$ where $p = (1 + \varepsilon)/n$. If v and w are two vertices of degree at least 3 in $\mathcal{L}_1^{(2)}$ chosen uniformly at random among all such vertices, then the distance between v and w is whp $(1/\varepsilon + O(1)) \log(\varepsilon^3 n)$. ■

Turning to the much more difficult question of the maximum distance between vertices, this approach now requires results concerning extreme-value estimates for first passage percolation. Using such ideas, Ding, Kim, Lubetzky and Peres [67] obtained the following result.

Theorem 25. Consider the random graph $G_{n,p}$ for $p = (1 + \varepsilon)/n$, where $\varepsilon^3 n \rightarrow \infty$ and $\varepsilon \rightarrow 0$. Let \mathcal{L}_1 be the largest component of $G_{n,p}$, with 2-core $\mathcal{L}_1^{(2)}$ and kernel K . Then, whp,

$$(8) \quad \begin{aligned} \text{diam}(\mathcal{L}_1) &= (3 + o(1)) (1/\varepsilon) \log(\varepsilon^3 n), \\ \text{diam}(\mathcal{L}_1^{(2)}) &= (2 + o(1)) (1/\varepsilon) \log(\varepsilon^3 n), \\ \max_{v,w \in K} d_{\mathcal{L}_1^{(2)}}(v, w) &= \left(\frac{5}{3} + o(1) \right) (1/\varepsilon) \log(\varepsilon^3 n). \quad \blacksquare \end{aligned}$$

In the last statement, the distance is measured in the 2-core $\mathcal{L}_1^{(2)}$, not in the kernel.

Stepping back a little, historically the diameter of $G_{n,p}$ was originally only studied when this graph is likely to be connected. After all, what is the

diameter of a disconnected graph? Burtin [57, 58] proved the first results, and Bollobás [28] gave a very precise formula for the diameter at exactly the point where the graph first becomes connected. Turning to the case of bounded average degree, in the subcritical regime Łuczak [122] proved very precise results for the maximum of the diameters of the components. For the supercritical case, with $\lambda > 1$ constant, the question was surprisingly neglected. Chung and Lu [61] proved a partial result, but the correct asymptotic formula was obtained only much later, independently by Fernholz and Ramachandran [83] and by Bollobás, Janson and Riordan [39], in both papers as a special case of a result for a much more general model.

Finally, Riordan and Wormald [149] obtained essentially best-possible estimates for the diameter of $G_{n,p}$ throughout the supercritical regime. For the young giant component, they proved the following.

Theorem 26. *Let $\varepsilon = \varepsilon(n)$ satisfy $0 < \varepsilon < \frac{1}{10}$ and $\varepsilon^3 n \rightarrow \infty$. Set $\lambda = \lambda(n) = 1 + \varepsilon$, and let $\lambda_* < 1$ be the dual of λ . Then*

$$(9) \quad \text{diam}(G_{n,\lambda/n}) = \frac{\log(\varepsilon^3 n)}{\log \lambda} + 2 \frac{\log(\varepsilon^3 n)}{\log(1/\lambda_*)} + O_p(1/\varepsilon). \quad \blacksquare$$

Of course, this result is consistent with, and thus refines, (8). As with the results of Bollobás [29] and Łuczak [118] on the order of the giant component, Theorem 26 is sharp in that the variation of the diameter is of order $1/\varepsilon$. However, one can of course ask for more: what is the limiting distribution of the error term? This is also answered in [149], but the statement is a little involved.

In the critical case, the behaviour of the diameter of $G_{n,p}$ is even more complicated. Addario-Berry, Broutin and Goldschmidt [3] found the limiting distribution of the diameter divided by $n^{1/3}$, by proving convergence of $\mathcal{L}_1(G_{n,p})$ to a certain (complicated) random continuum limit object in a certain sense.

7. NEW RANDOM GRAPH MODELS

One of the key developments in the study of phase transitions is the generalization of the results, and sometimes methods, of Erdős and Rényi to a vast array of different models of random graphs. Our main focus will be the (mostly) new ‘inhomogeneous’ models, but first we consider a classical homogeneous model.

7.1. The configuration model

Let $r \geq 3$ be fixed, and let G_r be a random r -regular graph on $[n]$, i.e., an r -regular graph on $[n]$ chosen uniformly at random from all such graphs. (We assume rn is even, of course.) This graph is connected whp, but what about a random *subgraph* of this graph? Let $G_r[p]$ be the random subgraph of G_r obtained by retaining each edge with probability p , independently of the others and of G_r itself. Is there a phase transition in the size of the largest component as p is varied? The answer is yes: the critical point, $p = 1/(r - 1)$, was found by Goerdts [91].

There is a natural family of inhomogeneous models extending G_r . Given a degree sequence $\mathbf{d} = (d_1, \dots, d_n)$ satisfying appropriate conditions, we may consider a graph $G_{\mathbf{d}}$ chosen uniformly at random from all graphs on $[n]$ with the property that each vertex i has degree d_i . The analysis of this graph, and indeed of G_r , is almost always based on a construction due to Bollobás [26], who defined the *configuration multigraph* with degree sequence \mathbf{d} using a random pairing of $\sum_i d_i$ objects. Sometimes one studies $G_{\mathbf{d}}$ using this model, and sometimes the multigraph directly. One can of course study random subgraphs of $G_{\mathbf{d}}$ but (at least for the multigraph) there is not too much reason to do so, since they can themselves be seen as new instances of the model with appropriate (random) degree sequences. (This is spelled out in detail by Fountoulakis [84].)

Molloy and Reed [127, 128] studied this *configuration model* $G_{\mathbf{d}}$ under fairly general conditions, finding the size of the largest component up to a $o(n)$ error under some mild assumptions. This result has been extended and generalized in various directions. On the one hand, one can ask for the weakest conditions under which the asymptotic size of the giant component can be established. Results of this type were proved by Janson and Luczak [94], for example. For asymptotic results we of course consider a sequence (\mathbf{d}_n) of degree sequences with (for notational simplicity) \mathbf{d}_n having length n . Let $n_i(\mathbf{d})$ denote the number of times degree i appears in \mathbf{d} , and $m(\mathbf{d})$ half the sum of the entries of \mathbf{d} , i.e., the number of edges of $G_{\mathbf{d}}$. In order to be able to say something about the limiting behaviour, it is natural to assume that

$$(10) \quad \lim_{n \rightarrow \infty} \frac{n_i(\mathbf{d}_n)}{n} = p_i$$

for each i , for some number p_i which we view as the probability $\mathbb{P}(D = i)$ that the limiting degree distribution takes the value i . It also turns out to be necessary to assume that

$$(11) \quad \frac{m(\mathbf{d}_n)}{n} \rightarrow \frac{\mathbb{E}(D)}{2} = \frac{1}{2} \sum_{i=0}^{\infty} i p_i.$$

Perhaps the minimal conditions are those of the following result from [49], whose proof uses several ideas from the original paper of Erdős and Rényi [75]; this result applies both to the random simple graph $G_{\mathbf{d}_n}$ and to the configuration multigraph.

Theorem 27. *Let D be a probability distribution on the non-negative integers with $0 < \mathbb{E}(D) < \infty$ and $\mathbb{P}(D \geq 3) > 0$, and let (\mathbf{d}_n) be a sequence of degree sequences converging to D in the sense that (10) and (11) hold, with \mathbf{d}_n having length n . Then*

$$L_1(G_{\mathbf{d}_n}) = \rho(D)n + o_p(n)$$

and $L_2(G_{\mathbf{d}_n}) = o_p(n)$. ■

Of course, one can ask for more precise results. For example, what is the width of the ‘scaling window’ of this phase transition? Results of this type, either for the special case of r -regular graphs or for the configuration model (with some assumptions) were given by Kang and Seierstad [104], Pittel [136] and Janson and Luczak [94], all with logarithmic gaps in the bounds. The precise width of the scaling window was first established for random r -regular graphs by Nachmias and Peres [130], using a variant of their ideas in [129]. This result was extended to the configuration model with bounded degrees by Riordan [143], who established not only the asymptotic size of the giant component in the subcritical, critical and weakly supercritical ranges, but also the scale and limiting distribution of its fluctuations. To state the supercritical case of this result, given a degree sequence \mathbf{d} let $\mu_r = \mu_r(\mathbf{d}) = n^{-1} \sum_{i=1}^n (d_i)_r$ be its r th factorial moment, and let $\lambda = \lambda(\mathbf{d}) = \mu_2/\mu_1$. Furthermore, let ρ and ρ^* be two quantities whose precise definition is given in [143] in terms of a certain branching process; these satisfy

$$\rho \sim \frac{2\mu_1^2}{\mu_3}\varepsilon \quad \text{and} \quad \rho^* \sim \frac{2\mu_1^3}{3\mu_3^2}\varepsilon^3$$

as $\varepsilon \rightarrow 0$, where $\varepsilon = \lambda - 1$.

Theorem 28. *Let $\Delta \geq 2$ and $c_0 > 0$ be fixed. For each n let $\mathbf{d} = \mathbf{d}_n$ be a degree sequence of length n with maximum degree at most Δ with at least c_0n vertices having degree not in $\{0, 2\}$. Define μ_i , λ , ρ and ρ^* as above, noting that these quantities depend on n . Setting $\varepsilon = \lambda - 1$, suppose that $\varepsilon \rightarrow 0$ and $\varepsilon^3n \rightarrow \infty$. Let L_1 and N_1 denote the order and nullity of the largest component of $G_{\mathbf{d}_n}$. Then $L'_1 = L_1 - \rho n$ and $N'_1 = N_1 - \rho^* n$ are asymptotically jointly normally distributed with mean 0,*

$$\text{Var}(L'_1) \sim 2\mu_1\varepsilon^{-1}n, \quad \text{Var}(N'_1) \sim 5\rho^*n \sim \frac{10\mu_1^3}{3\mu_3^2}\varepsilon^3n,$$

and

$$\text{Cov}(L'_1, N'_1) \sim \frac{2\mu_1^2}{\mu_3} \varepsilon n.$$

Furthermore,

$$L_2(G_{\mathbf{d}_n}) = O_p(\varepsilon^{-2} \log(\varepsilon^3 n)). \quad \blacksquare$$

Independently, for the critical case Joseph [100] proved precise results for the component sizes under weaker assumptions. In a similar vein, Hatami and Molloy [92] established the width of the phase transition for a more general set of degree sequences, including cases where it is no longer $n^{-1/3}$ as in the cases above.

7.2. The Watts–Strogatz model

One of the most prophetic statements of Erdős and Rényi in [75] was the suggestion that variants of their model in which edges are not equiprobable might be good models for real-world networks such as communication or electrical networks. The subject of using random graphs as models for real-world networks is of course much too large to cover here, but let us mention a few of the very many examples from the last 15 years, during which this has become a very active field.

In the 1960s, Milgram [126] and others noticed that many *social networks*, i.e., real-world graphs where the vertices are people and the edges represent, for example, who is acquainted with whom, exhibit something called the ‘small world phenomenon’, or ‘six degrees of separation’: the apparently surprising phenomenon that in graphs of this type where the number n of vertices is very large, even if the average degree is not that large, the typical distance between vertices is relatively small – of order $\log n$. If we view the graph as a random graph of something like the Erdős–Rényi type, then mathematically this is exactly what one would expect; see Section 6.2. However, as pointed out by Watts and Strogatz [166], social networks are in many respects very different from $G_{n,p}$: they exhibit high *clustering*, the phenomenon that two neighbours of a given vertex v are much more likely to be joined to each other than a random pair of vertices. This is a typical property of geometric networks where vertices are connected to other vertices within some given distance, say; such networks tend to have large diameter.

In a very influential paper [166], Watts and Strogatz proposed a new model with both features, clustering and small distances. The idea is very simple: start from the r -th power of a cycle, i.e., the graph with vertex set $\{1, \dots, n\}$ in which two vertices are joined if and only if their

distance around the cycle $12 \cdots n1$ is at most r . Then ‘rewire’ a fraction of the edges randomly: for each edge, with some probability β replace one or both endpoints by vertices chosen uniformly at random. Watts and Strogatz simulated such networks and showed that they do indeed have logarithmic diameter (with the implicit constant depending on β) as well as high clustering: they called networks with this property ‘small world’ networks.

Mathematically, once one thinks of this type of model, it is easy to see that it has the desired properties: high clustering is inherited from the initial graph before rewiring, and the small diameter essentially follows from that of $G_{n,p}$. Indeed, a very similar graph, namely a cycle plus a random matching, had been studied much earlier by Bollobás and Chung [36] who found its asymptotic diameter, which is indeed logarithmic.

7.3. Scale-free and other growing models

In 1999, Faloutsos, Faloutsos and Faloutsos [82] and others noticed that many real-world graphs are ‘scale-free’ in the sense that certain key features, and in particular the degree distribution, follow a power law. This contrasts sharply with the Poisson distribution seen in sparse instances of $G_{n,p}$. Although power-law distributions had been observed much earlier, for example by Lotka [115] in the distribution of citations in academic literature, around this time there was an explosion of activity based on modelling, or trying to explain, such power laws. In one direction, Aiello, Chung and Lu [4] proposed a model for ‘massive graphs’ of this type, namely the configuration model with a fixed, power-law degree sequence. In a very different direction, Barabási and Albert [14] proposed a model to explain how such power laws might arise, based on growth with preferential attachment:

“... starting with a small number (m_0) of vertices, at every time step we add a new vertex with $m (\leq m_0)$ edges that link the new vertex to m different vertices already present in the system. To incorporate preferential attachment, we assume that the probability Π that a new vertex will be connected to a vertex i depends on the connectivity k_i of that vertex, so that $\Pi(k_i) = k_i / \sum_j k_j$. After t steps the model leads to a random network with $t + m_0$ vertices and mt edges.”

The idea is to provide a highly simplified model of the evolution of, for example, the world-wide web. New webpages (or sites) are added one-at-a-time, and link to earlier pages chosen with probabilities depending on how many existing pages link to them. This mechanism represents the fact that the creator of a new page is more likely to know about an existing page

that has many links, or more generally the phenomenon that ‘popularity is attractive’.

Mathematically, the model as described by Barabási and Albert does not quite make sense. The fundamental problem is that one cannot describe the distribution of a random set simply by specifying the probability that it contains each given element. A precisely formulated version of the Barabási–Albert model, the *LCD model*, was introduced in [44]. This may be defined inductively as follows: start with $G_1^{(0)}$ the empty ‘graph’ with no vertices, or with $G_1^{(1)}$ the graph with one vertex and one loop. Given $G_1^{(t-1)}$, form $G_1^{(t)}$ by adding the vertex v_t together with a single edge between v_t and v_i , where i is chosen randomly with

$$(12) \quad \mathbb{P}(i = s) = \begin{cases} d_{G_1^{(t-1)}}(v_s)/(2t-1) & 1 \leq s \leq t-1, \\ 1/(2t-1) & s = t. \end{cases}$$

In other words, send an edge e from v_t to a random vertex v_i , where the probability that a vertex is chosen as v_i is proportional to its degree at the time, counting e as already contributing one to the degree of v_t . (The reason that this is convenient is that it allows an alternative ‘static’ description of the model in terms of linearized chord diagrams; see [44].) For $m > 1$, add m edges from v_t one-at-a-time, counting the previous edges as well as the ‘outward half’ of the edge being added as already contributing to the degrees. Equivalently, define the process $(G_m^{(t)})_{t \geq 0}$ by running the process $(G_1^{(t)})$ on a sequence v'_1, v'_2, \dots , and forming the graph $G_m^{(t)}$ from $G_1^{(mt)}$ by identifying the vertices v'_1, v'_2, \dots, v'_m to form v_1 , identifying $v'_{m+1}, v'_{m+2}, \dots, v'_{2m}$ to form v_2 , and so on.

Bollobás, Riordan, Spencer and Tusnády [51] showed that the LCD model does indeed have a power-law degree distribution, confirming experimental and heuristic predictions of Barabási and Albert [14]. Computer experiments of Barabási, Albert and Jeong [6, 15] and heuristic arguments given by Newman, Strogatz and Watts [132] suggested that models of the Barabási–Albert type should have diameter $\Theta(\log n)$. In fact, as shown in [44] the diameter is slightly smaller.

Theorem 29. *Fix an integer $m \geq 2$ and a positive real number ε . Then whp $G_m^{(n)}$ is connected and has diameter $\text{diam}(G_m^{(n)})$ satisfying*

$$(1 - \varepsilon) \log n / \log \log n \leq \text{diam}(G_m^{(n)}) \leq (1 + \varepsilon) \log n / \log \log n. \quad \blacksquare$$

Even more than the Watts–Strogatz model, the Barabási–Albert model has been extremely influential, stimulating the writing of hundreds of papers

on related models incorporating preferential attachment and similar mechanisms. Some early examples of models inspired by it are the ‘copying’ model of Kumar, Raghavan, Rajagopalan, Sivakumar, Tomkins and Upfal [114], the very general models defined by Cooper and Frieze [64], and a model for directed graphs with preferential attachment introduced by Bollobás, Borgs, Chayes and Riordan [34]. Within a very few years of the paper of Barabási and Albert, there was enough activity in this area to justify several survey papers (for example [5, 69]) and even several books [13, 70, 164, 165].

Although, as in the example above, it is possible to establish some properties of some of these new models rigorously, often the dependence inherent in the construction, with the distribution of each edge added depending on the arrangement of those already present, means that they are rather hard to analyse. For this reason, often the ‘mean-field’ versions are analysed instead: such models have (roughly) the same edge probabilities as the evolving model, but with different edges present independently. We shall return to this in the next subsection.

The work on scale-free random graphs led to renewed interest in evolving random graphs in general. For example, Callaway, Hopcroft, Kleinberg, Newman and Strogatz [59] proposed a random graph model based on growth without preferential attachment. Roughly speaking, vertices and edges are added at a uniform rate, with edges joining uniformly random vertices. Depending on the value of a density parameter, it turns out that there may or may not be a giant component. The authors of [59] gave heuristic arguments for the critical value of this parameter. In fact, as noted by Durrett [71] and Bollobás, Janson and Riordan [38], this ‘CHKNS’ model is very closely related to one proposed by Dubins in 1984 (see [102, 156]). To define this, let $c < 2$ be a density parameter. Given the number n of vertices, let $G_n(c)$ be the random graph with vertex set $[n]$ in which edges are present independently, and the probability of an edge ij , $i < j$, is c/j . Similarly, let $G_\infty(c)$ be the corresponding graph on $\{1, 2, \dots\}$. Dubins asked for which c the graph $G_\infty(c)$ is almost surely connected: this was answered by Kalikow and Weiss [102] and Shepp [156], who showed that $c = 1/4$ is the critical value. As pointed out in [71, 38], these much earlier results easily determine the critical value in the CHKNS model.

One particularly interesting feature of Dubins’ model is the nature of the phase transition. In the Erdős–Rényi model $G_{n,p}$, if the edge density is a factor $1 + \varepsilon$ times the critical one, then the ‘giant’ component has order $\Theta(\varepsilon n)$ (more precisely, $\sim 2\varepsilon n$ for $\varepsilon \rightarrow 0$ more slowly than $n^{-1/3}$). In Dubins’ model, just above the phase transition the giant component is *much* smaller: as shown in [38] it contains $\exp(-\Theta(1/\sqrt{\varepsilon}))n$ vertices. More precisely, the following result is proved in [141].

Theorem 30. *There is a function $f(c)$ such that for $c > 0$ fixed, $L_1(G_n(c)) = f(c)n + o_p(n)$. Furthermore, $f(c) = 0$ for $c \leq 1/4$, and*

$$(13) \quad f(1/4 + \varepsilon) = \exp\left(-\frac{\pi}{2} \frac{1}{\sqrt{\varepsilon}} + O(\log(1/\varepsilon))\right)$$

as $\varepsilon \rightarrow 0$ from above. ■

This result carries over to the CHKNS model, taking $c = 2\delta$, where δ is the edge-density parameter in [59].

7.4. The BJR model

Stepping backwards in time for a moment, let us recall that Erdős and Rényi themselves suggested that it might make sense to consider variants of their model in which vertices have different ‘types’, and edges are not equiprobable, presumably with the probability depending on the types of the vertices. Some such models are too obvious to be anything but ‘folklore’, for example, the random bipartite graph $G_{n,n,p}$ with two vertex classes of size n in which each possible edge between the classes has probability p .

Going further in this direction, Söderberg [157] proposed and studied the following model. Let $k \geq 1$, let A be a symmetric k -by- k matrix with non-negative entries, and let $\mu = (\mu_1, \dots, \mu_k)$ be a vector of probabilities summing to 1. Define a random graph $G(n, A, \mu)$ as follows: for each vertex $i = 1, 2, \dots, n$, first choose its type x_i from $\{1, 2, \dots, k\}$ according to the distribution μ , with these choices independent for different vertices. Then, conditional on these choices, let $G(n, A, \mu)$ be the random graph in which edges are present independently, with the probability of an edge joining i and j being $A_{x_i, x_j}/n$. (Here we assume n is larger than the maximum entry in A ; otherwise take the minimum of $A_{x_i, x_j}/n$ and 1, say.)

In 2007, Bollobás, Janson and Riordan [39] greatly generalized this model, to one whose special cases include many of the other inhomogeneous models considered to that point, as well as the ‘mean-field’ versions of many others. We shall not describe the model in full generality, as this takes some time; a not-too-restrictive special case may be described as follows. Let κ be an integrable symmetric non-negative function on $[0, 1]^2$, called a *kernel* in [39]. Also, let μ be a probability measure on $[0, 1]$ (Lebesgue measure being the most natural special case). Suppressing the dependence on μ in the notation, let $G(n, \kappa)$ be the random graph on $[n] = \{1, 2, \dots, n\}$ constructed as follows: first choose the vertex types $x_1, \dots, x_n \in [0, 1]$ independently at random according to the distribution μ . Then let $G(n, \kappa)$ be the random graph in which edges are present conditionally (on the types) independently,

with the conditional probability of an edge ij being $\kappa(x_i, x_j)/n$ if this is less than 1, and 1 otherwise. (This version of the model is not actually covered by the definitions in [39], but the results extend to it; see [41]).

Clearly, the BJR model generalizes that proposed by Söderberg: divide $[0, 1]$ into k intervals I_i with lengths μ_i and take κ to be piecewise constant, taking the value $A_{i,j}$ on $I_i \times I_j$. In the form introduced in [39], it also includes the finite version $G_n(c)$ of Dubins’ model described above. Indeed, the conditions in [39] allow the vertex types to be fixed, rather than random, as long as their distribution converges to μ in a suitable sense as $n \rightarrow \infty$. In this version some technical assumptions are needed, but these apply with, for example, $\kappa(x, y) = c/\max\{x, y\}$ and vertex i of $G(n, \kappa)$ having type i/n ; it is easy to see that then $G(n, \kappa)$ is exactly $G_n(c)$. Similarly, taking $\kappa(x, y) = c/\sqrt{xy}$ gives a mean-field version of the Barabási–Albert model. There are also choices for μ and κ giving, essentially, the dynamical random graph model introduced by Turova [160, 161], and so on.

Many properties of $G(n, \kappa)$ are established in [39]; the most important is the critical point of the phase transition, and the asymptotic size of the giant component when there is one. To state this result we need a few definitions. Firstly, a kernel κ is *reducible* if there is a set $A \subset [0, 1]$ with $0 < \mu(A) < 1$ such that $\kappa = 0$ a.e. on $A \times ([0, 1] \setminus A)$, and *irreducible* otherwise. Note that if κ is reducible then $G(n, \kappa)$ splits automatically into two or more pieces: vertices with types in A cannot be connected to those whose type is not in A .

Given a kernel κ , let T_κ be the integral operator on $([0, 1], \mu)$ with kernel κ , defined by

$$(T_\kappa f)(x) = \int_{[0,1]} \kappa(x, y) f(y) d\mu(y),$$

for any (measurable) function f such that this integral is defined (finite or $+\infty$) for a.e. x . Let

$$\|T_\kappa\| = \sup\{\|T_\kappa f\|_2 : f \geq 0, \|f\|_2 \leq 1\}.$$

When finite, $\|T_\kappa\|$ is the norm of T_κ as an operator on $L^2([0, 1], \mu)$; it is infinite if T_κ does not define a bounded operator on this space. Finally, define a non-linear operator Φ_κ by

$$\Phi_\kappa f = 1 - e^{-T_\kappa f}$$

for $f \geq 0$, let $\rho = \rho_\kappa : [0, 1] \rightarrow [0, 1]$ be the pointwise largest solution to the functional equation $\Phi_\kappa(\rho) = \rho$, and let

$$\rho(\kappa) = \int_{[0,1]} \rho_\kappa(x) d\mu(x).$$

A version of the main result of [39] may now be stated as follows.

Theorem 31. *Let κ be a kernel, i.e., an integrable symmetric non-negative function on $[0, 1]^2$, let μ be a probability measure on $[0, 1]$, and define the random graph $G(n, \kappa)$ as above. If κ is irreducible then*

$$L_1(G(n, \kappa)) = \rho(\kappa)n + o_p(n)$$

and $L_2(G(n, \kappa)) = o_p(n)$. Moreover, $\rho(\kappa) > 0$ if and only if $\|T_\kappa\| > 1$. ■

Since the BJR model includes as special cases many previously studied models, Theorem 31 generalizes, at least in part, many results about their phase transitions. Indeed, this result shows that to find the critical point of the phase transition we ‘merely’ need to find the norm of a corresponding operator, and to find the size of the giant component we need to solve a certain non-linear integral equation. Depending on the kernel κ , this may not be so easy. For example, Theorem 31 certainly implies the first statement (existence of $f(c)$) of Theorem 30, but for the main part one still needs to analyse solutions to the equation $\Phi_\kappa(\rho) = \rho$, and that is where the work in [141] is. For further examples see [39, Section 16].

One of the most interesting questions about the phase transition in any random graph model is the following: how big is the giant component just above the critical point? In the context of the BJR model, if we fix a kernel κ and vary a scaling parameter c , the question is how the function $c \mapsto \rho(c\kappa)$ behaves as c tends down to the critical value $c_0 = 1/\|T_\kappa\|$ from above. In [39, Section 16] examples are given of cases in which $\rho((c_0 + \varepsilon)\kappa) = \Theta(\varepsilon^d)$ for any $d \in [1, \infty)$ – in other words, we may have a phase transition of any finite order $d \geq 1$. Theorem 30 gives an example where the order is infinite: $\rho((c_0 + \varepsilon)\kappa) = o(\varepsilon^d)$ as $\varepsilon \rightarrow 0$ for any finite d .

What about the other direction? Are there phase transitions that are steeper than that in $G_{n,p}$? For the BJR model, the answer is no, at least under a certain additional assumption.

Theorem 32. *Let κ be an irreducible kernel on $([0, 1], \mu)$ such that*

$$(14) \quad \sup_x \int_{[0,1]} \kappa(x, y)^2 d\mu(y) < \infty,$$

and let $c_0 = \|T_\kappa\|^{-1} > 0$. Then $c_0 \rho'_+(c_0) \leq 2$, with equality in the classical Erdős–Rényi case; more precisely, equality holds if and only if

$$c_0 \int_{[0,1]} \kappa(x, y) d\mu(y) = 1 \quad \text{for a.e. } x. \quad \blacksquare$$

In [39], the above result is proved as a special case of the following result, establishing the initial rate of growth of the giant component for any kernel satisfying (14). Note that one can think of ε as constant or as tending to 0 slowly as $n \rightarrow \infty$.

Theorem 33. *Let κ be a kernel on $([0, 1], \mu)$. Suppose that κ is irreducible, and that (14) holds.*

1. *The function $c \mapsto \rho(c) = \rho(c\kappa)$ is analytic except at $c_0 = \|T_\kappa\|^{-1}$.*
2. *The linear operator T_κ has an eigenfunction ψ of eigenvalue $\|T_\kappa\| < \infty$, and every such eigenfunction is bounded and satisfies*

$$(15) \quad \rho(c_0 + \varepsilon) = 2c_0^{-1} \frac{\int_{[0,1]} \psi \int_{[0,1]} \psi^2}{\int_{[0,1]} \psi^3} \varepsilon + O(\varepsilon^2), \quad \varepsilon > 0,$$

so $\rho'_+(c_0) = 2c_0^{-1} \int_{[0,1]} \psi \int_{[0,1]} \psi^2 / \int_{[0,1]} \psi^3 > 0$ and ρ has a phase transition at c_0 with exponent 1. ■

Of course, there are many further questions that one can ask about the BJR model. For example, how large is the giant component in the subcritical case $\|T_\kappa\| < 1$? It seems that the model is too general for a single comprehensive answer to this, but certain special cases are relatively well understood. For example, in the *rank 1* case we assume that $\kappa(x, y) = \psi(x)\psi(y)$ for some $\psi : [0, 1] \rightarrow [0, \infty)$. In this case $G(n, \kappa)$ is a version of the Norros–Reittu model [133]. Under some additional assumptions (including that there are only countably many vertex types) Turova [163] showed that in the subcritical rank 1 case $L_1(G(n, \kappa))$ is whp asymptotic to $c_\kappa \log n$, for some constant c_κ that she determined. Again in the rank 1 case, under a third moment condition on ψ Bhamidi, van der Hofstad and van Leeuwen [23] and independently Turova [162] proved an analogue of Aldous’ result (Theorem 18) for the critical case. In a different direction, Janson and Riordan [98] studied the ‘susceptibility’ (the average component size, appropriately defined) in the sub- and super-critical cases, and in [97] proved a duality result showing that deleting the giant component from a supercritical instance of the model leaves a subcritical one, as one might expect.

Let us briefly mention one very interesting connection between the BJR model and the theory of graph limits developed by Lovász and Szegedy [116] and Borgs, Chayes, Lovász, Sós and Vesztegombi [54, 55]. Roughly speaking, these authors showed that any sequence of dense graphs (with $\Theta(n^2)$ edges, where n is the number of vertices) has a subsequence that converges in any one of several natural senses that they show to be equivalent; these results have immeasurably improved our understanding of the space of dense graphs. (In fact, the definitions make sense for arbitrary sequences of graphs, but the notion of convergence is informative only in the dense case: any sequence of graphs with $o(n^2)$ edges converges to the same (zero) limit object.) One way of describing the limit objects in this theory is as

graphons, i.e., symmetric measurable functions $\kappa : [0, 1]^2 \rightarrow [0, 1]$. Except that the values must be bounded by 1, a graphon is exactly a kernel. Moreover, the results of [116, 54, 55] can be seen as saying that any sequence of graphs has a subsequence that can be thought of as ‘inhomogeneous quasi-random’ (see the discussion in [46]). Unsurprisingly, the corresponding inhomogeneous *random* graphs play an important role in the theory. These were introduced in [116] and can be seen as a dense equivalent of the BJR model: given a graphon κ , the random graph $G_1(n, \kappa)$ is constructed by choosing x_1, \dots, x_n independently and uniformly from $[0, 1]$ and then inserting each possible edge ij with probability $\kappa(x_i, x_j)$. Of course, just as when studying $G_{n,p}$, the types of properties of these graphs studied in the dense and sparse cases are very different. Also, in the sparse case it is very important that the kernels need not be bounded.

Of course, there is a natural way of associating a sparse random graph (formally, a sequence of such graphs) to a dense graph (sequence of dense graphs): simply take a random subgraph, retaining edges independently with probability c/n , where n is the number of vertices. Bollobás, Borgs, Chayes and Riordan [35] showed that if we start with a sequence that converges (in the sense of [54, 55]) to a graphon κ , then the subgraphs have a giant component if and only if $c\|T_\kappa\| > 1$, and, assuming irreducibility, its order is then $\rho(c\kappa)n + o_p(n)$. Random graphs of this type, as well as the BJR model, can be seen in the following general framework: construct somehow a random sequence of matrices A_n , and then, given A_n , construct a random graph $G_n(A_n)$ on $[n]$ by taking the entries of A_n divided by n to give the edge probabilities. In both cases, the matrices A_n converge in some sense to a kernel κ . Bollobás, Janson and Riordan [40] established a common generalization of the main results of [35] and [39] by giving what is perhaps the weakest convergence condition under which it is sensible to expect to be able to describe the phase transition in terms of the kernel; this turns out to be exactly convergence in the cut metric of Borgs, Chayes, Lovász, Sós and Vesztergombi [54]! For the details, see [40].

Although the BJR model has many attractive features, like the Erdős–Rényi model it lacks an important feature of many real-world networks, namely significant clustering. Oversimplifying significantly, in a graph such as $G(n, \kappa)$ where the edge probabilities are of order $1/n$ and edges are (conditionally) independent, the expected number of triangles is $O(1)$. (This is not quite true since κ need not be bounded.) In contrast, many real-world networks with $\Theta(n)$ edges contain $\Theta(n)$ triangles. Models which do show significant clustering necessarily have some form of dependence between edges; it is not so easy to introduce this while keeping the model tractable. In the context of the configuration model, Newman [131] suggested one natural approach: each vertex starts with some edge stubs and some triangle

stubs, then the edge stubs are randomly paired and the triangle stubs randomly grouped into threes.

Generalizing the BJR model even further, Bollobás, Janson and Rioridan [41] introduced the ‘sparse inhomogeneous model with clustering’. A special case of this can be described roughly as follows. Start with a *kernel family*, i.e., a family (κ_F) of functions, one for each finite graph F , with κ_F a non-negative measurable function on $[0, 1]^{V(F)}$ that is invariant under the natural action of the automorphism group of F . Define a random graph $G(n, (\kappa_F))$ on $[n]$ by choosing vertex types x_1, \dots, x_n independently and uniformly from $[0, 1]$, and then for each r -tuple v_1, \dots, v_r of distinct vertices, and each graph F on r vertices, inserting a copy of F with these vertices with probability $\kappa_F(x_{v_1}, \dots, x_{v_r})/n^{r-1}$. The special case where $\kappa_F = 0$ for all F other than the single edge K_2 is simply $G(n, \kappa)$ as above. Perhaps surprisingly, even though the operator playing the role that T_κ does for the BJR model is now non-linear, much of the analysis of $G(n, \kappa)$ extends to this generalization; see [41]. Already using non-trivial functions κ_F for $F = K_2$ and $F = K_3$ is enough to generate a variety of random graph models with various types of degree distribution and various clustering coefficients. There is also a cut-metric generalization of this model; see [40].

7.5. Achlioptas processes

We close this section, and thus the paper, by briefly describing yet another new and very active area of research that can ultimately be traced back to the work of Erdős and Rényi. At a Fields Institute workshop in 2000, Dimitris Achlioptas suggested a class of variants of the random graph process \tilde{G}_n , which may be described as follows. Start with the empty graph with n vertices and no edges. At each time step, two potential edges e_1 and e_2 are chosen independently and uniformly at random from all $\binom{n}{2}$ possible edges (or from those edges not already present). One of these edges is selected according to some ‘rule’ \mathcal{R} and added to the graph. The result is a random graph process $\tilde{G}_n^{\mathcal{R}}$ whose distribution of course depends on the rule \mathcal{R} ; these processes are known as ‘Achlioptas processes’. Of course, always selecting e_1 , say, gives (exactly, or approximately, depending on the precise definitions) the classical random graph process \tilde{G}_n .

The original question posed by Achlioptas was whether one can shift the critical point of the random graph process by choosing an appropriate rule. It is easy to see that it can be brought forward (there are rules that lead to inhomogeneous instances of the BJR model, for example); it is more interesting that it can be delayed. Bollobás suggested that the rule most likely to achieve this is the ‘product rule’ – select the edge (of e_1 and e_2)

minimizing the product of the orders of the components it joins. Bohman and Frieze [24] soon showed that a much simpler rule (essentially to select e_1 if it joins two isolated vertices and e_2 otherwise) does indeed delay the appearance of a giant component. This Bohman–Frieze rule is an example of a ‘bounded size’ rule, i.e., one in which the only information used to decide which edge to choose is the sizes of the components that would be joined by e_1 and e_2 , with all sizes above some constant treated the same. (The constant is 1 for the Bohman–Frieze rule.) By now this class of rules is relatively well understood; for example, Bohman and Kravitz [25] and Spencer and Wormald [158] showed that using Wormald’s ‘differential equation method’ [167] one can find the asymptotic number of vertices in components of any given (constant) size at a given point in such a process.

The phase transition in these processes, especially for ‘unbounded’ rules such as the product rule, remained relatively resistant to analysis. It also seemed to be extremely interesting. In particular, making explicit an earlier belief of Achlioptas and others, Achlioptas, D’Souza and Spencer [1] conjectured in 2009 that for the product rule there exists a $\delta > 0$ (in fact, $\delta \geq 1/2$) such that with high probability the order of the largest component ‘jumps’ from $o(n)$ to at least δn in $o(n)$ steps of the process, a phenomenon known as ‘explosive percolation’. If true, this would have been a very interesting example of a random graph model in which the size of the giant component really does jump (though not quite as suddenly as the k -core does in $G_{n,p}$). However, as indicated heuristically for a variant of the product rule by da Costa, Dorogovtsev, Goltsev and Mendes [65] and proved (by a very different argument) rigorously by Riordan and Warnke [144, 146], *no* Achlioptas process can ‘jump’ in this way – they all have continuous phase transitions. These papers also settled various conjectures of Spencer and others, showing that for any bounded size rule there is a limiting function $\rho_{\mathcal{R}}(t)$ so that after tn steps of the process the largest component has size $\rho_{\mathcal{R}}(t)n + o_p(n)$; in addition, Riordan and Warnke [147] gave strong evidence that this need not be the case for general Achlioptas processes.

In the last few years, many papers have been written about Achlioptas processes. For example, results concerning the barely subcritical regime in the Bohman–Frieze process have been published by Janson and Spencer [99] and Kang, Perkins and Spencer [103]. Bhamidi, Budhiraja and Wang have obtained detailed results for bounded size rules in the subcritical [20] and, remarkably, critical regimes [21], using a generalization of Aldous’ multiplicative coalescent process to study the latter.

Despite this work, of which we have listed only a small fraction, many questions remain open. One is the following: what natural random graph processes *do* exhibit explosive percolation? There are trivial examples, such as the rule ‘always join the two smallest components in the whole graph’.

Recently, non-trivial examples were given by Panagiotou, Spöhel, Steger and Thomas [134] but these processes (for example involving choosing a vertex randomly from among the 50% (say) of vertices in the smallest components), are not as natural as one might hope for. Concerning Achlioptas processes themselves, although the phase transitions are always continuous, it seems that they can be very steep. In particular, the heuristics given by da Costa, Dorogovtsev, Goltsev and Mendes [65] suggest that for a rule closely related to the product rule, after $(t_c + \varepsilon)n$ steps of the process (where t_c is the critical value), the largest component has size around $\varepsilon^\beta n$ where the exponent β appears to have the value $0.055 \dots$, in contrast to $\beta = 1$ for the Erdős–Rényi model. (They suggest that perhaps $\beta = 1/18$.) Thus the giant component grows from size $o(n)$ to around $n/2$ in around $n/250000$ steps – a number that is linear in n but with an extremely small constant. It would be very interesting to establish even the most basic properties of this extremely steep phase transition rigorously. At the moment, for the product rule (which appears to have a similarly steep phase transition) it is not even known that the limiting rescaled size $\rho_{\mathcal{R}}(t)$ exists, although partial results have been given by Riordan and Warnke [145, 148]. This fascinating question is just one part of one of the very many active strands of research ultimately arising from the work of Erdős and Rényi on random graphs.

REFERENCES

- [1] Achlioptas, D., R. M. D’Souza and J. Spencer, Explosive percolation in random networks, *Science*, **323** (2009), 1453–1455.
- [2] Achlioptas, D. and A. Naor, The two possible values of the chromatic number of a random graph, *Ann. of Math. (2)*, **162** (2005), 1335–1351.
- [3] Addario-Berry, L., N. Broutin and C. Goldschmidt, The continuum limit of critical random graphs, *Probab. Theory Related Fields*, **152** (2012), 367–406.
- [4] Aiello, W., F. Chung and L. Lu, A random graph model for power law graphs, *Experiment. Math.*, **10** (2001), 53–66.
- [5] Albert, R. and A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Mod. Phys.*, **74** (2002), 47–97.
- [6] Albert, R., H. Jeong and A.-L. Barabási, Diameter of the world-wide web, *Nature*, **401** (1999), 130–131.
- [7] Aldous, D., Brownian excursions, critical random graphs and the multiplicative coalescent, *Ann. Probab.*, **25** (1997), 812–854.
- [8] Alon, N., S. Friedland and G. Kalai, Regular subgraphs of almost regular graphs, *J. Combin. Theory Ser. B*, **37** (1984), 79–91.
- [9] Alon, N. and M. Krivelevich, The concentration of the chromatic number of random graphs, *Combinatorica*, **17** (1997), 303–313.

- [10] Arratia, R. and E. S. Lander, The distribution of clusters in random graphs, *Adv. in Appl. Math.*, **11** (1990), 36–48.
- [11] Austin, T. L., R. E. Fagen, W. F. Penney and J. Riordan, The number of components in random linear graphs, *Ann. Math. Statist.*, **30** (1959), 747–754.
- [12] Babai, L., P. Erdős and S. M. Selkow, Random graph isomorphism, *SIAM J. Comput.*, **9** (1980), 628–635.
- [13] Barabási, A.-L., *Linked: the new science of networks*. Perseus Books; First Printing edition (2003), 288 pages.
- [14] Barabási, A.-L. and R. Albert, Emergence of scaling in random networks, *Science*, **286** (1999), 509–512.
- [15] Barabási, A.-L., R. Albert and H. Jeong, Scale-free characteristics of random networks: the topology of the world-wide web, *Physica A*, **281** (2000), 69–77.
- [16] Barbour, A. D., Poisson convergence and random graphs, *Math. Proc. Cambridge Philos. Soc.*, **92** (1982), 349–359.
- [17] Barbour, A. D., S. Janson, M. Karoński and A. Ruciński, Small cliques in random graphs, *Random Struct. Alg.*, **1** (1990), 403–434.
- [18] Barbour, A. D., M. Karoński and A. Ruciński, A central limit theorem for decomposable random variables with applications to random graphs, *J. Combin. Theory Ser. B*, **47** (1989), 125–145.
- [19] Behrisch, M., A. Coja-Oghlan and M. Kang, The order of the giant component of random hypergraphs, *Random Struct. Alg.*, **36** (2010), 149–184.
- [20] Bhamidi, S., A. Budhiraja and X. Wang, Bounded-size rules: The barely subcritical regime, preprint (2012) arXiv:1212.5480
- [21] Bhamidi, S., A. Budhiraja and X. Wang, The augmented multiplicative coalescent and critical dynamic random graph models, preprint (2012) arXiv:1212.5493
- [22] Bhamidi, S., R. van der Hofstad and G. Hooghiemstra, First passage percolation on random graphs with finite mean degrees, *Ann. Appl. Probab.*, **20** (2010), 1907–1965.
- [23] Bhamidi, S., R. van der Hofstad and J. S. H. van Leeuwen, Scaling limits for critical inhomogeneous random graphs with finite third moments, *Electron. J. Probab.*, **15** (2010), 1682–1703.
- [24] Bohman, T. and A. Frieze, Avoiding a giant component, *Random Struct. Alg.*, **19** (2001), 75–85.
- [25] Bohman, T. and D. Kravitz, Creating a giant component, *Combin. Probab. Comput.*, **15** (2006), 489–511.
- [26] Bollobás, B., A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European J. Combin.*, **1** (1980), 311–316.
- [27] Bollobás, B., Threshold functions for small subgraphs, *Math. Proc. Cambridge Philos. Soc.*, **90** (1981), 197–206.
- [28] Bollobás, B., The evolution of sparse graphs, *Graph theory and combinatorics (Cambridge, 1983)*, Academic Press (1984), 35–57.
- [29] Bollobás, B., The evolution of random graphs, *Trans. Amer. Math. Soc.*, **286** (1984), 257–274.

- [30] Bollobás, B., Paul Erdős and probability theory, Proceedings of the Eighth International Conference “Random Structures and Algorithms” (Poznań, 1997), *Random Struct. Alg.*, **13** (1998), 521–533.
- [31] Bollobás, B., *Modern Graph Theory*, Graduate Texts in Mathematics **184**, Springer-Verlag, New York, 1998, xiv+394 pp.
- [32] Bollobás, B., *Random Graphs*, Second edition, Cambridge Studies in Advanced Mathematics, **73**, Cambridge University Press, Cambridge, 2001, xviii+498 pp.
- [33] Bollobás, B., The Erdős–Rényi theory of random graphs, in *Paul Erdős and his mathematics, II (Budapest, 1999)*, pp. 79–134, Bolyai Soc. Math. Stud., **11**, János Bolyai Math. Soc., Budapest, 2002.
- [34] Bollobás, B., C. Borgs, T. Chayes and O. Riordan, Directed scale-free graphs. Proc. 14th ACM-SIAM Symposium on Discrete Algorithms, 132–139 (2003).
- [35] Bollobás, B., C. Borgs, J. Chayes and O. Riordan, Percolation on dense graph sequences, *Annals of Probability*, **38** (2010), 150–183.
- [36] Bollobás, B. and F. R. K. Chung, The diameter of a cycle plus a random matching, *SIAM J. Discrete Math.*, **1** (1988), 328–333.
- [37] Bollobás, B. and A. M. Frieze, On matchings and Hamiltonian cycles in random graphs, in *Random Graphs '83 (Poznań, 1983)*, pp. 23–46, North-Holland Math. Stud., **118**, North-Holland, Amsterdam, 1985.
- [38] Bollobás, B., S. Janson and O. Riordan, The phase transition in the uniformly grown random graph has infinite order, *Random Struct. Alg.*, **26** (2005), 1–36.
- [39] Bollobás, B., S. Janson and O. Riordan, The phase transition in inhomogeneous random graphs, *Random Struct. Alg.*, **31** (2007), 3–122.
- [40] Bollobás, B., S. Janson and O. Riordan, The cut metric, random graphs and branching processes, *J. Statist. Phys.*, **140** (2010), 289–335.
- [41] Bollobás, B., S. Janson and O. Riordan, Sparse random graphs with clustering, *Random Struct. Alg.*, **38** (2011), 269–323.
- [42] Bollobás, B., J. H. Kim and J. Verstraëte, Regular subgraphs of random graphs, *Random Struct. Alg.*, **29** (2006), 1–13.
- [43] Bollobás, B. and O. Riordan, Mathematical results on scale-free random graphs, in *Handbook of Graphs and Networks*, pp. 1–34, Wiley-VCH, Weinheim, 2003.
- [44] Bollobás, B. and O. Riordan, The diameter of a scale-free random graph, *Combinatorica*, **24** (2004), 5–34.
- [45] Bollobás, B. and O. Riordan, Random graphs and branching processes, in *Handbook of large-scale random networks*, Bolyai Soc. Math. Stud., **18**, B. Bollobás, R. Kozma and D. Miklós eds (2009), pp. 15–115.
- [46] Bollobás, B. and O. Riordan, Metrics for sparse graphs, in *Surveys in Combinatorics 2009*, London Math. Soc. Lecture Note Series, **365**, S. Huczynska, J. D. Mitchell and C. M. Roney-Dougal eds, CUP (2009), pp. 212–287.
- [47] Bollobás, B. and O. Riordan, Asymptotic normality of the size of the giant component via a random walk, *J. Combin. Theory (B)*, **102** (2012), 53–61.
- [48] Bollobás, B. and O. Riordan, Asymptotic normality of the size of the giant component in a random hypergraph, *Random Struct. Alg.*, **41** (2012), 441–450.

- [49] Bollobás, B. and O. Riordan, An old approach to the giant component problem, preprint (2012) arXiv:1209.3691
- [50] Bollobás, B. and O. Riordan, Global to local limit theorems for giant components in hypergraphs, in preparation.
- [51] Bollobás, B., O. Riordan, J. Spencer and G. Tusnády, The degree sequence of a scale-free random graph process, *Random Struct. Alg.*, **18** (2001), 279–290.
- [52] Bollobás, B. and A. G. Thomason, Random graphs of small order, in *Random Graphs '83 (Poznań, 1983)*, North-Holland Math. Studies, **118**, North Holland, Amsterdam, 1985, pp. 47–97.
- [53] Bollobás, B. and A. Thomason, Threshold functions, *Combinatorica*, **7** (1987), 35–38.
- [54] Borgs, C., J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztegombi, Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing, *Advances in Math.*, **219** (2008), 1801–1851.
- [55] Borgs, C., J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztegombi, Convergent sequences of dense graphs II. Multiway cuts and statistical physics, *Ann. of Math. (2)*, **176** (2012), 151–219.
- [56] Bourgain, J., J. Kahn, G. Kalai, Y. Katznelson and N. Linial, The influence of variables in product spaces, *Israel J. Math.*, **77** (1992), 55–64.
- [57] Burtin, Ju. D., Asymptotic estimates of the diameter and the independence and domination numbers of a random graph, *Dokl. Akad. Nauk SSSR*, **209** (1973), 765–768, transl. in *Soviet Math. Dokl.*, **14** (1973), 497–501.
- [58] Burtin, Ju. D., Extremal metric characteristics of a random graph. I, *Teor. Veroyatnost. i Primenen.*, **19** (1974), 740–754.
- [59] Callaway, D. S., J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman and S. H. Strogatz, Are randomly grown graphs really random?, *Phys. Rev. E*, **64** (2001), 041902.
- [60] Chan, S. O. and M. Molloy, $(k + 1)$ -cores have k -factors, *Combin. Probab. Comput.*, **21** (2012), 882–896.
- [61] Chung, F. and L. Lu, The diameter of sparse random graphs, *Adv. Appl. Math.*, **26** (2001), 257–279.
- [62] Chvátal, V., Almost all graphs with $1.44n$ edges are 3-colorable, *Random Struct. Alg.*, **2** (1991), 11–28.
- [63] Coja-Oghlan, A., C. Moore and V. Sanwalani, Counting connected graphs and hypergraphs via the probabilistic method, *Random Struct. Alg.*, **31** (2007), 288–329.
- [64] Cooper, C. and A. Frieze, A general model of web graphs, *Random Struct. Alg.*, **22** (2003), 311–335.
- [65] da Costa, R. A., S. N. Dorogovtsev, A. V. Goltsev and J. F. F. Mendes, Explosive percolation transition is actually continuous, *Phys. Rev. Lett.*, **105** (2010), 255701 (4 pages).
- [66] DeVilleville, R. E. L. and C. S. Peskin, Synchrony and asynchrony for neuronal dynamics defined on complex networks, *Bull. Math. Biol.*, **74** (2012), 769–802.

- [67] Ding, J., J. H. Kim, E. Lubetzky and Y. Peres, Diameters in supercritical random graphs via first-passage percolation, *Combin. Probab. Comput.*, **19** (2010), 729–751.
- [68] Ding, J., J. H. Kim, E. Lubetzky and Y. Peres, Anatomy of a young giant component in the random graph, *Random Struct. Alg.*, **39** (2011), 139–178.
- [69] Dorogovtsev, S. N. and J. F. F. Mendes, Evolution of networks, *Adv. Phys.*, **51** (2002), 1079.
- [70] Dorogovtsev, S. N. and J. F. F. Mendes, *Evolution of networks. From biological nets to the Internet and WWW*. Oxford University Press, Oxford, 2003, x+264 pp.
- [71] Durrett, R., Rigorous result for the CHKNS random graph model, Proceedings, Discrete Random Walks 2003, Cyril Banderier and Christian Krattenthaler, Eds. *Discrete Mathematics and Theoretical Computer Science*, **AC** (2003), 95–104.
- [72] Erdős, L., A. Knowles, H.-T. Yau and J. Yin, Spectral statistics of Erdős–Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues, *Comm. Math. Phys.*, **314** (2012), 587–640.
- [73] Erdős, L. and H.-T. Yau, Universality of local spectral statistics of random matrices, *Bull. Amer. Math. Soc. (N.S.)*, **49** (2012), 377–414.
- [74] Erdős, P. and A. Rényi, On random graphs, I., *Publ. Math. Debrecen*, **6** (1959), 290–297.
- [75] Erdős, P. and A. Rényi, On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **5** (1960), 17–61.
- [76] Erdős, P. and A. Rényi, On the strength of connectedness of a random graph, *Acta Math. Acad. Sci. Hungar.*, **12** (1961), 261–267.
- [77] Erdős, P. and A. Rényi, On the evolution of random graphs, *Bull. Inst. Internat. Statist.*, **38** (1961), 343–347.
- [78] Erdős, P. and A. Rényi, Asymmetric graphs, *Acta Math. Acad. Sci. Hungar.*, **14** (1963), 295–315.
- [79] Erdős, P. and A. Rényi, On random matrices, *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **8** (1964), 455–461.
- [80] Erdős, P. and A. Rényi, On the existence of a factor of degree one of a connected random graph, *Acta Math. Acad. Sci. Hungar.*, **17** (1966), 359–368.
- [81] Erdős, P. and A. Rényi, On random matrices, II., *Studia Sci. Math. Hungar.*, **3** (1968), 459–464.
- [82] Faloutsos, M., P. Faloutsos and C. Faloutsos, On power-law relationships of the internet topology, SIGCOMM 1999, *Comput. Commun. Rev.*, **29** (1999), 251.
- [83] Fernholz, D. and V. Ramachandran, The diameter of sparse random graphs, *Random Struct. Alg.*, **31** (2007), 482–516.
- [84] Fountoulakis, N., Percolation on sparse random graphs with given degree sequence, *Internet Mathematics*, **4** (2007), 329–356.
- [85] Friedgut, E., Sharp thresholds of graph properties and the k -sat problem, with an appendix by Jean Bourgain, *J. Amer. Math. Soc.*, **12** (1999), 1017–1054.
- [86] Friedgut, E. and G. Kalai, Every monotone graph property has a sharp threshold, *Proc. Amer. Math. Soc.*, **124** (1996), 2993–3002.

- [87] Frieze, A., Perfect matchings in random bipartite graphs with minimal degree at least 2, *Random Struct. Alg.*, **26** (2005), 319–358.
- [88] Frieze, A. and B. Pittel, Perfect matchings in random graphs with prescribed minimal degree, in *Mathematics and Computer Science III*, pp. 95–132, Trends Math., Birkhäuser, Basel, 2004.
- [89] Gilbert, E. N., Enumeration of labelled graphs, *Canad. J. Math.*, **8** (1956), 405–411.
- [90] Gilbert, E. N., Random graphs, *Ann. Math. Statist.*, **30** (1959), 1141–1144.
- [91] Goerdt, A., The giant component threshold for random regular graphs with edge faults, *Theoret. Comput. Sci.*, **259** (2001), 307–321.
- [92] Hatami, H. and M. Molloy, The scaling window for a random graph with a given degree sequence, *Random Struct. Alg.*, **41** (2012), 99–123.
- [93] Janson S., D. Knuth, T. Łuczak and B. Pittel, The birth of the giant component, with an introduction by the editors, *Random Struct. Alg.*, **4** (1994), 231–358.
- [94] Janson, S. and M. J. Łuczak, A new approach to the giant component problem, *Random Struct. Alg.*, **34** (2009), 197–216.
- [95] Janson, S., T. Łuczak and A. Ruciński, An exponential bound for the probability of nonexistence of a specified subgraph in a random graph, in *Random Graphs '87* (Poznań, 1987), Wiley, Chichester (1990), 73–87.
- [96] Janson, S., T. Łuczak and A. Ruciński, *Random Graphs*, John Wiley and Sons, New York, 2000, xi+333 pp.
- [97] Janson, S. and O. Riordan, Duality in inhomogeneous random graphs and the cut metric, *Random Struct. Alg.*, **39** (2011), 399–411.
- [98] Janson, S. and O. Riordan, Susceptibility in inhomogeneous random graphs, *Electron. J. Combin.*, **19** (2012), Paper 31, 59 pp.
- [99] Janson, S. and J. Spencer, Phase transitions for modified Erdős–Rényi processes, *Ark. Mat.*, **50** (2012), 305–329.
- [100] Joseph, A., The component sizes of a critical random graph with given degree sequence, preprint (2010), arXiv:1012.2352
- [101] Kahn, J., G. Kalai and N. Linial, The influence of variables on Boolean functions, in *Proc. 29-th Ann. Symp. on Foundations of Comp. Sci.*, pp. 68–80, Computer Society Press, 1988.
- [102] Kalikow, S. and B. Weiss, When are random graphs connected?, *Israel J. Math.*, **62** (1988), 257–268.
- [103] Kang, M., W. Perkins and J. Spencer, The Bohman–Frieze process near criticality, preprint (2011) arXiv:1106.0484
- [104] Kang, M. and T. G. Seierstad, The critical phase for random graphs with a given degree sequence, *Combin. Probab. Comput.*, **17** (2008), 67–86.
- [105] Karoński, M. and T. Łuczak, The phase transition in a random hypergraph, *J. Comput. Appl. Math.*, **142** (2002), 125–135.
- [106] Karoński, M. and A. Ruciński, On the number of strictly balanced subgraphs of a random graph, in *Graph Theory (Lagów, 1981)*, *Lecture Notes in Math.*, **1018**, Springer, Berlin (1983), 79–83.

- [107] Karoński, M. and A. Ruciński, Poisson convergence and semi-induced properties of random graphs, *Math. Proc. Cambridge Philos. Soc.*, **101** (1987), 291–300.
- [108] Karoński, M. and A. Ruciński, The origins of the theory of random graphs, in *The Mathematics of Paul Erdős, I, Algorithms Combin.*, **13** Springer, 1997, pp. 311–336.
- [109] Karp, R. M., The transitive closure of a random digraph, *Random Struct. Alg.*, **1** (1990), 73–93.
- [110] Katona, G., A theorem of finite sets, *Theory of graphs (Proc. Colloq., Tihany, 1966)*, Academic Press, New York, 1968, 187–207.
- [111] Kim, J. H., B. Sudakov and V. H. Vu, On the asymmetry of random regular graphs and random graphs, *Random Struct. Alg.*, **21** (2002), 216–224.
- [112] Krivelevich, M., E. Lubetzky and B. Sudakov, Cores of random graphs are born Hamiltonian, preprint (2013) arXiv:1303.3524.
- [113] Kruskal, J. B., The number of simplices in a complex, in *Mathematical Optimization Techniques*, University of California Press, Berkeley, California, 1963, pp. 251–278.
- [114] Kumar, R., P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins and E. Upfal, Stochastic models for the web graph, FOCS 2000.
- [115] Lotka, A. J., The frequency distribution of scientific productivity, *J. of the Washington Acad. of Sci.*, **16** (1926), 317.
- [116] Lovász, L. and B. Szegedy, Limits of dense graph sequences, *J. Combin. Theory B*, **96** (2006), 933–957.
- [117] Łuczak, M. and T. Łuczak, The phase transition in the cluster-scaled model of a random graph, *Random Struct. Alg.*, **28** (2006), 215–246.
- [118] Łuczak, T., Component behavior near the critical point of the random graph process, *Random Struct. Alg.*, **1** (1990), 287–310.
- [119] Łuczak, T., Size and connectivity of the k -core of a random graph, *Discrete Math.*, **91** (1991), 61–68.
- [120] Łuczak, T., A note on the sharp concentration of the chromatic number of random graphs, *Combinatorica*, **11** (1991), 295–297.
- [121] Łuczak, T., Cycles in a random graph near the critical point, *Random Struct. Alg.*, **2** (1991), 421–439.
- [122] Łuczak, T., Random trees and random graphs, *Random Struct. Alg.*, **13** (1998), 485–500.
- [123] Łuczak, T., B. Pittel and J. C. Wierman, The structure of a random graph at the point of the phase transition, *Trans. Amer. Math. Soc.*, **341** (1994), 721–748.
- [124] Łuczak, T. and J. C. Wierman, The chromatic number of random graphs at the double-jump threshold, *Combinatorica*, **9** (1989), 39–49.
- [125] Martin-Löf, A., Symmetric sampling procedures, general epidemic processes and their threshold limit theorems, *J. Appl. Probab.*, **23** (1986), 265–282.
- [126] Milgram, S., The small world phenomenon, *Psychol. Today*, **2** (1967), 60–67.
- [127] Molloy, M. and B. Reed, A critical point for random graphs with a given degree sequence, *Random Struct. Alg.*, **6** (1995), 161–179.

- [128] Molloy, M. and B. Reed, The size of the giant component of a random graph with a given degree sequence, *Combin. Probab. Comput.*, **7** (1998), 295–305.
- [129] Nachmias, A. and Y. Peres, Component sizes of the random graph outside the scaling window, *ALEA Lat. Am. J. Probab. Math. Stat.*, **3** (2007), 133–142.
- [130] Nachmias, A. and Y. Peres, Critical percolation on random regular graphs, *Random Struct. Alg.*, **36** (2010), 111–148.
- [131] Newman, M. E. J., Random graphs with clustering, *Phys. Rev. Lett.*, **103** (2009), 058701 [4 pages].
- [132] Newman, M. E. J., S. H. Strogatz and D. J. Watts, Random graphs with arbitrary degree distribution and their applications, *Physical Review E*, **64** (2001), 026118.
- [133] Norros, I. and H. Reittu, On a conditionally Poissonian graph process, *Adv. Appl. Probab.*, **38** (2006), 59–75.
- [134] Panagiotou, K., R. Spöhel, A. Steger and H. Thomas, Explosive percolation in Erdős–Rényi-like random graph processes, *Combin. Probab. Comput.*, **22** (2013), 133–145.
- [135] Pittel, B., On tree census and the giant component in sparse random graphs, *Random Struct. Alg.*, **1** (1990), 311–342.
- [136] Pittel, B., Edge percolation on a random regular graph of low degree, *Ann. Probab.*, **36** (2008), 1359–1389.
- [137] Pittel, B., J. Spencer and N. Wormald, Sudden emergence of a giant k -core in a random graph, *J. Combin. Theory Ser. B*, **67** (1996), 111–151.
- [138] Pittel, B. and N. Wormald, Counting connected graphs inside-out, *J. Combinatorial Theory B*, **93** (2005), 127–172.
- [139] Prałat, P., J. Verstraëte and N. Wormald, On the threshold for k -regular subgraphs of random graphs, *Combinatorica*, **31** (2011), 565–581.
- [140] Riddell, R. J. Jr. and G. E. Uhlenbeck, On the theory of virial development of the equation of state of monoatomic gases, *J. Chem. Phys.*, **21** (1953), 2056–2064.
- [141] Riordan, O., The small giant component in scale-free random graphs, *Combin. Probab. Comput.*, **14** (2005), 897–938.
- [142] Riordan, O., The k -core and branching processes, *Combin. Probab. Comput.*, **17** (2008), 111–136.
- [143] Riordan, O., The phase transition in the configuration model, *Combin. Probab. Comput.*, **21** (2012), 265–299.
- [144] Riordan, O. and L. Warnke, Explosive percolation is continuous, *Science*, **333** (2011), 322–324.
- [145] Riordan, O. and L. Warnke, Convergence of Achlioptas processes via differential equations with unique solutions, preprint (2011), arXiv:1111.6179
- [146] Riordan, O. and L. Warnke, Achlioptas process phase transitions are continuous, *Ann. Appl. Probab.*, **22** (2012), 1450–1464.
- [147] Riordan, O. and L. Warnke, Achlioptas processes can be nonconvergent, *Phys. Rev. E*, **86** (2012), 011129 (4 pages).

- [148] Riordan, O. and L. Warnke, The evolution of subcritical Achlioptas processes, preprint (2012), arXiv:1204.5068
- [149] Riordan, O. and N. Wormald, The diameter of sparse random graphs, *Combin. Probab. Comput.*, **19** (2010), 835–926.
- [150] Robinson, R. W. and N. C. Wormald, Almost all cubic graphs are Hamiltonian, *Random Struct. Alg.*, **3** (1992), 117–125.
- [151] Ruciński, A. and A. Vince, Balanced graphs and the problem of subgraphs of random graphs, *Proc. of the 16th Southeastern International conference on Combinatorics, Graph Theory and Computing* (Boca Raton, Fla, 1985), *Congr. Numer.*, **49** (1985), 181–190.
- [152] Ruciński, A. and A. Vince, Strongly balanced graphs and random graphs, *J. Graph Theory*, **10** (1986), 251–264.
- [153] Ruciński, A. and A. Vince, Balanced extensions of graphs and hypergraphs, *Combinatorica*, **8** (1988), 279–291.
- [154] Ruciński, A. and A. Vince, The solution to an extremal problem on balanced extensions of graphs, *J. Graph Theory*, **17** (1993), 417–431.
- [155] Schmidt-Pruzan, J. and E. Shamir, Component structure in the evolution of random hypergraphs, *Combinatorica*, **5** (1985), 81–94.
- [156] Shepp, L. A., Connectedness of certain random graphs, *Israel J. Math.*, **67** (1989), 23–33.
- [157] Söderberg, B., General formalism for inhomogeneous random graphs, *Phys. Rev. E*, **66** (2002), 066121 [6 pages].
- [158] Spencer, J. and N. C. Wormald, Birth control for giants, *Combinatorica*, **27** (2007), 587–628.
- [159] Stepanov, V. E., Phase transitions in random graphs, (in Russian) *Teor. Veroyatnost. i Primenen.*, **15** (1970), 200–216. Translated in *Theory Probab. Appl.*, **15** (1970), 55–67.
- [160] Turova, T. S., Dynamical random graphs with memory, *Phys. Rev. E*, **65** (2002), 066102. Erratum: *Phys. Rev. E*, **70** (2004), 059902(E).
- [161] Turova, T. S., Phase transitions in dynamical random graphs, *J. Statist. Phys.*, **123** (2006), 1007–1032.
- [162] Turova, T. S., Diffusion approximation for the components in critical inhomogeneous random graphs of rank 1, preprint (2009), arXiv:0907.0897
- [163] Turova, T. S., The largest component in subcritical inhomogeneous random graphs, *Combin. Probab. Comput.*, **20** (2011), 131–154.
- [164] Watts, D. J., *Small worlds. The dynamics of networks between order and randomness*. Princeton Studies in Complexity. Princeton University Press, Princeton, NJ, 1999. xvi+262 pp.
- [165] Watts, D. J., *Six degrees. The science of a connected age*. W. W. Norton & Co. Inc., New York, 2003. 368 pp.
- [166] Watts, D. J. and S. H. Strogatz, Collective dynamics of ‘small-world’ networks, *Nature*, **393** (1998), 440–442.

- [167] Wormald, N. C., The differential equation method for random graph processes and greedy algorithms, in *Lectures on approximation and randomized algorithms*, pages 73–155. PWN, Warsaw, 1999.

Béla Bollobás

*Department of Pure Mathematics and
Mathematical Statistics,*

Wilberforce Road,

Cambridge CB3 0WB, UK,

*Department of Mathematical
Sciences,*

University of Memphis,

Memphis TN 38152, USA,

and

*London Institute for Mathematical
Sciences,*

35a South St,

Mayfair, London W1K 2XF, UK

e-mail: `bb12@cam.ac.uk`

Oliver Riordan

*Mathematical Institute,
University of Oxford,*

24–29 St Giles’,

Oxford OX1 3LB, UK

e-mail: `riordan@maths.ox.ac.uk`

AROUND THE SUM-PRODUCT PHENOMENON

JEAN BOURGAIN*

(Dedicated to P. Erdős)

1. INTRODUCTION

The purpose of this exposé is to give a sample of P. Erdős many contributions to combinatorics that turned out to be unexpectedly seminal. He was indeed a master in recognizing seemingly elementary questions which require new insights, often with far reaching consequences. The results discussed below originate from his papers “Problems and results in combinatorial number theory, III” ([32]), “On sums and products of integers”, ([33]), jointly with Szemerédi, and “Additive Gruppen mit vorgegebener Hausdorffscher Dimension” ([34]), jointly with Volkmann. These papers led to numerous developments over the past decade and influenced other parts of mathematics, including number theory, theoretical computer science, ergodic theory and group theory. Giving a fair account of them would be a considerable task and we limit ourselves to citing just a few. The choice only reflects the author’s interests and research; many related contributions and contributors will not be cited and the bibliography strictly serves this presentation.

2. SOME COMMENTS ON THE ORIGINAL PROBLEMS AND THEIR STATUS

If A is a finite subset of \mathbb{Z} , Erdős conjectured in [32] that for all $\varepsilon > 0$

$$(2.1) \quad |A + A| + |A \cdot A| > c_\varepsilon |A|^{2-\varepsilon}.$$

*The research was partially supported by NSF grants DMS-0808042 and DMS-0835373.

It is proved in [33] (among other things) that

$$(2.2) \quad |A + A| + |A.A| > c|A|^{1+c}$$

for some $c > 0$; the best result to date is due to Solymosi [59], with

$$(2.3) \quad |A + A| + |A.A| > \frac{1}{2}|A|^{4/3}(\log |A|)^{-\frac{1}{3}}.$$

More generally, related to h -fold sumsets hA and product sets $A^{(h)}$, [32] puts forward the problem of showing that for A as above

$$(2.4) \quad |hA| + |A^{(h)}| > c_{h,\varepsilon}|A|^{h-\varepsilon}.$$

These problems may be considered for finite subsets A of \mathbb{R} as well (cf. [33]) and (2.3) is equally valid in this setting. Note that the statement (2.1) is carefully formulated, since (cf. [33])

$$(2.5) \quad \min_{|A|=k} (|A + A| + |A.A|) \ll k^{2-\frac{c}{\log \log k}}.$$

Problem 2.1 remains widely open and, if correct, is likely a deep statement. Some hints of this are provided by the treatment of certain special cases, that rely on methods from algebraic number theory; see [27], [28]. We cite two such results.

Let $A \subset \mathbb{R}$ or $A \subset \mathbb{C}$ be a finite set. Assume $|A + A| < K|A|$ where, for simplicity, we view K as a fixed large constant. Using Freiman’s theorem and divisor theory in number fields, it is shown in [27] that $|A^{(h)}| > c_{h,\varepsilon}|A|^{h-\varepsilon}$ for $h = 2, 3, \dots$. Conversely, if $|A.A| < K|A|$, the work of [35] on additive relations in multiplicative subgroups of \mathbb{C}^* (based on extensions of the ‘subspace theorem’) permits to deduce that $|hA| > c_h|A|^h$ for $h = 2, 3, \dots$, see [28].

It was also proven in [10] that

$$(2.6) \quad |hA| + |A^{(h)}| > |A|^{c(h)}$$

with $c(h) \rightarrow \infty$ for $h \rightarrow \infty$, provided we assume $A \subset \mathbb{Z}$.

Statement (2.6) was generalized to sets A consisting of algebraic numbers of bounded degree, [11] but so far remains unsettled for $A \subset \mathbb{R}$.

In [33], a further rather fascinating generalization of (2.1) is suggested, where one considers sumsets and product sets restricted to a graph $G \subset A \times A$. Thus

$$A +_G A = \{x + y; (x, y) \in G\}$$

and $A \times A$ is defined similarly. The reader is referred to [1] for an extensive discussion and recent developments around this question.

Let us turn next to the ‘continuous’ counterpart of the sum-product problem. One version is embodied in the following conjecture going back to [34]:

“A measurable subring R of \mathbb{R} (in the algebraic sense) is either of zero Hausdorff dimension or $R = \mathbb{R}$ ”.

It was proven in [36] that $R = \mathbb{R}$ if $\dim_H R > \frac{1}{2}$. The question remained somewhat dormant until in [47] its relevance (more precisely, the discretized version formulated in terms of box-dimension) to various other issues, such as Falconer’s distance problem (a dimensional version of Erdős’ distance problem) and questions raised by Furstenberg, was noted; these were in fact largely motivated by attempts to progress on the 3-dimensional Kakeya set problem, still unsolved to date. On the other hand, the Erdős–Volkman ring problem got settled in an elegant paper by Edgar and Miller [31]. While it did not capture the problems discussed in [47], that have to do with box-dimension, it is the starting point of the ‘sum-product theory’ in finite fields which had its own rather remarkable impact. Returning to the [47] paper, the so-called ‘discretized ring theorem’ was established in [3] and underlies a different set of developments.

3. SUM-PRODUCT IN FINITE FIELDS

The basic philosophy of the sum-product theorem is that either the sum set $A + A = \{x + y | x \in A, y \in A\}$ or the product set $A \cdot A = \{x \cdot y : x, y \in A\}$ will be substantially “larger” than A , putting aside obvious obstructions of algebraic (or metrical) nature.

The following result was proven in [21] and in a more precise form in [20]. See also [61].

Theorem 3.1 ([21] and [20]). *For all $\varepsilon > 0$, there is $\delta > 0$ such that if $A \subset \mathbb{F}_p$ and $|A| < p^{1-\varepsilon}$, then*

$$|A + A| + |A \cdot A| > c|A|^{1+\delta}$$

where $c > 0$ is an absolute constant.

To be pointed out that formulations with explicit exponents have been obtained, but we will not discuss them here.

If we try to generalize Theorem 3.1 to arbitrary finite fields, there is the obvious obstruction of nontrivial subfields. As is clear from the next result, this is the only one.

Theorem 3.2 ([21]). *Assume $S \subset \mathbb{F}_q$ and $|S| > q^\delta$ where $\delta > 0$ is arbitrary and*

$$|S + S| + |S \cdot S| < K|S|.$$

Then there is a subfield G of \mathbb{F}_q and $\xi \in \mathbb{F}_q^$ such that*

$$|G| < K^C|S| \text{ and } |S \setminus \xi G| < K^C,$$

where $C = C(\delta)$.

Further generalizations (with an appropriate formulation) to Cartesian products $\mathbb{F}_p \times \mathbb{F}_p$, residue rings $\mathbb{Z}/q\mathbb{Z}$ and, more generally, O/I with I an ideal in the integers O of a number field, followed. See in particular [9].

While their formulation is unavoidably more technical, they rigorously conform with the ‘philosophy’ stated above. These results have equally significant consequences, some of which are discussed below (see also the survey paper of M. Garaev [42]).

4. EXPONENTIAL SUMS: BEYOND WEIL AND STEPANOV

A first significant application of the results of Section 3 is to the theory of exponential sums over finite fields, leading to nontrivial results in situations where classical methods do not seem to apply. The first progress obtained along these lines appear in [22] and [20].

Theorem 4.1. *For all $\varepsilon > 0$, there is $\delta > 0$ such that if H is a multiplicative subgroup of \mathbb{F}_p^* ($H < \mathbb{F}_p^*$ for short) and $|H| > p^\varepsilon$, then*

$$(4.2) \quad \max_{(a,p)=1} \left| \sum_{x \in H} e_p(ax) \right| < cp^{-\delta}|H|.$$

Earlier results cover the range up to $\varepsilon > \frac{1}{4}$; see [44], [49]. The technique used in those papers are variants of Stepanov’s method.

Nontrivial bounds of the form

$$(4.3) \quad \max_{(a,p)=1} \left| \sum_{x \in H} e_p(ax) \right| = o(|H|)$$

may be obtained provided $\log |H| > C \frac{\log p}{\log \log p}$, for some constant C . This seems to be the limitation of our method. It is a challenging problem to obtain estimates below this threshold.

Following [53], one can state

Problem 1. Is (4.3) valid under the assumption that $\frac{|H|}{\log p} \rightarrow \infty$?

H. Furstenberg’s famous $\times 2, \times 3$ problem for invariant measures on the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ naturally leads to the following question.

Problem 2. Does (4.3) hold if we let $H < \mathbb{F}_p^*$ be the multiplicative group generated by 2 and 3?

Theorem 4.1 is of course equivalent to the following formulation for Gauss sums.

Corollary 4.4. *For all $\delta > 0$, there is $\delta' > 0$ such that if $(k, p - 1) < p^{1-\delta}$, then*

$$(4.5) \quad \max_{(a,p)=1} \left| \sum_{x=1}^p e_p(ax^k) \right| < cp^{1-\delta'}.$$

Note that Gauss classical bound by $(k, p - 1)\sqrt{p}$ is trivial if $(k, p - 1) \geq \sqrt{p}$.

More generally, one has Weil’s inequality for $f(x) \in \mathbb{F}_p[X]$ of degree d , namely

$$(4.6) \quad \left| \sum_{1 \leq x \leq p} e_p(f(x)) \right| \leq d\sqrt{p}.$$

This inequality is again trivial for $d \geq \sqrt{p}$. Obtaining nontrivial exponential sum bounds for general polynomials when $d \geq \sqrt{p}$ is a major open problem.

For certain applications, as described for instance in the book [50], it is useful to have the following version of Theorem 4.1 for ‘almost groups’.

Theorem 4.7. *For all $\delta > 0$, there is $\delta' > 0$ such that if $\theta \in \mathbb{Z}_+$ satisfies*

$$(4.8) \quad (\theta, p) = 1 \quad \text{and} \quad \mathcal{O}_p(\theta) \geq t > p^\delta$$

where we denote $\mathcal{O}_p(\theta)$ the multiplicative order of $\theta \pmod p$, then

$$(4.9) \quad \max_{(a,p)=1} \left| \sum_{s=1}^t e_p(a\theta^s) \right| < tp^{-\delta'}.$$

This type of statement is obviously relevant to the distributional properties of the linear congruential pseudo-random number generator for instance.

A similar result, with the appropriate necessary assumptions, may be obtained for arbitrary finite fields \mathbb{F}_q .

Let $q = p^m$ and denote for $x \in \mathbb{F}_q$ the trace

$$Tr(x) = x + x^p + \dots + x^{p^{m-1}}.$$

Let $\psi(x) = e_p(Tr(x))$ be the additive character.

Theorem 4.10 [12]. *Let $\theta \in \mathbb{F}_q^*$ be of order t and let $t \geq t_1 > q^\varepsilon$. Assume*

$$\max_{\substack{1 \leq \nu < m \\ \nu | m}} (p^\nu - 1, t) < q^{-\varepsilon} t$$

where $\varepsilon > 0$ is arbitrary and fixed. Then

$$\max_{a \in \mathbb{F}_q^*} \left| \sum_{j \leq t_1} \psi(ag^j) \right| < Cq^{-\delta} t_1,$$

where $\delta = \delta(\varepsilon) > 0$.

Already for $\varepsilon = \frac{1}{2}$, the result seems new (there does not seem a version of Stepanov’s method available beyond prime fields).

Theorem 4.1 has an extension to subgroups of the unit group $(\mathbb{Z}/q\mathbb{Z})^*$ of the ring $\mathbb{Z}/q\mathbb{Z}$ of residues modulo q , see [6].

Theorem 4.11. *Let q be an arbitrary modulus. For all $\varepsilon > 0$, there is $\delta = \delta(\varepsilon)$ such that if $H < \mathbb{Z}_q^*$ satisfies*

$$(4.12) \quad |H| > q^\varepsilon,$$

then

$$(4.13) \quad \max_{\xi \in \mathbb{Z}_q^*} \left| \sum_{x \in H} e_q(\xi x) \right| < q^{-\delta} |H|.$$

Note that the statement in Theorem 4.11 is uniform in the modulus q , which may be highly composite. In this setting, only the case $\varepsilon > \frac{1}{2}$ was previously known, as a consequence of Gauss’ estimate.

Let us cite one more generalization.

Let R be a finite commutative ring with unit and assume $|R| = q$ where q has no small prime divisors (hence Theorem 4.14 below does not cover Theorem 4.11). Denote R^* the group of invertible elements of R . The following trichotomy holds, see [5].

Theorem 4.14. *Let $H < R^*$ and $|H| > q^\delta$, where δ is arbitrarily and fixed. For all $\varepsilon > 0$, there is $\varepsilon' = \varepsilon'(\varepsilon) \rightarrow 0$, as $\varepsilon \rightarrow 0$, such that one of the following alternatives holds:*

(i) *We have*

$$(4.15) \quad \max_{\chi \neq \chi_0} \left| \sum_{x \in H} \chi(x) \right| < |H|^{1-\varepsilon},$$

where χ refers to the additive characters of R .

(ii) *There is a nontrivial ideal I in R with*

$$(4.16) \quad |H \cap (1 + I)| > |H|^{1-\varepsilon'}.$$

(iii) *There is a nontrivial subring R_1 of R such that*

$$(4.17) \quad |H \cap R_1| > |H|^{1-\varepsilon'}.$$

A word of explanation about the relation between the results in Section 3, which are purely set-theoretical, and those in Section 4 that depend on bounds on additive and multiplicative energy

$$(4.18) \quad E_+(A, B) = |\{(a, a', b, b') \in A^2 \times B^2; a + b = a' + b'\}|$$

$$(4.19) \quad E_\times(A, B) = |\{(a, a', b, b') \in A^2 \times B^2; ab = a'b'\}|.$$

The link is provided by the Balog-Szemerédi-Gowers theorem (cf. [61]), which is basically a general result from graph theory. It was originally proven by Balog and Szemerédi, using Szemerédi’s uniformity lemma and the estimates were quantitatively poor. A simpler argument with better bounds was obtained more recently by Gowers and is an essential ingredient in his work on arithmetic progressions.

Combining Theorem 3.1 with the [17] result, one establishes multi-linear exponential sum bounds of the following type

Theorem 4.20. *Given $\varepsilon > 0$, there is $\delta = \delta(\varepsilon)$ such that for arbitrary sets $A_1, \dots, A_k \subset \mathbb{F}_p$ satisfying $|A_j| > p^\varepsilon$ for $1 \leq j \leq k$ and $|A_1| \cdots |A_k| > p^{1+\varepsilon}$,*

$$(4.21) \quad \max_{a \in \mathbb{F}_p^*} \left| \sum_{x_1 \in A_1} \cdots \sum_{x_k \in A_k} e_p(ax_1 \dots x_k) \right| < p^{-\delta} |A_1| \cdots |A_k|.$$

Observe that when $k = 2$ the statement is elementary and well-known. Indeed one has

$$(4.22) \quad \max_{a \in \mathbb{F}_p^*} \left| \sum_{x \in A} \sum_{y \in B} e_p(axy) \right| \leq (p|A||B|)^{1/2}.$$

Also, the condition $|A_1| \dots |A_k| > p^{1+\varepsilon}$ is essentially optimal.

Apart from bounding exponential sums over multiplicative subgroups, Theorem 4.20 has other number theoretic applications that are worth to be mentioned. We cite a few consequences to incomplete Kloosterman sums obtained in [19]. The following statement relates to results from [52] but the assumptions are much less restrictive.

Theorem 4.23. *There are an absolute constant $C > 0$ such that for any positive integer n , arbitrary intervals $I_1, \dots, I_n \subset \mathbb{F}_p$ of size $N > p^{C/n^2}$ and arbitrary subsets $A_1 \subset I_1, \dots, A_n \subset I_n$, one has the estimate*

$$(4.24) \quad \max_{(a,p)=1} \left| \sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} e_p(ax_1^* \cdots x_n^*) \right| < p^{-\delta} N^n$$

with x^* the multiplicative inverse of $x \pmod{p}$. Here $\delta = \delta(n) > 0$.

Remark. In fact, one may take $C = 4$ in the above statement.

Bounds on multi-linear Kloosterman sums play a role in obtaining good remainder estimates in sieving theory; cf. [37]. Denote

$$(4.25) \quad \pi(x; q, a) = \{p < x; p \equiv a \pmod{q}\}.$$

The following result from [19] improves the Brunn–Titchmarsh theorem derived in [38] (see Ch. 13).

Theorem 4.26. *Let $q = x^\theta$, where $\theta < 1$ is close to 1. Then, for sufficiently large x*

$$(4.27) \quad \pi(x; q, a) < \frac{cx}{\phi(q) \log \frac{x}{q}}$$

with $c = 2 - c_1(1 - \theta)^2$, for some absolute constant $c_1 > 0$

Other applications of sum-product theory appear in relation to incomplete character sums. This is perhaps not surprising, taking into consideration the amplification step, based on Vinogradov's shifted product argument, in the proof of Burgess' theorem. Depending on the problem, geometry of numbers or sum-product techniques may be more effective here.

5. GROWTH AND EXPANSION IN SEMI-SIMPLE GROUPS

It turned out that sum-product theorems in finite fields lead to product theorems in semi-simple Lie groups. A first breakthrough result in this direction was obtained by Helfgott.

Theorem 5.1 ([45]). *Let $A \subset SL_2(\mathbb{F}_p)$, $|A| < p^{3-\delta}$ and assume A is not contained in any proper subgroup of $SL_2(\mathbb{F}_p)$. Then*

$$|A \cdot A \cdot A| > c|A|^{1+\varepsilon},$$

with $c, \varepsilon > 0$ only depending on δ .

Theorem 5.1 is a key ingredient in the proof of the following result on expansion in $SL_2(\mathbb{F}_p)$ Cayley graphs.

Theorem 5.2 ([14]). *Let S be a symmetric generating subset of $SL_2(\mathbb{F}_p)$ satisfying the girth condition*

$$\text{girth}(\mathcal{G}(SL_2(p), S)) > \rho \log p,$$

where $\rho > 0$ is an arbitrary fixed constant. Then the Busemann-Cheeger expansion coefficient $c(\mathcal{G})$ satisfies

$$c(\mathcal{G}) > c(\rho) > 0.$$

Recall that the ‘girth’ is the size of the smallest Hamiltonian cycle. The expansion coefficient is $\min \frac{|\partial A|}{|A|}$ for A a subset of the vertex set V , $|A| \leq \frac{1}{2}|V|$ and ∂A refers to the edges joining A and $V \setminus A$.

Theorem 5.2 relates to Theorem 5.1 the way Theorem 4.1 relates to Theorem 3.1 and again the Balog–Szemerédi–Gowers result is involved in deriving one from the other. Taking $S \subset SL_2(\mathbb{Z})$ generating a free group, the Cayley graphs

$$\{\mathcal{G}(SL_2(p), \pi_p(S)) : p \geq p_0(S)\}$$

from an expander family, according to Theorem 5.2 and the strong approximation property.

Problem 3 (A. Lubotzky). Is there for given $k \geq 2$ an absolute constant $\rho > 0$ such that the expansion coefficient $c(\mathcal{G}(SL_2(p)), S) > \rho$, whenever $S \subset SL_2(p)$, $|S| = k$ and S generates $SL_2(p)$?

The work of Breuillard and Gamburd [24] comes tentatively close to a positive answer.

Theorem 5.1 and 5.2 have been vastly generalized, leading to a powerful and fairly complete theory. In particular, Theorem 5.1 was extended by Helfgott and subsequently, in the works of Pyber–Szabó [54] and Breuillard–Green–Tao [25]. Similar results in $SL_d(\mathbb{Z}/q\mathbb{Z})$ were obtained by Gamburd, Sarnak and the author and, more recently, by P. Varju. In terms of group expansion, we cite the following extensions of Theorem 5.2.

Theorem 5.3 [57]. *Let $\Gamma \leq SL_n(\mathbb{Q})$ be a finitely generated group with a symmetric generating set S . Then the congruence graphs $(\pi_q(\Gamma), S)$ for q squarefree and coprime to a finite set of primes (depending on Γ), are an expander family, provided G^0 the identity component of the Zariski closure G of Γ , is perfect, i.e. $[G^o, G^o] = G^o$.*

Theorem 5.4 ([23]). *Let $S \subset SL_d(\mathbb{Z})$ be finite and symmetric. Assume that S generates a subgroup of $\Gamma < SL_d(\mathbb{Z})$, which is Zariski dense in SL_d . Then the Cayley graphs $\mathcal{G}(\pi_q(\Gamma), \pi_q(S))$ with q running over the integers, forms an expander family. Moreover, there is an integer q_0 such that $\pi_q(\Gamma) = SL_d(\mathbb{Z}/q\mathbb{Z})$ if q is coprime to q_0 .*

Note that the last part of the statement in Theorem 5.4 is just the strong approximation property. In particular, setting $d = 2$, Theorem 5.4 provides the full generalization of Selberg’s theorem to non-elementary subgroups of $SL_2(\mathbb{Z})$.

Part of the the motivation for this research comes from number theory and diophantine problems related to group actions. A first line of applications has to do with prime number sieving in the orbits of ‘thin’ groups, a popularized example being the curvatures in integral Apollonian circle packings (cf. [17] and the Bourbaki exposé by E. Kowalski [51]). Further applications (as discussed in [4] and in A. Kontorovich’ expository paper [48]) emerged with the elaboration of a version of the Hardy–Littlewood circle method in the study of global properties of group (or semi-group) orbits. The major arcs analysis in the application of the circle method requires indeed precise counting asymptotics in the Archimedian balls of congruence subgroups. Those are provided by Lax–Phillips theory or thermo-dynamical methods applied to thin groups and the spectral input is provided by the expansion theory discussed above. (See [18]). Note that an implementation in the circle method requires expansion with unrestricted modulus (as stated in Theorem 5.4) while for most sieving applications, it suffices to consider square-free moduli.

Returning to the combinatorial aspect of Theorem 5.1, an important notion (introduced by Tao in [60]) is that of an ‘approximate group’. (See

also [61]). Roughly speaking, a subset A of a group G is called a K -approximate group, provided $A = A^{-1}$ and there is a subset $X \subset G, |X| < K$ such that $X = X^{-1}$ and $A.A \subset A.X$. A natural line of research is then to explore the arithmetic structure of K -approximate groups, in various ranges of K (the quantitative aspects are essential in applications). For instance, Freiman’s classical theorem states that if $A \subset \mathbb{Z}$ is a finite K -approximate group, say with K a fixed constant, then A is commensurable with a generalized arithmetic progression – the ultimate quantitative version in terms of dependence on K , known as the polynomial Freiman-Ruzsa conjecture being still unsettled at the time of this writing (see [56] for the strongest results in this direction). On the other hand, Theorem 5.1 implies that if $A \subset SL_2(p)$ is a generating K -approximate group, then either $|A| < K^C$ or $|A| > K^{-C}|SL_2(p)|$ for some absolute constant C .

Note that while Helfgott’s approach was based on the scalar sum-product theory, the subsequent developments in [54], [25] rely solely on the ambient group structure. The key combinatorial insights in these works indeed are purely group theoretical. They provide a quantitative version of earlier work due to E. Hrushovski [46].

The reader may wish to consult B. Green’s survey paper [43] on approximate groups for an introduction; but it does not discuss the work in [54] and [25] that came slightly later.

Finally, a complete structural description of K -approximate groups in the general setting (for fixed K) is to be found in [26].

6. THE DISCRETIZED RING THEOREM

While in [34] Hausdorff dimension is considered, it turns out that for many applications (starting from the ones discussed in [47]), it is rather a sum-product principle for box-dimension that is useful. Let us point out that the results presented in this section also have p -adic versions. Another comment is that while it is standard to derive statements involving Hausdorff dimension from their counterpart for box-dimension, the other way around is by no means automatic and may require different methods.

Theorem 6.1 ([3]). *For all $0 < \sigma < 1$ and $\kappa > 0$, there is $\varepsilon = \varepsilon(\sigma, \kappa) > 0$ such that if $A \subset [0, 1]$ is a union of δ -intervals, where $\delta > 0$ is small, satisfying*

$$|A| = \delta^{1-\sigma},$$

and for all $\delta < \rho < \delta^\varepsilon$,

$$(6.2) \quad \max_t |A \cap B(t, \rho)| < \rho^k |A|,$$

then

$$|A + A| + |A \cdot A| > \delta^{1-\sigma-\varepsilon}.$$

Remark. A “non-concentration” assumption such as (6.2) is easily seen to be necessary for such a statement to hold.

From the above result, one may deduce new Marstrand-type projection theorems;

Their original motivation lies in the work [13] discussed in the next section. Our first statement uses box-dimension; the next one is in terms of Hausdorff dimension (see [8]).

Theorem 6.3. *Given $0 < \alpha < 2, \beta > 0$ and $\kappa > 0$, there exist $\tau_0 > 0$ and $\eta > \alpha/2$ such that the following holds.*

Let μ_1 be a probability measure on S^1 such that

$$(6.4) \quad \max_{\theta} \mu_1([\theta - \rho, \theta + \rho]) < C\rho^{\kappa}.$$

Let $\delta > 0$ be chosen sufficiently small and let $\mathcal{A} \subset [1, 2] \times [1, 2]$ be a union of size- δ squares satisfying

$$(6.5) \quad |\mathcal{A}| = \delta^{2-\alpha}$$

and

$$(6.6) \quad \max_x |\mathcal{A} \cap \mathcal{B}(x, \rho)| < \rho^{\beta} |\mathcal{A}| \quad \text{for } \delta < \rho < \delta^{\tau_0}.$$

Then there exists $\theta \in \text{supp } \mu_1$ such that

$$(6.7) \quad |\pi_{\theta}(\mathcal{A})| > \delta^{1-\eta},$$

where π_{θ} denotes the orthogonal projection on the line $y = (tg\theta)x$.

Stated in terms of Hausdorff-dimension, one has the following.

Theorem 6.8. *Given $0 < \alpha < 2$ and $\kappa > 0$, there exists $\eta > \alpha/2$ such that, if $\mathcal{A} \subset \mathbb{R} \times \mathbb{R}$ is a set of Hausdorff dimension*

$$(6.9) \quad \text{H-dim } \mathcal{A} > \alpha.$$

then

$$(6.10) \quad \text{H-dim } \pi_{\theta}(\mathcal{A}) \geq \eta$$

for all $\theta \in S^1$ except in an exceptional set E satisfying

$$(6.11) \quad \text{H-dim } E \leq \kappa.$$

Similar results hold for 1-dimensional projections in an ambient space of arbitrary dimension $d \geq 2$. It is also possible to generalize to higher dimensional projections, with the proper assumptions, though this is still work in progress. There is a version of Theorem 6.1 for subsets of the complex numbers, but more technical to formulate. We rather state a consequence, used in [16] for instance. Let $d \geq 1$ and \mathbb{C}^d be equipped with its product ring structure.

Theorem 6.12. *Given $\sigma > 0$, there is a constant $C(d, \sigma)$ such that the following holds. Let $\delta > 0$ be sufficiently small and $A \subset \mathbb{C}^d \cap B(0, 1)$ containing at least $\delta^{-\sigma}$ points that are δ -separated. Then there is a unit vector $\xi \in \mathbb{C}^d$ and $0 \leq \gamma < C(d, \sigma)$ such that for some positive integers $s < C(d, \sigma)$*

$$(6.13) \quad [0, \delta^\gamma] \xi \subset sA^{(s)} - sA^{(s)} + B(0, \delta^{\gamma+1})$$

where $sA^{(s)}$ denotes the s -fold sumset of the s -fold product set $A^{(s)}$.

7. SPECTRAL GAPS IN LIE-GROUPS AND RANDOM MATRIX PRODUCTS

The following result conjectured in [41] may be seen as the $SU(2)$ counterpart of Theorem 5.2. Its proof uses Theorem 6.1 which replaces Theorem 3.1 in the continuous setting

First we need to recall the non-Abelian diophantine condition from [41] that will play the role of the “large girth” assumption.

Definition 7.1. For $k \geq 2$, we say that the set of elements $g_1, \dots, g_k \in SU(2)$ are diophantine if there is $D > 0$ such that for any $m \geq 1$ and any word R_m in g_1, \dots, g_k of length m , and such that $R_m \neq \pm e$, we have

$$\|R_m \pm e\| \geq D^{-m}.$$

Example. Take $g_1, \dots, g_k \in SU(2) \cap M_2(\overline{\mathbb{Q}})$ generating a free group.

Theorem 7.2. *Let $\{g_1, \dots, g_k\}$ be a set of elements in $SU(2)$ generating a free group and satisfying a diophantine property. Then*

$$z_{g_1, \dots, g_k} = g_1 + g_1^{-1} + \dots + g_k + g_k^{-1}$$

has a spectral gap, with lower bound depending on k and D only.

Corollary 7.3. *If $g_1, \dots, g_k \in SU(2) \cap M_2(\bar{\mathbb{Q}})$ generate a free group, then z_{g_1, \dots, g_k} has a spectral gap.*

This result is proven in [15], combining Theorem 6.1 with the scheme elaborated by Helfgott to prove his product theorem in $SL_2(p)$. A different approach exploiting the Lie-algebra appears in [16], where Theorem 7.2 is generated as follows.

Theorem 7.4. *Corollary 7.3 generalizes to systems $g_1, \dots, g_k \in SU(d) \cap M_d(\bar{\mathbb{Q}})$ generating a subgroup of $SU(d)$ which is topologically dense.*

Note that the assumption in Theorem 7.4 is equivalent to $\langle g_1, \dots, g_k \rangle$ being Zariski dense in $SL_d(\mathbb{C})$.

Also the method used to prove Theorem 7.4 is likely to generalize to other Lie groups.

The above results have several applications, including to the stochastic tilings, studied in [29], [30], [55] for instance, and to quantum computation (ε -approximation by words of length $\sim \log \frac{1}{\varepsilon}$ in ‘fault tolerant’ gates, improving on the Solovay-Kitaev algorithm.)

Compared with classical methods, one is now able to establish spectral gaps when the group $\langle g_1, \dots, g_k \rangle$ is not arithmetic, as may occur in the applications above. See also [58] for a discussion around ‘thin groups’.

Theorem 6.3 is an essential ingredient in the work [13] on equidistribution for toral actions of linear groups. Let $S = \{g_1, \dots, g_k\}$ be elements in $SL_d(\mathbb{Z})$ generating a semi-group Γ_+ which action on \mathbb{R}^d is strongly irreducible (no finite union of proper linear subspaces is invariant under Γ_+) and proximal (Γ_+ contains an element with a single largest eigenvalue.) (These assumptions are satisfied if $\langle S \rangle$ is Zariski dense in SL_d for instance). Denote

$$(7.5) \quad \nu = \frac{1}{|S|} \sum_{g \in S} \delta_g$$

which is a finitely supported measure on $SL_d(\mathbb{Z})$.

The following statement answering an equidistribution problem due to Guivarch in a quantitative way, is the main result from [13].

Theorem 7.6. *Given ν as above, there are constants $c > 0$ and $C < \infty$ such that if $\theta \in \mathbb{T}^d \setminus \{0\}$, $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, and $b \in \mathbb{Z}^d \setminus \{0\}$, $\|b\| < e^{cn}$, then*

$$(7.7) \quad (*) = \left| \sum_g \nu^{(n)}(g) e^{2\pi i \langle b, g\theta \rangle} \right| < e^{-cn}$$

unless $\|\theta - \frac{a}{q}\| < e^{-cn}$ with $q < e^{\frac{c}{4}n}$, in which case

$$(7.8) \quad (*) < \frac{|b|^C}{q^c}.$$

A probability measure μ on \mathbb{T}^d is called ν -stationary if

$$(7.9) \quad \mu * \nu \equiv \sum_g \nu(g)g_*[\mu].$$

Corollary 7.10. *With ν as in Theorem 7.6, any ν -stationary measure μ on \mathbb{T}^d is a combination of Haar measure and an atomic measure supported by rational points and μ is $\langle \nu \rangle$ -invariant.*

The last part of the above statement answers Furstenberg’s ‘stiffness conjecture’ [40]. Subsequent work [2] due to Y. Benoist and J. Quint has vastly generalized Corollary 7.10, but their equidistribution results are not quantitative. We observe for instance that Theorem 7.6 in its full strength is essential in the proof of Theorem 5.4 above.

Returning to $SU(2)$, we conclude this section with the following problem that in some sense is the analogue of Problem 3 above.

Problem 4. Do Corollary 7.3 and Theorem 7.4 hold without assuming that $g_1, \dots, g_k \in M_d(\mathbb{Q})$?

REFERENCES

- [1] N. Alon, O. Angel, I. Benjamini, E. Lubetzky, *Sums and products along sparse graphs*, Israel J. Math **188** (2012), 353–384.
- [2] Y. Benoist, J. Quint, *Measures stationnaires et fermés invariants des espaces homogènes*, Ann. of Math. (2) **174** (2011), no 2, 1111-1162.
- [3] J. Bourgain, *On the Erdős-Volkmann and Katz-Tao ring conjecture*, Geom. Funct. Anal. **13** (2003), 334–365.
- [4] J. Bourgain, *Some diophantine applications of the theory of group expansion*, to appear in MSRI publications.
- [5] J. Bourgain, *Exponential sum estimates in finite commutative rings and applications*, J. Anal. Math. **101** (2007), 325–355.
- [6] J. Bourgain, *Exponential sum estimates over subgroups of \mathbb{Z}_q^* , q arbitrary*, J. Anal. Math. **97** (2005), 317–355.
- [7] J. Bourgain, *Multilinear exponential sums in prime fields under optimal entropy condition on the sources*, Geom. Funct. Anal. **18** (2009), no 5, 1477–1502.

- [8] J. Bourgain, *The discretized ring and projection theorems*, J. Analyse, Vol. **112** (2010), 193–236.
- [9] J. Bourgain, *The sum-product theorem in \mathbb{Z}_q with q arbitrary*, J. Analyse Math. **106** (2008), 1–93.
- [10] J. Bourgain, M. Chang, *On the size of k -fold sum and product sets of integers*, J. AMS **17** (2004), no 2, 473–497.
- [11] J. Bourgain, M. Chang, *Sum-product theorems in algebraic number fields*, J. Anal. Math. **109** (2009), 253–277.
- [12] J. Bourgain, M. Chang, *A Gauss sum estimate in arbitrary finite fields*, C.R. Math. Acad. Sci. Paris **342** (2006), 643–646.
- [13] J. Bourgain, A. Furman, E. Lindenstrauss, S. Mozes, *Stationary measures and equidistribution for orbits of non-abelian semigroups on the torus*, JAMS **24** (2011), 231–280.
- [14] J. Bourgain, A. Gamburd, *Uniform expansion bounds for Cayley graphs $SL_2(\mathbb{F}_p)$* , Ann. of Math. (2) **167** (2008), no. 2, 625–642.
- [15] J. Bourgain, A. Gamburd, *On the spectral gap for finitely-generated subgroups of $SU(2)$* , Invent. Math. **171** (2008), no. 1, 83–121.
- [16] J. Bourgain, A. Gamburd, *A spectral gap theorem in $SU(d)$* , JEMS **14** (2012), 1455–1511.
- [17] J. Bourgain, A. Gamburd, P. Sarnak, *Affine linear sieve, expanders and sum-product*, Invent. Math. **179** (2010), 559–644.
- [18] J. Bourgain, A. Gamburd, P. Sarnak, *Generalization of Selberg’s theorem and Selberg’s sieve*, Acta Math. **207** (2011), no 2, 255–290.
- [19] J. Bourgain, M. Garaev, *Sumsets of reciprocals in prime fields and multilinear Kloosterman sums*, arXiv:1211.4184.
- [20] J. Bourgain, A. Glibichuk, S. V. Konyagin, *Estimate for the number of sums and products and for exponential sums in fields of prime order*, J. London Math. Soc. (2), **73** (2006), 380–398.
- [21] J. Bourgain, N. Katz, T. Tao, *A sum-product estimate in finite fields and their applications*, Geom. Funct. Anal., **14** (2003), 27–57.
- [22] J. Bourgain, S. V. Konyagin, *Estimates for the number of sums and products and for exponential sums over subgroups in fields of prime order*, C.R. Acad. Sci. Paris, **337** (2003), 75–80.
- [23] J. Bourgain, P. Varju, *Expansion in $SL_d(\mathbb{Z}/q\mathbb{Z})$, q arbitrary*, Inventiones Math. **188** (2012), no 1, 151–173.
- [24] E. Breuillard, A. Gamburd, *Strong uniform expansion in $SL(2, p)$* , GAFA **20** (2010), no 5, 1201–1209.
- [25] E. Breuillard, B. Green, T. Tao, *Approximate subgroups of linear groups*, GAFA (2011), no 4, 774–819.
- [26] E. Breuillard, B. Green, T. Tao, *The structure of approximate groups*, IHES Publ. **116**, 115–221 (2012).
- [27] M. Chang, *Factorization in generalized arithmetic progressions and applications to the Erdős-Szemerédi sum-product problems*, GAFA **13** (2003), no 4, 720–736.

- [28] M. Chang *Sum and product of different sets*, Contrib. Discrete Math. **1** (2006), no 1, 47–56.
- [29] J. Conway, C. Radin, *Quaquaversal tilings and rotations*, Inv. Math. **132** (1998), 179–188.
- [30] B. Draco, L. Sadun, D. Van Wieren, *Growth rates in the quaquaversal tiling*, Discrete Comput. Geom. **23** (2000), 419–435.
- [31] G. Edgar, C. Miller, *Borel subrings of the reals*, Proc. AMS **131** (2003), no 4, 1121–1129.
- [32] P. Erdős, *Problems and results in combinatorial number theory, III*, Number Theory Day, New York 1976.
- [33] P. Erdős, E. Szemerédi, *On Sums and Products of Integers*, in Studies in pure mathematics (ed. by L. Alpár, P. Erdős, G. Halász, A. Sárközy) Birkhäuser, Basel, 1983, 213–218.
- [34] P. Erdős, B. Volkmann, *Additiven Gruppen mit vorgegebener Hausdorffscher Dimension*, J. Reine Angew. Math. **221** (1966), 203–208.
- [35] J. Evertse, H. Schlickewei, W. Schmidt, *Linear equations in variables which lie in a multiplicative group*, Annals Math. **155** (2002), 807–836.
- [36] K. Falconer, *Rings of fractional dimension*, Mathematika **31** (1989), 201, 25–27.
- [37] J. Friedlander, H. Iwaniec, *Opera de Cribro*, AMS, Colloquium Publ. **57**, 2010.
- [38] J. Friedlander, H. Iwaniec, *The Brun–Titchmarsh theorem*, Analytic number theory (Kyoto, 1996), 85–93, London Math Soc. Lecture Note, Ser. **247**, Cambridge UP (1997).
- [39] J. Friedlander, H. Iwaniec, *Analytic Number Theory*, AMS Colloquium Publ. **53** (2004).
- [40] H. Furstenberg, *Stiffness of group actions*, in ‘Lie groups and ergodic theory’, Mumbai, Tata Inst. Fund. Res. Stud. in Math. **14**, Tata Inst. Fund. Res., Bombay, 1998, 105–117.
- [41] A. Gamburd, D. Jakobson, P. Sarnak, *Spectra of elements in the group ring of $SU(2)$* , H. Eur. Math. Soc. **1** (1999), 51–85.
- [42] M. Garaev, *Sums and products of sets and estimates for rational trigonometric sums in fields of prime order*, Russian Math. Surveys **65** (2010), no 4, 599–658.
- [43] B. Green, *Approximate groups and their applications: work of Bourgain, Gamburd, Helfgott and Sarnak*, Current Events Bulletin of the AMS, 2010.
- [44] R. Heath-Brown, S. V. Konyagin, *New bounds for Gauss sum derived from k th powers and for Heilbronn’s exponential sum*, Quart J. Math. **51** (2003), 221–335.
- [45] H. A. Helfgott, *Growth and generation in $SL_2(\mathbb{Z}/p\mathbb{Z})$* , Annals of Math. (2) **167** (2008), no. 2, 601–623.
- [46] E. Hrushovski, *Stable group theory and approximate subgroups*, arXiv:0909.2190.
- [47] N. Katz, T. Tao, *Some connections between Falconer’s distance set conjecture and sets of Furstenberg type*, New York J. Math. **7** (2001), 149–187.
- [48] A. Kontorovich, *From Apollonius to Zaremba: local-global phenomena in thin orbits*, arXiv:1208.5460, to appear in Bulletin AMS.

- [49] S. V. Konyagin, *Estimates for trigonometric sums and for Gaussian sums*, in IV Intern. Conf. on Modern Problems of Number Theory and its Applications: Current Problems, Part III, Moscow State Univ., 2002, 86–114 (in Russian).
- [50] S. V. Konyagin, I. Shparlinski, *Character sums with exponential functions and their applications*, Cambridge Tracts in Mathematics, **136**, Cambridge UP, Cambridge 1999.
- [51] E. Kowalski, *Sieve in expansion*, arXiv:1012.2793.
- [52] W. Luo, *Bounds on incomplete multiple Kloosterman sums*, J. Number Th. **75** (1999), 41–46.
- [53] H. Montgomery, R. Vaughan, T. Wooley, *Some remarks on Gauss sums associated to k^{th} powers*, Math. Proc. Cambridge Philo. Soc. **118** (1995), 21–33.
- [54] L. Pyber, E. Szabó, *Growth in finite simple groups of the Lie type*, arXiv:1208.2538.
- [55] C. Radin, L. Sadun, *Subgroups of $SO(3)$ associated with tilings*, J. Algebra **202** (1998), no 2, 611–633.
- [56] T. Sanders, *On the Bogolyubov-Ruzsa lemma*, arXiv:1011-0107.
- [57] A. Salehi, P. Varju, *Expansion in perfect groups*, to appear in GAFA.
- [58] P. Sarnak, *Notes on thin matrix groups*, to appear in MSRI publications.
- [59] J. Solymosi, *Bounding multiplicative energy by the sumset*, Adv. Math. **222** (2009), no. 2, 402–408.
- [60] T. Tao, *Product set estimates for non-commutative groups*, Combinatorica **28** (2008), no 5, 547–594.
- [61] T. Tao, V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics **105**, Cambridge University Press, 2006.

Jean Bourgain

*School of Mathematics,
Institute for Advanced Study,
1 Einstein Drive,
Princeton,
NJ 08540*

e-mail: bourgain@math.ias.edu

SMALL DOUBLING IN GROUPS

EMMANUEL BREUILLARD, BEN GREEN and TERENCE TAO

Let A be a subset of a group $G = (G, \cdot)$. We will survey the theory of sets A with the property that $|A \cdot A| \leq K|A|$, where $A \cdot A = \{a_1 a_2 : a_1, a_2 \in A\}$. The case $G = (\mathbb{Z}, +)$ is the famous Freiman–Ruzsa theorem.

1. SMALL DOUBLING IN ABELIAN GROUPS

Let $G = (G, +)$ be an abelian group, the group operation being written with the $+$ symbol. If $A \subseteq G$ is a finite set, we may consider the sumset $A + A := \{a_1 + a_2 : a_1, a_2 \in A\}$. We have the trivial bounds

$$(1.1) \quad |A| \leq |A + A| \leq \min\left(\frac{1}{2}|A|(|A| + 1), |G|\right)$$

on the cardinality $|A + A|$ of this sumset. One expects the trivial upper bound to be attained with equality (or near-equality) unless A has some special additive structure. For example, it is certainly attained when $A = \{1, 2, 2^2, \dots, 2^{n-1}\}$ consists of powers of two.

Clarifying what exactly is meant by *special additive structure* turns out to be very interesting, and is the main topic of this survey. Specifically, we will be interested in describing as carefully as we can the structure of non-empty finite sets A for which $\sigma[A] := |A + A|/|A|$ is at most K , where $K \in \mathbb{R}^+$ is some constant. We say that such a set A has *doubling* at most K . If $\sigma[A]$ is “small”, we informally say that A has small doubling.

Let us begin with some examples of sets with small doubling. The simplest example is that of a finite subgroup, or a subset of one.

Example 1. Suppose that A is a finite subgroup of G . Then $|A + A| = |A|$, and so $\sigma[A] = 1$. Similarly if A is a coset of a subgroup of G . If A is not a whole subgroup but occupies a non-zero proportion δ of some finite subgroup $H \leq G$ then $A + A \subseteq H$, and so $\sigma[A] \leq 1/\delta$.

It is a nice exercise to prove that the *only* finite non-empty sets with doubling 1 are cosets of subgroups. On the other hand it is very easy to come up with an example of a set A with small doubling which is not related to a subgroup.

Example 2. Suppose that $G = \mathbb{Z}$, and let u_1 and $N_1 > 0$ be integers. We define the arithmetic progression¹

$$P(u_1; N_1) := \{n_1 u_1 : 0 \leq n_1 < N_1\}$$

If $A = P(u_1; N_1)$ then $A + A$ is given by

$$A + A = P(u_1; 2N_1 - 1) = \{n_1 u_1 : 0 \leq n_1 < 2N_1 - 1\},$$

and hence $\sigma[A] < 2$. If A occupies a proportion δ of some arithmetic progression then $\sigma[A] \leq 2/\delta$.

There are multidimensional constructions of a similar nature.

Example 3. Suppose that $G = \mathbb{Z}$, let u_1, \dots, u_d and $N_1, \dots, N_d > 0$ be integers. We introduce the d -dimensional progression

$$P(u_1, \dots, u_d; N_1, \dots, N_d) := \{n_1 u_1 + \dots + n_d u_d : 0 \leq n_i < N_i\}$$

If $A = P(u_1, \dots, u_d; N_1, \dots, N_d)$ then the sumset $A + A$ is given by

$$\begin{aligned} A + A &= P(u_1, \dots, u_d, 2N_1 - 1, \dots, 2N_d - 1) \\ &= \{n'_1 u_1 + \dots + n'_d u_d : 0 \leq n'_i < 2N_i - 1\}. \end{aligned}$$

Thus $A + A$ can be covered by 2^d translates of A , so that $\sigma[A] \leq 2^d$. If A occupies a proportion δ of some d -dimensional progression then $\sigma[A] \leq 2^d/\delta$.

Finally, one can combine any of these examples using a direct product construction.

Example 4. Suppose that $A_1 \subseteq G_1$, $A_2 \subseteq G_2$ and that $\sigma[A_i] \leq K_i$ for $i = 1, 2$. Consider $A_1 \times A_2$ as a subset of $G_1 \times G_2$. Then $\sigma[A_1 \times A_2] \leq K_1 K_2$.

It turns out that, qualitatively at least, the above four examples provide a complete description of sets with small doubling in abelian groups. In the case $G = \mathbb{Z}$ this was established by Freiman [23] and Ruzsa [61].

¹The notation here may seem slightly odd. The point is that an arithmetic progression is a very special case of a much more general object called a *coset nilprogression*, which we will discuss in detail in what follows using an elaboration of the same notation.

Theorem 1.1 (Freiman’s theorem). *Let A be a finite non-empty set of integers \mathbb{Z} such that $\sigma[A] \leq K$. Then A is contained within a generalised arithmetic progression*

$$P(u_1, \dots, u_r; N_1, \dots, N_r) := \{n_1 u_1 + \dots + n_r u_r : n_1, \dots, n_r \in \mathbb{Z}, 0 \leq n_i < N_i\}.$$

Here $u_1, \dots, u_r \in \mathbb{Z}$ are integers, the rank r is $O_K(1)$ and the volume $V := N_1 \dots N_r$ satisfies² $V \ll_K |A|$.

Note that \mathbb{Z} does not have any interesting subgroups, so only Examples 2 and 3 are relevant here. At the other (high-torsion) extreme, Ruzsa [63] gave a very short and elegant proof of the following statement. Here, \mathbb{F}_2^ω is the direct product of countably many copies of the finite field \mathbb{F}_2 .

Theorem 1.2 (Ruzsa). *Let A be a finite non-empty subset of \mathbb{F}_2^ω , and suppose that $\sigma[A] \leq K$. Then there exists a subgroup H containing A such that $|H| \ll_K |A|$.*

Ruzsa’s theorem works in \mathbb{F}_p^ω for an arbitrary prime p , although the dependence of the \ll_K constant is not uniform³ in p . Ruzsa and the second author [33] combined these two results to get a result valid for all abelian groups.

Theorem 1.3 (Green–Ruzsa). *Let A be a finite non-empty subset of an additive group G such that $\sigma[A] \leq K$. Then there exists a coset progression $H + P$, where H is a finite subgroup of G and $P = P(u_1, \dots, u_r; N_1, \dots, N_r)$ is a generalised arithmetic progression of rank $O_K(1)$, such that $A \subseteq H + P$ and $|H|N_1 \dots N_r \ll_K |A|$.*

These theorems completely resolve the qualitative question of describing the structure of sets A whose doubling constant $\sigma[A]$ is at most K . There are many very interesting quantitative issues in connection with this question, and we will address these in §5.

Let us give a brief selection of other results connected with small doubling in abelian groups.

The first does not concern finite sets (although there are variants of it that do, such as Propositions 5.3 and 5.4). If $A \subseteq \mathbb{R}^d$ is compact and of positive measure then we define its doubling constant to be $\sigma[A] := \mu(A + A)/\mu(A)$, where μ is Lebesgue measure.

² $O_K(1)$ means a quantity bounded by C_K for some constant C_K depending only on K . The notation $X \ll_K Y$ means that $X \leq C_K Y$ for some C_K depending only on K . We might equivalently write $X = O_K(Y)$.

³The optimal value of this constant was worked out recently in [51].

Proposition 1.4. *Suppose that A is a compact subset of \mathbb{R}^d of positive measure and that $\sigma[A] \leq K$. Then $d \leq \log_2 K$.*

Proof. This is essentially trivial: $A + A$ contains the dilate $2 \cdot A$, which has 2^d times the volume of A . The claim also follows from the more general *Brunn-Minkowski inequality* $\mu(A + B)^{1/d} \geq \mu(A)^{1/d} + \mu(B)^{1/d}$ (see e.g. [26]).

There are still further results connected with *very* small doubling, when $K < 2$, or with moderately small doubling (when $2 \leq K \leq 3$ say). Let us finish this section by giving a very small and incomplete selection of them. Perhaps the most famous is the Cauchy-Davenport-Chowla theorem [11, 16]. When combined with Vosper’s theorem [86], this gives the following result.

Theorem 1.5. *Suppose that p is a prime and that $A \subseteq \mathbb{Z}/p\mathbb{Z}$. Suppose that $\sigma[A] < 2$. Then either $|A| > p/2$ and $A + A$ is all of $\mathbb{Z}/p\mathbb{Z}$, or else A is an arithmetic progression.*

See e.g. [82, Theorem 5.4, Theorem 5.9] for a proof. Kneser’s theorem [47] is a generalisation of the Cauchy-Davenport theorem to arbitrary abelian groups. A consequence of it is the following.

Theorem 1.6. *Suppose that G is an arbitrary abelian group and that $A \subseteq G$. Suppose that $\sigma[A] \leq K$ where $K < 2$. Then $A + A$ is a union of cosets of a subgroup $H \leq G$ of size at least $(2 - K)|A|$.*

Finally let us mention a result [23] known as *Freiman’s $3k - 3$ theorem*, concerning sets of integers with doubling at most (roughly) 3. See [50] for a simpler proof and a generalisation to pairs of sets. It gives a very precise version of Theorem 1.1 in this regime.

Theorem 1.7. *Suppose that $A \subseteq \mathbb{Z}$ is a finite set with $|A| \geq 3$ and with doubling constant $K = \sigma[A]$. Suppose that $K < 3 - \frac{3}{|A|}$. Then A is contained in an arithmetic progression P of length at most $(K - 1)|A| + 1$.*

Results by Stanchescu [75, 76] make various assertions, more precise than Theorem 1.1, for values of K in the range $3 \leq K < 4$.

Finally we remark that in many of the above theorems the hypothesis $\sigma[A] \leq K$ may be varied to other, related, conditions such as $|A - A| \leq K|A|$ or $|A + B| \leq K|A|^{1/2}|B|^{1/2}$ using standard additive combinatorial lemmas; see [82, Chapter 2]. There are also variants when one replaces the full sumset $A + A$ by a partial sumset $A \overset{G}{+} A := \{a + b : (a, b) \in G\}$ for some (dense) subset G of $A \times A$, using what is now known as the *Balog-Szemerédi-Gowers lemma*. Again, see [82, Chapter 2] for details and further references, and §4 for further comments.

2. SMALL DOUBLING IN ARBITRARY GROUPS – EXAMPLES

We now turn to the main focus of this survey, which is to study inverse sum-set theorems in the *noncommutative* setting, in which one works with finite nonempty subsets A of an arbitrary group G . To emphasise the fact that G is not necessarily abelian, we write the group operation multiplicatively.

We are now interested in the structure of finite sets $A \subseteq G$ with the property that $\sigma[A] := |A \cdot A|/|A|$ is at most K , where $A \cdot A := \{a_1 a_2 : a_1, a_2 \in A\}$. The trivial bounds are now $|A| \leq |A \cdot A| \leq \min(|A|^2, |G|)$. Equality can occur in the upper bound, for example if $A = \{xy^i : i = 0, 1 \dots, n - 1\}$ where x, y are generators of a non-abelian free group. As in the first section, we begin with some examples. The first few of these are parallel to the abelian examples we discussed before.

Example 5. Suppose that A is a subgroup of G . Then $|A \cdot A| = |A|$, and $\sigma[A] = 1$. Similarly if $A = Hx$ is a coset of some subgroup $H \leq G$ where x lies in the normaliser of H (that is to say $xH = Hx$). If A is not a whole subgroup but occupies a proportion δ of some finite subgroup $H \leq G$ then $A \cdot A \subseteq H$, so $\sigma[A] \leq 1/\delta$.

It is a nice exercise to prove the converse to this, namely that the only sets with doubling 1 are cosets Hx , where H is a subgroup and x normalises H .

Example 6. The nonabelian analogue of an arithmetic progression is a geometric progression $P(u_1; N_1) := \{u_1^{n_1} : 0 \leq n_1 < N_1\}$. Assuming all N_1 elements are distinct, we have $A \cdot A = \{u_1^{n'_1} : 0 \leq n'_1 < 2N_1 - 1\}$, and so $\sigma[A] < 2$. If A occupies a proportion δ of some geometric progression then $\sigma[A] \leq 2/\delta$.

As in the abelian case, there are multidimensional constructions of a similar nature, but one must be a little careful in the absence of commutativity.

Example 7. Let A be a set of the form

$$P(u_1, \dots, u_d; N_1, \dots, N_d) := \{u_1^{n_1} u_2^{n_2} \dots u_d^{n_d} : 0 \leq n_i < N_i\},$$

where the u_i commute and $N_1, \dots, N_d > 0$ are integers. We call this a d -dimensional progression. Then $A \cdot A$ is equal to

$$P(u_1, \dots, u_d; 2N_1 - 1, \dots, 2N_d - 1) = \{u_1^{n'_1} u_2^{n'_2} \dots u_d^{n'_d} : 0 \leq n'_i < 2N_i - 1\}.$$

Thus, $A \cdot A$ can be covered by 2^d dilates of A , so that $\sigma[A] \leq 2^d$. If A occupies a proportion δ of some (proper) d -dimensional progression then $\sigma[A] \leq 2^d/\delta$.

Just as in the abelian case, we may consider direct products of sets with small doubling and thereby obtain new examples. In the non-abelian case, however, there are two genuinely new examples of sets with small doubling such as the following.

Example 8. Let $N_1, N_2, N_{1,2}$ be positive integers, and let A be the set of 3×3 matrices defined as follows. Let

$$u_1 := \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad u_2 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

and set

$$A = P(u_1, u_2, [u_1, u_2]; N_1, N_2, N_{1,2}) \\ := \{u_1^{n_1} u_2^{n_2} [u_1, u_2]^{n_{1,2}} : 0 \leq n_1 < N_1, 0 \leq n_2 < N_2, 0 \leq n_{1,2} < N_{1,2}\}.$$

Here,

$$[u_1, u_2] := u_1 u_2 u_1^{-1} u_2^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is the commutator of u_1 and u_2 . Noting that

$$u_1^{n_1} u_2^{n_2} [u_1, u_2]^{n_{1,2}} = \begin{pmatrix} 1 & n_1 & n_1 n_2 + n_{1,2} \\ 0 & 1 & n_2 \\ 0 & 0 & 1 \end{pmatrix},$$

it follows that $|A| = N_1 N_2 N_{1,2}$. Furthermore one may easily check that

$$A \cdot A \subseteq \left\{ \begin{pmatrix} 1 & n'_1 & n'_{1,2} \\ 0 & 1 & n'_2 \\ 0 & 0 & 1 \end{pmatrix} : \begin{array}{l} 0 \leq n'_1 < 2N_1 \\ 0 \leq n'_2 < 2N_2 \\ 0 \leq n'_{1,2} < 3N_1 N_2 + 2N_{1,2} \end{array} \right\}.$$

Thus if $N_{1,2} \geq N_1 N_2$ then $\sigma[A] \leq 20$.

We call the preceding example a *nilprogression*. The name comes from the fact that the group of 3×3 upper-triangular matrices (the Heisenberg group) is nilpotent of class 2, which means that the group is nonabelian and that higher-order commutators such as $[u_1, [u_2, u_3]]$ are all equal to the identity. We will define nilprogressions in general later on. The second new type of example combines subgroups with progressions in a manner which is not a direct product. The next example was shown to us by Helfgott.

Example 9. Let p be a large prime, let $r, s, t \in \mathbb{F}_p$ be fixed generators of \mathbb{F}_p^* , let N_1, N_2, N_3 be positive integers, and define A to be a set of 3×3 matrices over \mathbb{F}_p as follows. Set

$$A = H \cdot P(u_1, u_2; u_3; N_1, N_2, N_3)$$

where

$$H := \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{F}_p \right\},$$

$$u_1 := \begin{pmatrix} r & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad u_2 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad u_3 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & t \end{pmatrix}$$

and

$$P(u_1, u_2, u_3; N_1, N_2, N_3) := \{u_1^{n_1} u_2^{n_2} u_3^{n_3} : 0 \leq n_i < N_i\},$$

as in Example 7 above. Thus

$$A = \left\{ \begin{pmatrix} r^{n_1} & x & z \\ 0 & s^{n_2} & y \\ 0 & 0 & t^{n_3} \end{pmatrix} : x, y, z \in \mathbb{F}_p, 0 \leq n_i < N_i \right\}.$$

One may easily check that

$$A \cdot A \subseteq \left\{ \begin{pmatrix} r^{n'_1} & x & z \\ 0 & s^{n'_2} & y \\ 0 & 0 & t^{n'_3} \end{pmatrix} : x, y, z \in \mathbb{F}_p, 0 \leq n'_i < 2N_i \right\},$$

and so $\sigma[A] \leq 8$.

Examples 8 and 9 (and in fact all of the other examples we have mentioned) are *coset nilprogressions*, which turn out to be the appropriate generalisation of a coset progression (cf. Theorem 1.3) to the nonabelian setting.

To conclude this section we discuss the general notion of a coset nilprogression, that is to say the natural generalisation of all the preceding examples. There are several roughly equivalent definitions, which turn out to be “essentially the same”, meaning that a coset nilprogression of one type is efficiently covered by coset nilprogressions of another type. We begin by giving one definition of a *nilprogression*, following [10, Definition 2.5].

Definition 2.1 (Nilprogression). Let u_1, \dots, u_r be elements in a nilpotent group of step s , that is to say a group in which commutators of order $s + 1$ or greater are all trivial. Let N_1, \dots, N_r be positive integers. Define the *nilprogression* $P^*(u_1, \dots, u_r; N_1, \dots, N_r)$ to consist of all products of the u_i and their inverses u_i^{-1} for which the letter u_i and its inverse u_i^{-1} appear at most N_i times between them, and the terms in the product may be arranged in an arbitrary order (e.g. $P^*(u_1, u_2; 1, 1)$ contains $u_1 u_2, u_2 u_1, u_1^{-1} u_2$, etc.).

We have written P^* instead of P to emphasise the fact that this is not *quite* the same as the objects $P(u_1, \dots, u_r, N_1, \dots, N_r)$ we considered earlier. It can be shown that if A is a nilprogression then $\sigma[A] \ll_{r,s} 1$. With this in hand it is quite straightforward to define a coset nilprogression.

Definition 2.2 (Coset nilprogression). Let G be a group and suppose that $u_1, \dots, u_r \in G$. Suppose that H is a finite subgroup of G which is normal in $G_0 := \langle u_1, \dots, u_r \rangle$. Suppose that G_0/H is nilpotent of step s . Then the set $H \cdot P^*(u_1, \dots, u_r; N_1, \dots, N_r)$ is called a *coset nilprogression* of rank r and step s .

Once again one may show that if A is a coset nilprogression then $\sigma[A] \ll_{r,s} 1$, that is to say coset nilprogressions are examples of sets with small doubling constant. An alternative way to define nilprogressions is as the image of a “ball” in a free nilpotent group. This gives objects which generalise our examples more directly, but requires quite a lot of nomenclature concerning basic commutators. For more details see [5, Definition 1.4] and also Tointon’s paper [83], which clarifies aspects of the relation between the two types of nilprogression⁴.

3. SMALL DOUBLING IN ARBITRARY GROUPS – THEOREMS

A basic aim in the subject, first suggested by Helfgott and Lindenstrauss, is to show that all sets of small doubling are related to one of the examples discussed in the preceding section, namely coset nilprogressions. Theorems 1.1, 1.2 and 1.3 in §1 were results of this type in the abelian case. While such results have now been established by the authors in full generality, for applications such as the ones in §6 one often needs a result with good bounds, and in the general case none are known. Much work, then, has been done on specific groups (for example matrix groups) where additional structure is extremely helpful and quite precise results have been obtained.

⁴Note that Tointon calls the objects of [5, Definition 1.4] *nilpotent progressions* but otherwise his nomenclature is essentially the same as ours.

Furthermore one does not always need (in fact one essentially never needs) to see the full structure of a coset nilprogression to draw useful applications.

With the aim of clarity of exposition, we will in this section only examine results of the following type, which we call “The structure theorem”.

Prototype Theorem 3.1. *Suppose that A is a set in some group G , belonging to a specific class (matrix group, nilpotent group, solvable group, free group ...) and that $\sigma[A] \leq K$. Then there is a set $A' \subseteq A$, $|A'| \geq |A|/K'$, with A' contained in some set P lying in a “structured” class of sets \mathcal{C} .*

These results are, therefore, a little weaker than the theorems of §1, which cover the whole of A by a structured object. However theorems of that type can be obtained from results having the form of Prototype Theorem 3.1, and to an extent this amounts only to additive-combinatorial “book-keeping”, although some more precise variants of this type are both deep and interesting.

Doubling less than 2. The case of very small doubling, in which $\sigma[A]$ is close to 1, received attention at the hands of Freiman [24] almost 50 years ago (see also [25]). He showed (among other things) that if $\sigma[A] < 3/2$, then $H := A \cdot A^{-1}$ is a finite group of order $|A \cdot A| = \sigma[A]|A|$, and that $A \subseteq xH = Hx$ for some x . In a similar vein, an argument of Hamidoune [41, 81] shows that if $\sigma[A] < 2 - \varepsilon$ for some $\varepsilon > 0$, then there exists a finite group H of order $|H| \leq \frac{2}{\varepsilon}|A|$, such that A can be covered by at most $\frac{2}{\varepsilon} - 1$ right-cosets Hx of H . See also [67] for a different proof of a related result. Very recently, a more complete classification of the sets A with $\sigma[A] < 2$ was achieved in [17].

Small doubling in nilpotent and solvable groups. Results along the lines of Theorem 3.1 with G nilpotent or solvable of fixed step have been developed in various papers [5, 6, 22, 27, 68, 79, 80, 83], there being a tradeoff in each case between generality and the quality of the bounds obtained. A quite satisfactory recent result of Tointon [83] is the following, which applies to arbitrary nilpotent groups of fixed step.

Theorem 3.2 (Tointon). *Suppose that G is nilpotent of step s . Then the structure theorem holds with $K' \sim \exp(K^{O_s(1)})$ and with \mathcal{C} consisting of coset nilprogressions of rank $K^{O_s(1)}$ and size no more than $\exp(K^{O_s(1)})|A|$.*

Let us also mention a short note of the authors [8], which adapts an argument of Gleason from the theory of locally compact groups to show that the s -dependence is unnecessary when G is torsion-free.

Small doubling in matrix groups. When G is a group of matrices over a field, one has available a wide variety of machinery. One may attempt to exploit the fact that matrix multiplication involves both addition and multiplication of the entries, and hence bring into play results from *sum-product theory* such as [3]. One may also try to involve algebraic geometry and particularly the theory of algebraic groups. All of these techniques have enjoyed success.

A pioneering result about small doubling in matrix groups was that of Elekes and Király [20].

Theorem 3.3 (Elekes–Király). *Suppose that $G = \mathrm{SL}_2(\mathbb{R})$. Then the structure theorem holds for some $K' = O_K(1)$ and with \mathcal{C} consisting of cosets of abelian subgroups of $\mathrm{SL}_2(\mathbb{R})$.*

One interesting consequence of this is that it implies the same result with G equal to the free group. This is because $\mathrm{SL}_2(\mathbb{R})$ contains two elements (and hence k elements, for any k) generating a free group. Further important work on additive combinatorics in the free group was done by Razborov [60] and Safin [65].

Elekes and Király did not obtain useful bounds for K' . Subsequent advances have addressed this issue, and have also led to the replacement of $\mathrm{SL}_2(\mathbb{R})$ by an arbitrary matrix group. Significant progress in this regard was made by Helfgott [42], who applied the sum-product theorem of Bourgain, Katz and Tao [3] to show⁵ that Theorem 3.3 holds with G replaced by $\mathrm{SL}_2(\mathbb{C})$ and with K' having polynomial dependence on K . He subsequently generalised this to $\mathrm{SL}_3(\mathbb{C})$, with \mathcal{C} consisting of cosets of nilpotent subgroups of SL_3 . Chang [13] also proved various results in this direction, for example obtaining a structure theorem in the case $G = \mathrm{SL}_3(\mathbb{Z})$ prior to the work of Helfgott.

A breakthrough came in 2009 with a paper of Hrushovski [45], who applied model-theoretic arguments to generalise Elekes–Király’s result to SL_n .

Theorem 3.4 (Hrushovski). *Suppose that $G = \mathrm{SL}_n(\mathbb{C})$. Then the structure theorem holds for some $K' = O_K(1)$ and with \mathcal{C} consisting of cosets of solvable subgroups of $\mathrm{SL}_n(\mathbb{C})$.*

Combining this with the main result of [6], “solvable” can be replaced by “nilpotent”. More down-to-earth proofs are now known due to work of Pyber-Szabó [58] and the authors [7, 9], and furthermore these arguments

⁵In fact Helfgott does not quite state this result or the one for $\mathrm{SL}_3(\mathbb{C})$, his concern having been with $\mathrm{SL}_2(\mathbb{F}_p)$ and $\mathrm{SL}_3(\mathbb{F}_p)$, but it follows very easily from his methods.

show that K' can be taken to be $K^{O_n(1)}$. These arguments use some of Helfgott's ideas as well as some more purely algebraic group theoretic facts. The argument of [7] was, in addition, heavily influenced by groundbreaking work of Larsen and Pink [49] in their work on finite subgroups (as opposed to finite subsets of small doubling) of linear groups.

We have discussed the case $G = \mathrm{SL}_n(\mathbb{C})$. However the analogous question over finite fields is more interesting, enjoys wider application, and was the historical motivation for much of the work we have just mentioned. In this setting Helfgott's result states the following.

Theorem 3.5 (Helfgott). *Suppose that $G = \mathrm{SL}_2(\mathbb{F}_p)$. Then either $|A| \geq K^{-C}|G|$, or else the structure theorem holds with K' polynomial in K and with \mathcal{C} consisting of cosets of solvable (or upper-triangular) subgroups of $\mathrm{SL}_2(\mathbb{F}_p)$.*

This was generalised to \mathbb{F}_q , q arbitrary, by Dinai [18]. Subsequently, Helfgott [43] generalised his previous argument to the setting of $\mathrm{SL}_3(\mathbb{F}_p)$ and finally Pyber-Szabó [58] obtained the same result in $\mathrm{SL}_n(\mathbb{F}_q)$, as well as in more general finite simple groups of Lie type and bounded rank, while the authors [7] obtained simultaneously closely related results. See the related surveys [4, 59].

General groups. A qualitatively complete classification of sets of small doubling in arbitrary nonabelian groups was obtained by the authors [10].

Theorem 3.6. *Suppose that G is an arbitrary group. Then the structure theorem holds for some $K' = O_K(1)$ and with \mathcal{C} consisting of coset nilprogressions P with rank and step $O_K(1)$, and with $|P| \leq |A|$.*

In fact, the theorem applies to sets with small doubling in *local* groups, which we will not define here (and in fact this generalisation is an important part of the proof). However the dependence of K' on K is not known explicitly at all. On the other hand, the rank and step of P can be chosen to be at most $6 \log_2 K$; see [10, Theorem 10.1].

We do not have the space to say much about the proof of this theorem here. A crucial theme is a certain ‘‘correspondence principle’’, due to Hrushovski [45], linking sets with small doubling (or rather approximate groups, as discussed in the next section) and *locally compact topological groups*. One may then bring into play the extensive theory, largely developed in the 1950s and before, of locally compact groups in connection with Hilbert's fifth problem. See [10, 78] for considerably more on this.

4. APPROXIMATE GROUPS

This survey is about sets with small doubling constant. However for technical reasons many papers consider a closely-related object called an *approximate group*, first introduced in [79].

Definition 4.1. Let A be a finite set in a group G . We say that A is a K -approximate group if A is symmetric (that is to say if $a \in A$ then $a^{-1} \in A$), contains the identity and if $A \cdot A$ is contained in $X \cdot A$ for some set X of size at most K .

This notion has some technical advantages over the notion of a set with small doubling; for example, it is immediately clear that the image of a K -approximate group under a homomorphism is also a K -approximate group, but there are examples (even in the abelian case) which show that the doubling constant $\sigma[\pi(A)]$ of a finite set A under a homomorphism π is strictly greater than that for A ; see [82, Exercise 2.2.10]. By construction, it is clear that any finite K -approximate group A has $\sigma[A] \leq K$. It is also easily seen that one has control over the higher product sets $A^n := A \cdot A \cdots A$, specifically $|A^n| \leq K^{n-1}|A|$, whereas no such bound is generally available for sets of small doubling⁶.

Although sets with small doubling need not be approximate groups, we do have the following converse result, obtained in [79, Theorem 4.6]. It asserts in some sense that sets of small doubling are essentially “controlled” by approximate groups:

Theorem 4.2. *Let A be a finite non-empty subset of a multiplicative group $G = (G, \cdot)$ such that $\sigma[A] \leq K$, where $K \geq 2$. Then one can find a $K^{O(1)}$ -approximate group H in G of cardinality $\ll K^{O(1)}|A|$ such that A can be covered by $K^{O(1)}$ left (or right) translates of H .*

The arguments used to prove this are fairly elementary, and are non-commutative variants of some arguments of Ruzsa [62]. By contrast the majority of the structural results discussed in §3 rely on considerable machinery. Thus, in a certain sense, one should think of the theory of sets with small doubling and the theory of approximate groups as equivalent.

⁶Although one does have $|A^n| \leq K^n|A|$ in the abelian case, a result of Plünnecke and Ruzsa [57, 64] for which a very elegant proof was recently provided by Petridis [56].

5. QUANTITATIVE ASPECTS

Thus far we have talked mainly about qualitative results concerning sets with small doubling, with the notable exception of Helfgott's work and its successors, where one obtains polynomial dependencies. A particularly acute example of this is Theorem 3.6, where no explicit bounds are known at all.

Even in (in fact *especially* in) the abelian case, quantitative issues are very interesting. We mention some of these now, deferring to the excellent recent survey [70] for considerably more detail.

For the most part we discuss the quantitative issues related to Ruzsa's Theorem 1.2, concerning sets with small doubling in \mathbb{F}_2^ω . This most abelian of all settings has acted as a significant test case for ideas. Theorem 1.2 stated that if $A \subseteq \mathbb{F}_2^\omega$ is a finite set with $\sigma[A] \leq K$ then there exists a subgroup $H \leq \mathbb{F}_2^\omega$ containing A with $|H| \leq F(K)|A|$. Ruzsa himself obtained the bound $F(K) \leq K^2 2^{K^4}$. This was subsequently refined by Green-Ruzsa [33] and then by Sanders, and after that by Green-Tao [35], who showed that one can take $F(K) \leq 2^{2K+o(K)}$, a bound which is sharp up to the $o(K)$ term. This was further refined by Konyagin [48], and finally Chaim Even-Zohar [21] obtained the precise value of $F(K)$. The techniques here are those of extremal combinatorics, specifically the technique of *compressions*.

It was already realised by Ruzsa, however, that trying to cover A by a subgroup is an inefficient endeavour. He attributes to Katalin Marton the following question, which has since become known as the *Polynomial Freiman-Ruzsa Conjecture* (PFR).

Conjecture 5.1. *Suppose that $A \subseteq \mathbb{F}_2^\omega$ is a set with $\sigma[A] \leq K$. Then A is covered by K^C translates of some subspace $H \leq \mathbb{F}_2^\omega$ with $|H| \leq |A|$.*

At present this conjecture is unresolved, although there has been spectacular recent progress by Sanders [69], building on work of Schoen [73]. Sanders shows that this conjecture does hold with K^C replaced by $\exp(\log^{4+o(1)} K)$. Ruzsa formulated a number of equivalent statements to this conjecture, which were written up in [32]. Perhaps the most attractive is the following statement concerning *almost homomorphisms*.

Conjecture 5.2. *Suppose that $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^{n'}$ is a map with the property that $f(x+y) - f(x) - f(y) \in S$ for all x, y , where S is some set of size K . Is it true that $f = \tilde{f} + g$, where $\tilde{f} : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^{n'}$ is linear, and $|\text{im}(g)| \leq K^C$?*

This statement is trivial with K^C replaced by 2^K (define \tilde{f} to equal f on a basis of \mathbb{F}_2^n). Sanders's result establishes that it holds with

$$\exp(\log^{4+o(1)} K).$$

Let us turn now to the structure of sets with small doubling in Euclidean spaces, and in particular⁷ in \mathbb{Z} . In this setting extra tools coming from geometry can be brought into play. In particular we have the following result [23, Lemma 1.13], known as Freiman's lemma.

Proposition 5.3 (Freiman's lemma). *Suppose that A is a finite subset of some Euclidean space \mathbb{R}^d , and that $\sigma[A] \leq K$. Suppose that A is not contained in any proper affine subspace of \mathbb{R}^d . Then*

$$d \leq K - 1 + \frac{d(d+1)}{2|A|}.$$

In particular, if $|A| \gg_{\varepsilon, d} 1$ then $d \leq K - 1 + \varepsilon$.

The proof of this result is very short, but makes crucial use of convexity. A writeup may be found in, for example, [31]. A follow-up to this is the next result, also originally due to Freiman [23], with a simplified proof given subsequently by Bilu [1].

Proposition 5.4 (Freiman–Bilu lemma). *Suppose that A is a finite subset of some Euclidean space \mathbb{R}^d , and that $\sigma[A] \leq K$. Then there is a subset $A' \subseteq A$, $|A'| \gg_{\varepsilon, K} |A|$, which is contained in an affine subspace of \mathbb{R}^d of dimension at most $\log_2 K + \varepsilon$.*

Moderately good dependencies in this theorem are now known [34]. There is a version of the Polynomial Freiman–Ruzsa conjecture for subsets of \mathbb{R}^d , a question closely related to that of finding the correct dependencies on K in Proposition 5.4. We are not certain that this has been stated in the literature before.

Conjecture 5.5. *Suppose that A is a finite subset of some Euclidean space \mathbb{R}^d , and that $\sigma[A] \leq K$. Then A can be covered by K^C translates of some generalised progression $P = P(u_1, \dots, u_r; N_1, \dots, N_r)$ with $|P| \leq |A|$ and $r = O(\log K)$.*

⁷In fact the theories of small doubling of finite sets in \mathbb{R}^d and in \mathbb{Z} are essentially equivalent, since any finite subset of \mathbb{R}^d is *Freiman isomorphic* to a set of integers.

For a slightly more cautious conjecture one might take, instead of the “box-like” generalised progression P , a set obtained from the set of lattice points in a convex body in \mathbb{R}^r . Whether this is really a more general statement seems slightly unclear and perhaps deserves to be clarified (it is an issue in the geometry of numbers).

Much work has been done on quantitative results in \mathbb{Z} , starting with Ruzsa’s work and the important paper of Chang [12]. A comprehensive history and summary of results may be found in Sanders [70].

Green and Tao [36] and independently Lovett [51], building on work of Gowers [29], demonstrated a fairly tight equivalence between Conjectures 5.2 and 5.5 and quantitative versions of the inverse conjectures for the Gowers U^3 -norm. This has yet to reveal itself a viable way to attack these Conjectures 5.2 and 5.5, since the only known strategies for proving the inverse conjectures for the U^3 -norm either *use* results about approximate subgroups of \mathbb{Z} , or else are essentially qualitative in nature.

Almost no work has been done on quantitative questions in general groups. It may be, for all we know, that Theorem 3.6 holds with rather good quantitative dependencies.

6. APPLICATIONS AND OPEN QUESTIONS

We conclude this survey by briefly mentioning some applications of the theory of sets with small doubling (or, usually more accurately, the theory of approximate groups), in various contexts. We encourage the reader to look for more.

Expanders. Helfgott’s paper [42] was soon followed by an application due to Bourgain and Gamburd [2] concerning the construction of *expanders*. We offer a very brief discussion: for more details, see [52, 77]. In particular Bourgain and Gamburd’s results have now been very substantially generalised, culminating in an almost final result of Varjú [85] and Salehi-Golsefidy and Varjú [28].

For the purposes of this brief discussion an expander graph is a $2k$ -regular graph Γ on n vertices for which there is a constant $c > 0$ such that for any set X of at most $n/2$ vertices of Γ , the number of vertices outside X which are adjacent to X is at least $c|X|$. Expander graphs share many of the properties of random regular graphs, and this is an important reason why they are of great interest in theoretical computer science (and would have been of interest to Paul Erdős, one imagines). There are many

excellent articles on expander graphs ranging from the very concise [71] to the seriously comprehensive [44, 52].

A key issue is that of constructing explicit expander graphs, and in particular that of constructing *families* of expanders in which k and c are fixed but the number n of vertices tends to infinity. Many constructions have been given, and several of them arise from Cayley graphs. Let G be a finite group and suppose that $S = \{g_1^{\pm 1}, \dots, g_k^{\pm 1}\}$ is a symmetric set of generators for G . The Cayley graph $\mathcal{C}(G, S)$ is the $2k$ -regular graph on vertex set G in which vertices x and y are joined if and only if $xy^{-1} \in S$. Such graphs provided some of the earliest examples of expanders [54, 55]. A natural way to obtain a family of such graphs is to take some large “mother” group \tilde{G} admitting many homomorphisms π from \tilde{G} to finite groups, a set $\tilde{S} \subseteq \tilde{G}$, and then to consider the family of Cayley graphs $\mathcal{C}(\pi(\tilde{G}), \pi(\tilde{S}))$ as π ranges over a family of homomorphisms. The work under discussion concerns the case $\tilde{G} = \mathrm{SL}_2(\mathbb{Z})$, which of course admits homomorphisms $\pi_p : \mathrm{SL}_2(\mathbb{Z}) \rightarrow \mathrm{SL}_2(\mathbb{F}_p)$ for each prime p . For certain sets $\tilde{S} \subseteq \tilde{G}$, for example

$$\tilde{S} = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{\pm 1}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^{\pm 1} \right\}$$

or

$$\tilde{S} = \left\{ \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}^{\pm 1}, \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}^{\pm 1} \right\},$$

spectral methods from the theory of automorphic forms may be used to show that $(\mathcal{C}(\pi_p(\tilde{G}), \pi_p(\tilde{S})))_{p \text{ prime}}$ is a family of expanders. See [52] and the references therein. These methods depend on the fact that the group $\langle \tilde{S} \rangle$ has finite index in $\tilde{G} = \mathrm{SL}_2(\mathbb{Z})$ and they fail when this is not the case, for example when

$$(6.1) \quad \tilde{S} = \left\{ \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}^{\pm 1}, \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix}^{\pm 1} \right\}.$$

In [53] Lubotzky asked whether the corresponding Cayley graphs in this and other cases might nonetheless form a family of expanders, the particular case of (6.1) being known as his “1-2-3 question”. The paper of Bourgain and Gamburd under discussion answers this quite comprehensively.

Theorem 6.1 (Bourgain–Gamburd). *Let $\tilde{G} = \mathrm{SL}_2(\mathbb{Z})$ as above and suppose that \tilde{S} is a finite symmetric set generating a free subgroup of $\mathrm{SL}_2(\mathbb{Z})$. Then $(\mathcal{C}(\pi_p(\tilde{G}), \pi_p(\tilde{S})))_{p \text{ prime}}$ is a family of expanders.*

We shall only say a very few words about the proof. It is well-known (see [44]) that proving expansion is equivalent to showing mixing in time $\sim C \log |G|$ of the random walk on S , or in other words showing that the convolution power μ_S^n is highly uniform for $n \sim C \log |G|$. Here, $\mu_S := \frac{1}{2k} \sum_{i=1}^k (\delta_{g_i} + \delta_{g_i^{-1}})$. This is analysed in three stages: the *early stage*, where $n \leq c \log |G|$, the *middle stage* where $c \log |G| \leq n \leq \frac{1}{10} C \log |G|$, and the *late stage* where $\frac{1}{10} C \log |G| \leq n \leq C \log |G|$. During the early stage the walk behaves in a very tree-like manner, and in particular at the end of that stage it has already visited a reasonable fraction of G . The input from the theory of sets with small doubling/approximate groups comes during the middle stage: here, one must show that the walk does not get “stuck”, and that by time $\frac{1}{10} C \log |G|$ it has filled out a large portion of G . The crucial point is to show, in a certain sense, that it is impossible to have $\mu^{(n)} \approx \mu^{(2n)}$. If this *did* happen then the support of $\mu^{(n)}$ would behave very much like a set with doubling ≈ 1 , a scenario that can be ruled out using Helfgott’s classification of such sets. Finally, a little representation theory is used in the analysis of the *late stage*, specifically the fact that $\mathrm{SL}_2(\mathbb{F}_p)$ is *quasirandom* in the sense of Gowers [30], that is to say has no nontrivial irreducible representations of small dimension. This observation was first employed in a related context by Sarnak and Xue [72].

There are many open problems connected with expanders, and we refer the reader to the literature cited above. Let us just mention one (well-known) question which we hope Paul Erdős (who wrote several foundational papers in probabilistic group theory) would have liked.

Problem 6.2. Suppose that k elements g_1, g_2, \dots, g_k are selected at random from the alternating group A_n , and set $S := \{g_1^{\pm 1}, \dots, g_k^{\pm 1}\}$. Is it true that, almost surely as $n \rightarrow \infty$, S gives an expander with expansion constant $\varepsilon = \varepsilon(k) > 0$?

It could be the case that this is so even for $k = 2$. By a result of Dixon [19], g_1 and g_2 do almost surely generate A_n as $n \rightarrow \infty$, certainly a prerequisite for expansion. By a tour de force result of Kassabov [46], there does exist $k = O(1)$ and generators g_1, \dots, g_k of A_n for which S gives an expander with $\varepsilon = \varepsilon(k) > 0$, but these are not random generators. Finally, we note that in the case of the alternating group none of the three parts of the Bourgain-Gamburd argument goes through in their current form. In particular, classification of sets with small doubling in A_n (which would be needed with good quantitative bounds) is just as hard as the classification of sets with small doubling in general.

Gromov’s theorem and Varopoulos’s result. There is a close link between sets with small doubling and Gromov’s theorem on groups of polynomial

growth. Suppose that a group G is generated by a finite symmetric set S (thus $S = S^{-1}$). We say that G has *polynomial growth* if there are C and d such that $|S^n| \leq Cn^d$ for all n . Gromov [40] proved that a group has polynomial growth if and only if it is virtually nilpotent, that is to say if and only if some finite index subgroup of it is nilpotent. The link between Gromov's result and sets with small doubling is that infinitely many of the "balls" S^n will have $\sigma[S^n] < 2^d + 1$. This is very easy to see: if not, then by induction we have $|S^{2^k}| \geq c(2^d + 1)^k$, which is a contradiction for large k . By elaborating slightly on this idea, one may fairly easily show that the general theorem for sets with small doubling, Theorem 3.6, implies Gromov's theorem. Conversely, large parts of the proof of Theorem 3.6 are motivated by Gromov's argument, in particular the use of ultrafilters to construct a locally compact group (which closely parallels the Wilkie and van der Dries [84] construction of the *asymptotic cone* of a finitely-generated group).

Theorem 3.6 allows for some strengthenings of Gromov's theorem. For example ([10, Theorem 1.13]) one need only assume that $|S^n| \leq Cn^d$ for *one* value of $n > n_0(C, d)$. Several other such results are given in Section 11 of [10], where some applications to differential geometry are also discussed. A possibility, not yet realised, is that a proper understanding of sets with small doubling could be used to study groups of polynomial growth from a quantitative viewpoint. In particular the following conjecture of Grigorchuk [37, 39] remains wide open.

Problem 6.3. Is there some constant c (perhaps even $c = \frac{1}{2}$) such that the following is true: if G is generated by a symmetric set S , and if $|S^n| \leq e^{n^c}$ for all large n , then G has polynomial growth?

Famous examples of Grigorchuk [38] show that this is not true for all $c < 1$. So far, the best result known is due to Shalom and Tao [74], who show that if $|S^n| \leq n^{(\log \log n)^c}$ for large n then G has polynomial growth. (This result does not, however, make use of the connection with approximate groups.) See [37] for the state of the art on the above problem regarding special classes of groups.

One lovely application of Gromov's theorem is the following result of Varopoulos [14] (see also [87]).

Theorem 6.4. *Let G be a group generated by a finite set S . Suppose that the (simple) random walk with generating set S is recurrent. Then G is finite or has a finite-index subgroup isomorphic to \mathbb{Z} or \mathbb{Z}^2 .*

Although Varopoulos's theorem uses Gromov's theorem, it actually only uses that theorem for some value of $d > 2$. It seems, however, that no simpler

proof of the theorem is known in that special case. From the point of view of sets with small doubling, one is interested in statements about sets A with $\sigma[A] < 4 + \varepsilon$. However, no analysis of this case is known which is simpler than the general analysis of [10].

An open problem. We conclude with a very simply-stated open question. We said very little about the proof of the classification of approximate groups in general [10]. An important ingredient in it (used to establish the correspondence between approximate groups and locally compact groups) was the following result of Croot-Sisask [15] and Sanders [66].

Theorem 6.5. *Suppose that A is a K -approximate group. Then there is a set S , $|S| \gg_K |A|$, such that $S^8 \subseteq A^4$.*

Problem 6.6. In the preceding theorem, can we take $|S| \gg K^{-O(1)}|A|$?

This is open even in the abelian case.

Acknowledgments. EB is supported in part by the ERC starting grant 208091-GADA. BG is supported by an ERC starting grant. TT is supported by NSF grant DMS-0649473 and by a Simons Investigator Award.

REFERENCES

- [1] Y. Bilu. Structure of sets with small sumset. *Astérisque*, (258):xi, 77–108, 1999. Structure theory of set addition.
- [2] J. Bourgain and A. Gamburd. Uniform expansion bounds for Cayley graphs of $\mathrm{SL}_2(\mathbb{F}_p)$. *Ann. of Math. (2)*, 167(2):625–642, 2008.
- [3] J. Bourgain, N. Katz, and T. Tao. A sum-product estimate in finite fields, and applications. *Geom. Funct. Anal.*, 14(1):27–57, 2004.
- [4] E. Breuillard. A mini-course on approximate groups. *to appear in Proceedings of the Mathematical Sciences Research Institute (MSRI)*, 2013. Preprint.
- [5] E. Breuillard and B. Green. Approximate groups. I: the torsion-free nilpotent case. *J. Inst. Math. Jussieu*, 10(1):37–57, 2011.
- [6] E. Breuillard and B. Green. Approximate groups, II: The solvable linear case. *Q.J. Math.*, 62(3):513–521, 2011.
- [7] E. Breuillard, B. Green, and T. Tao. Approximate subgroups of linear groups. *Geom. Funct. Anal.*, 21(4):774–819, 2011.
- [8] E. Breuillard, B. Green, and T. Tao. A nilpotent Freiman dimension lemma. <http://arxiv.org/abs/1112.4174>, 2011. to appear in special volume of EJC in honour of Yahya Ould Hamidoune.

- [9] E. Breuillard, B. Green, and T. Tao. A note on approximate subgroups of $GL_n(\mathbb{C})$ and uniformly nonamenable groups. [arXiv:1101.2552](#), 2011. Preprint.
- [10] E. Breuillard, B. Green, and T. Tao. The structure of approximate groups. [arXiv:1110.5008](#), 2011. Preprint.
- [11] A. L. Cauchy. Recherches sur les nombres. *J. École Polytech.*, 9:99–116, 1813.
- [12] M.-C. Chang. A polynomial bound in Freiman’s theorem. *Duke Math. J.*, 113(3):399–419, 2002.
- [13] M.-C. Chang. Product theorems in SL_2 and SL_3 . *J. Inst. Math. Jussieu*, 7(1):1–25, 2008.
- [14] T. Coulhon, L. Saloff-Coste, and N. T. Varopoulos. *Analysis and geometry on groups*, volume 100 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1992.
- [15] E. Croot and O. Sisask. A probabilistic technique for finding almost-periods of convolutions. *Geom. Funct. Anal.*, 20(6):1367–1396, 2010.
- [16] H. Davenport. On the addition of residue classes. *J. London Math. Soc.*, 10:30–32, 1935.
- [17] M. DeVos. The structure of critical product sets. [arXiv:1301.0096](#), 2013. Preprint.
- [18] O. Dinai. Expansion properties of finite simple groups. [arXiv:1001.5069](#), 2010. Preprint.
- [19] J. D. Dixon. The probability of generating the symmetric group. *Math. Z.*, 110:199–205, 1969.
- [20] G. Elekes and Z. Király. On the combinatorics of projective mappings. *J. Algebraic Combin.*, 14(3):183–197, 2001.
- [21] C. Even-Zohar. On sums of generating sets in $(\mathbb{Z}_2)^n$. *Combin. Probab. Comput.*, 21(6):916–941, 2012.
- [22] D. Fisher, N. H. Katz, and I. Peng. Approximate multiplicative groups in nilpotent Lie groups. *Proc. Amer. Math. Soc.*, 138(5):1575–1580, 2010.
- [23] G. A. Freiman. *Foundations of a structural theory of set addition*. American Mathematical Society, Providence, R. I., 1973. Translated from the Russian, Translations of Mathematical Monographs, Vol 37.
- [24] G. A. Freĭman. Groups and the inverse problems of additive number theory. In *Number-theoretic studies in the Markov spectrum and in the structural theory of set addition (Russian)*, pages 175–183. Kalinin. Gos. Univ., Moscow, 1973.
- [25] G. A. Freiman. On finite subsets of nonabelian groups with small doubling. *Proc. Amer. Math. Soc.*, 140(9):2997–3002, 2012.
- [26] R. J. Gardner. The Brunn-Minkowski inequality. *Bull. Amer. Math. Soc. (N.S.)*, 39(3):355–405, 2002.
- [27] N. Gill and H. Helfgott. Growth in solvable subgroups of $GL_r(\mathbb{Z}/p\mathbb{Z})$. [arXiv:1008.5264](#), 2010. Preprint.
- [28] A. S. Golesefidy and P. P. Varjú. Expansion in perfect groups. *Geom. Funct. Anal.*, 22(6):1832–1891, 2012.
- [29] W. T. Gowers. A new proof of Szemerédi’s theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

- [30] W. T. Gowers. Quasirandom groups. *Combin. Probab. Comput.*, 17(3):363–387, 2008.
- [31] B. Green. Edinburgh lecture notes on Freiman’s theorem.
- [32] B. Green. Finite field models in additive combinatorics. In *Surveys in combinatorics 2005*, volume 327 of *London Math. Soc. Lecture Note Ser.*, pages 1–27. Cambridge Univ. Press, Cambridge, 2005.
- [33] B. Green and I. Z. Ruzsa. Freiman’s theorem in an arbitrary abelian group. *J. Lond. Math. Soc. (2)*, 75(1):163–175, 2007.
- [34] B. Green and T. Tao. Compressions, convex geometry and the Freiman-Bilu theorem. *Q.J. Math.*, 57(4):495–504, 2006.
- [35] B. Green and T. Tao. Freiman’s theorem in finite fields via extremal set theory. *Combin. Probab. Comput.*, 18(3):335–355, 2009.
- [36] B. Green and T. Tao. An equivalence between inverse sumset theorems and inverse conjectures for the U^3 norm. *Math. Proc. Cambridge Philos. Soc.*, 149(1):1–19, 2010.
- [37] R. Grigorchuk. On the Gap Conjecture concerning group growth. <http://arxiv.org/abs/1202.6044>, 2012.
- [38] R. I. Grigorchuk. On Burnside’s problem on periodic groups. *Funktsional. Anal. i Prilozhen.*, 14(1):53–54, 1980.
- [39] R. I. Grigorchuk. On growth in group theory. In *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990)*, pages 325–338, Tokyo, 1991. Math. Soc. Japan.
- [40] M. Gromov. Groups of polynomial growth and expanding maps. *Inst. Hautes Études Sci. Publ. Math.*, (53):53–73, 1981.
- [41] Y. O. Hamidoune. Two inverse results. [arXiv:1006.5074](https://arxiv.org/abs/1006.5074), 2010. Preprint.
- [42] H. A. Helfgott. Growth and generation in $SL_2(\mathbb{Z}/p\mathbb{Z})$. *Ann. of Math. (2)*, 167(2):601–623, 2008.
- [43] H. A. Helfgott. Growth in $SL_3(\mathbb{Z}/p\mathbb{Z})$. *J. Eur. Math. Soc. (JEMS)*, 13(3):761–851, 2011.
- [44] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561 (electronic), 2006.
- [45] E. Hrushovski. Stable group theory and approximate subgroups. *J. Amer. Math. Soc.*, 25(1):189–243, 2012.
- [46] M. Kassabov. Symmetric groups and expanders. *Electron. Res. Announc. Amer. Math. Soc.*, 11:47–56 (electronic), 2005.
- [47] M. Kneser. Abschätzung der asymptotischen Dichte von Summenmengen. *Math. Z.*, 58:459–484, 1953.
- [48] S. V. Konyagin. On Freiman’s theorem in finite fields. *Mat. Zametki*, 84(3):472–474, 2008.
- [49] M. J. Larsen and R. Pink. Finite subgroups of algebraic groups. *J. Amer. Math. Soc.*, 24(4):1105–1158, 2011.
- [50] V. F. Lev and P. Y. Smeliansky. On addition of two distinct sets of integers. *Acta Arith.*, 70(1):85–91, 1995.

- [51] S. Lovett. Equivalence of polynomial conjectures in additive combinatorics. [arXiv:1001.3356](#), 2010. Preprint.
- [52] A. Lubotzky. Expander graphs in pure and applied mathematics (notes for the colloquium lectures, AMS annual meeting 2001).
- [53] A. Lubotzky. Cayley graphs: eigenvalues, expanders and random walks. In *Surveys in combinatorics, 1995 (Stirling)*, volume 218 of *London Math. Soc. Lecture Note Ser.*, pages 155–189. Cambridge Univ. Press, Cambridge, 1995.
- [54] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.
- [55] G. A. Margulis. Explicit constructions of expanders. *Problemy Peredači Informacii*, 9(4):71–80, 1973.
- [56] G. Petridis. Plünnecke’s inequality. *Combin. Probab. Comput.*, 20(6):921–938, 2011.
- [57] H. Plünnecke. *Eigenschaften und Abschätzungen von Wirkungsfunktionen*. BMWF-GMD-22. Gesellschaft für Mathematik und Datenverarbeitung, Bonn, 1969.
- [58] L. Pyber and E. Szabó. Growth in finite simple groups of Lie type of bounded rank. [arXiv:1005.1858](#), 2010. Preprint.
- [59] L. Pyber and E. Szabó. Growth in linear groups. *to appear in Proceedings of the Mathematical Sciences Research Institute (MSRI)*, 2013. Preprint.
- [60] A. Razborov. A product theorem in free groups.
- [61] I. Z. Ruzsa. Generalized arithmetical progressions and sumsets. *Acta Math. Hungar.*, 65(4):379–388, 1994.
- [62] I. Z. Ruzsa. Sums of finite sets. In *Number theory (New York, 1991–1995)*, pages 281–293. Springer, New York, 1996.
- [63] I. Z. Ruzsa. An analog of Freiman’s theorem in groups. *Astérisque*, (258):xv, 323–326, 1999. Structure theory of set addition.
- [64] I. Z. Ruzsa. Sumsets and structure. In *Combinatorial number theory and additive group theory*, Adv. Courses Math. CRM Barcelona, pages 87–210. Birkhäuser Verlag, Basel, 2009.
- [65] S. R. Safin. Powers of subsets of free groups. *Mat. Sb.*, 202(11):97–102, 2011.
- [66] T. Sanders. On a nonabelian Balog-Szemerédi-type lemma. *J. Aust. Math. Soc.*, 89(1):127–132, 2010.
- [67] T. Sanders. An analytic approach to a weak non-Abelian Kneser-type theorem. [arXiv:1212.0457](#), 2012. Preprint.
- [68] T. Sanders. Approximate groups and doubling metrics. *Math. Proc. Cambridge Philos. Soc.*, 152(3):385–404, 2012.
- [69] T. Sanders. On the Bogolyubov-Ruzsa lemma. *Anal. PDE*, 5(3):627–655, 2012.
- [70] T. Sanders. The structure theory of set addition revisited. *Bull. Amer. Math. Soc. (N.S.)*, 50(1):93–127, 2013.
- [71] P. Sarnak. What is... an expander? *Notices Amer. Math. Soc.*, 51(7):762–763, 2004.
- [72] P. Sarnak and X. X. Xue. Bounds for multiplicities of automorphic representations. *Duke Math. J.*, 64(1):207–227, 1991.
- [73] T. Schoen. Near optimal bounds in Freiman’s theorem. *Duke Math. J.*, 158(1):1–12, 2011.

- [74] Y. Shalom and T. Tao. A finitary version of Gromov’s polynomial growth theorem. *Geom. Funct. Anal.*, 20(6):1502–1547, 2010.
- [75] Y. Stanchescu. On the structure of sets with small doubling property on the plane. I. *Acta Arith.*, 83(2):127–141, 1998.
- [76] Y. Stanchescu. On the structure of sets with small doubling property on the plane. II. *Integers*, 8(2):20, 2008.
- [77] T. Tao. *Expansion in groups of Lie type*. Manuscript in preparation.
- [78] T. Tao. *Hilbert’s fifth problem and related topics*. Manuscript in preparation.
- [79] T. Tao. Product set estimates for non-commutative groups. *Combinatorica*, 28(5): 547–594, 2008.
- [80] T. Tao. Freiman’s theorem for solvable groups. *Contrib. Discrete Math.*, 5(2):137–184, 2010.
- [81] T. Tao. Noncommutative sets of small doubling. [arXiv:1106.2267](https://arxiv.org/abs/1106.2267), 2011. Preprint.
- [82] T. Tao and V. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.
- [83] M. Tointon. Freiman’s theorem in an arbitrary nilpotent group. [arXiv:1211.3989](https://arxiv.org/abs/1211.3989), 2012. Preprint.
- [84] L. van den Dries and A. J. Wilkie. Gromov’s theorem on groups of polynomial growth and elementary logic. *J. Algebra*, 89(2):349–374, 1984.
- [85] P. P. Varjú. Expansion in $SL_d(O_K/I)$, I square-free. *J. Eur. Math. Soc. (JEMS)*, 14(1):273–305, 2012.
- [86] A. G. Vosper. The critical pairs of subsets of a group of prime order. *J. London Math. Soc.*, 31:200–205, 1956.
- [87] W. Woess. *Random walks on infinite graphs and groups*, volume 138 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2000.

Emmanuel Breuillard

*Laboratoire de Mathématiques,
Bâtiment 425,
Université Paris Sud 11,
91405 Orsay
France*

e-mail:

`emmanuel.breuillard@math.u-psud.fr`

Ben Green

*Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WA
England*

e-mail: `b.j.green@dpmms.cam.ac.uk`

Terence Tao

*Department of Mathematics, UCLA,
405 Hilgard Ave,
Los Angeles CA 90095,
USA*

e-mail: `tao@math.ucla.edu`

ERDŐS AND MULTIPLICATIVE NUMBER THEORY

HAROLD G. DIAMOND

In memory of E. P.

1. INTRODUCTION

Paul Erdős was a prolific writer of letters as well as articles. Along with many other mathematicians in areas such as number theory, combinatorics, and set theory, I was on his “mailing list.” Paul’s letters arrived several times a year from mathematics centers near and far. They typically began, *I hope you are well and things are OK in Samland. I am visiting A right now, and leave next week to preach in B. Let $f(n)$ be a function* This article reviews some of the topics we discussed: estimates of prime number counts, distribution questions for the Euler φ function, and elementary methods in prime number theory.

Also, I shall examine in detail Erdős’ lower bound for primes in a short interval, a result which played a key role in the first elementary proof of the prime number theorem (PNT). As C. L. Siegel noted in his Zentralblatt review of Erdős’ paper [10], the result is valid, but the gaps and inaccuracies make it hard to understand. This is a good place to put on record a more accessible version of the short-interval result.

2. CHEBYSHEV PRIME COUNTING ESTIMATES

To begin, here are two articles of Erdős relating to P. L. Chebyshev’s famous approximation to the PNT. With $\pi(x)$ denoting the number of primes not exceeding x , Chebyshev found

$$x/\log x \ll \pi(x) \ll x/\log x,$$

inequalities that are today named for him.

One of Erdős' earliest papers, *Beweis eines Satzes von Tchebyschef* [8], gives simpler proofs of such bounds and of the so-called Bertrand Postulate. The "Postulate" asserts that there exists at least one prime in each interval $(n, 2n]$ for $n \in \mathbb{N}$. We recount the main ideas of his clever and interesting argument.

The prime upper bound is based on two observations about the binomial coefficient

$$(1) \quad \binom{2a}{a} = \frac{(2a)!}{a! a!}.$$

On the one hand, it is divisible by each prime $p \in (a, 2a]$, since these primes divide the numerator but not the denominator. On the other hand, the binomial coefficient is smaller than 4^{a-1} for $a = 5$, and by induction this inequality holds for each integer $a > 5$. These two facts together imply that $\vartheta(x) := \sum_{p \leq x} \log p$ satisfies

$$(2) \quad \vartheta(2a) - \vartheta(a) = \log \left\{ \prod_{a < p \leq 2a} p \right\} < (a-1) \log 4$$

for each number $a \geq 5$; by inspection this relation holds also for all $a \geq 2$. Adding together differences $\vartheta(2^{n+1}) - \vartheta(2^n)$ and making simple inequalities, we get, for all positive x , Erdős' upper bound

$$\vartheta(x) < x \log 4.$$

The upper bound for $\pi(x)$ follows by summation by parts.

The proof of Bertrand's Postulate and the lower bound estimate for $\vartheta(x)$ start with A. M. Legendre's formula for $\alpha(p)$, the exact power of p that divides $\binom{2a}{a}$. We have

$$\alpha(p) = \sum_{k \geq 1} \left(\left[\frac{2a}{p^k} \right] - 2 \left[\frac{a}{p^k} \right] \right) \leq \left[\frac{\log 2a}{\log p} \right],$$

since each summand is either 0 or 1 and all are 0 for $k > (\log 2a)/(\log p)$. Thus $p^{\alpha(p)} \leq 2a$, and if $p > \sqrt{2a}$, then $\alpha(p) \leq 1$. For $a \geq 2$ and all primes $p \in (2a/3, a]$ we have $[2a/p] - 2[a/p] = 0$ and so, by another application of Legendre's formula, such primes do not divide $\binom{2a}{a}$.

It follows that

$$\binom{2a}{a} \leq \prod_{p \leq \sqrt{2a}} 2a \prod_{\sqrt{2a} < p \leq 2a/3} p \prod_{a < p \leq 2a} p.$$

The first product on the right contains at most $\sqrt{2a}$ factors. Using the $\vartheta(x)$ upper bound, the second product is at most

$$\prod_{p \leq 2a/3} p < 4^{2a/3}.$$

Also

$$2a \binom{2a}{a} = \frac{2}{1} \cdot \frac{3}{1} \cdot \frac{4}{2} \cdot \frac{5}{2} \cdots \frac{2a-2}{a-1} \cdot \frac{2a-1}{a-1} \cdot \frac{2a}{a} \cdot \frac{2a}{a} > 2^{2a}.$$

Thus

$$2^{2a} < (2a)^{1+\sqrt{2a}} 2^{4a/3} \prod_{a < p \leq 2a} p$$

or

$$(3) \quad 2^{2a/3} < (2a)^{1+\sqrt{2a}} \prod_{a < p \leq 2a} p.$$

If there are no primes in an interval $(a, 2a]$, then the last product is empty and (3) becomes $2^{2a/3} < (2a)^{1+\sqrt{2a}}$, a relation that cannot hold for sufficiently large a , e.g. for $a \geq 500$. Thus Bertrand’s Postulate holds for all sufficiently large numbers a , and by inspection, it is found to hold for the entire range.

We can deduce a lower bound for $\vartheta(x)$ by another application of (3). Taking logarithms, we have

$$\vartheta(2a) - \vartheta(a) > (a/3) \log 4 - (1 + \sqrt{2a}) \log(2a) \gg a.$$

Again adding up theta differences, we find that $\vartheta(x) \gg x$ and hence $\pi(x) \gg x/\log x$ for $x \geq 2$.

For comparison and our use below, here are the classical bounds of Chebyshev: With $b := \log\{2^{1/2} 3^{1/3} 5^{1/5}/30^{1/30}\}$, he found

$$(4) \quad \liminf_{x \rightarrow \infty} \vartheta(x)/x \geq b = 0.921292\dots,$$

$$\limsup_{x \rightarrow \infty} \vartheta(x)/x \leq 6b/5 = 1.105550\dots$$

The second article with a Chebyshev theme is a joint work of ours, *On Sharp Elementary Prime Number Estimates* [5], in which it is shown that, in principle, Chebyshev could have given explicit estimates of $\pi(x)/\{x/\log x\}$ that are arbitrarily close to 1. Are we saying that the PNT could be proved this way? The answer is No; using the variant argument we describe, one could make estimates of the ratio close to 1. However, any finite calculation would have some “wobble,” and to show that the ratio can be made *arbitrarily* close to 1 depends on a form of the PNT. The precise statement of the result is as follows.

Theorem 1. *Let $\varepsilon > 0$ be given. There exists a positive integer $T = T(\varepsilon)$ and a construction using the values of the Moebius μ function on just the interval $[1, T)$ that yields the estimate*

$$\limsup_{x \rightarrow \infty} |\pi(x)/\{x/\log x\} - 1| < \varepsilon.$$

According to Erdős, the theorem has a long and curious history. It was discovered in 1937 by him and László Kalmár and about the same time by J. B. Rosser. Erdős and Kalmár decided not to publish their article when they learned that Rosser had already submitted for publication a manuscript that treated also primes in arithmetic progression. Unfortunately, the referee of Rosser's article was an expert in another of his research areas who may not have appreciated the point of this piece and rejected it. The theorem lived only by word of mouth until we published this version.

Chebyshev's method used an approximation to the Moebius μ function that can be described as follows. For k and n positive integers, define an arithmetic function

$$e_k(n) = \begin{cases} 1, & \text{if } n = k, \\ 0, & \text{if } n \neq k, \end{cases}$$

and set

$$\mu_C := e_1 - e_2 - e_3 - e_5 + e_{30}.$$

This function has the properties that $\sum_n \mu_C(n)/n = 0$,

$$0 \leq F_C(x) := \sum_{n \leq x} (1 * \mu_C)(n) \leq 1, \quad x \geq 1,$$

(* denotes multiplicative convolution) and $F_C(x) = 1$ for $1 \leq x < 6$.

The function underlying the present method is an elaboration of μ_C . For T and n positive integers, take as our finite approximation of μ

$$\mu_T(n) = \begin{cases} \mu(n), & \text{if } 1 \leq n < T, \\ -T \sum_{i < T} \mu(i)/i, & \text{if } n = T, \\ 0, & \text{if } n > T. \end{cases}$$

By its construction $\sum_n \mu_T(n)/n = 0$ and, by the main property of μ , $F_T(x) := \sum_{n \leq x} (1 * \mu_T)(n) = 1$ for $1 \leq x < T$.

Chebyshev's weighted prime counting function $\psi(x) := \sum_{p^i \leq x} \log p$ is expressed in terms of $L(n) := \log n$ and $\mu(n)$ as

$$\psi(x) = \sum_{n \leq x} (L * \mu)(n).$$

The PNT is equivalent to $\psi(x) \sim x$, as is the case for $\vartheta(x)$.

Our method is to replace $\psi(x)$ by a function $\psi_T(x)$, created by using μ_T in place of μ above, and to show that for suitable large T , $\psi_T(x)$ is close to both $\psi(x)$ and x . The proof that $\psi_T(x) \sim x$ as $T \rightarrow \infty$ (along a special sequence of T values) is unconditional; the PNT is invoked to show that $\psi_T(x) \sim \psi(x)$.

3. ELEMENTARY PROOF OF THE PRIME NUMBER THEOREM

The main goal in studying $\pi(x)$ from Chebyshev's time onward was to prove the celebrated PNT, i.e. that $\pi(x) \sim x/\log x$ as $x \rightarrow \infty$. Repeated attempts were made to establish the PNT by extending Chebyshev's real variable methods, which we today call "elementary." (This term does not suggest they are simple!) However, decades of work yielded only small numerical improvements, but not the PNT itself. A survey of such elementary methods is given in [3].

In another direction, Bernhard Riemann laid out a program for studying $\pi(x)$ by using properties of an analytic function, today called the Riemann zeta function $\zeta(s)$. These ideas, in the hands of Jacques Hadamard and others, led to a proof of the PNT some 40 years later. There were two reasons for believing that analytic methods provided the *only* path to the PNT. First, on a practical level, elementary methods had not made significant progress, despite considerable effort. Second, it was long known that the truth of the PNT implies that the Riemann zeta function $\zeta(s)$ does not vanish anywhere on the line $\Re s = 1$ and, in the other direction, the work of Wiener showed that the PNT was a consequence of this single – very analytic – relation.

About 50 years after the PNT was first proved, Atle Selberg established, in the course of his research in sieve theory, the formula

$$(5) \quad \vartheta(x) \log x + \sum_{p \leq x} \vartheta(x/p) \log p = 2x \log x + O(x),$$

where, again, $\vartheta(x)$ denotes the logarithmically weighted prime-counting function. Today, this relation is called Selberg's Formula and is regarded as fundamental in elementary prime number theory.

It was quickly noted that (5) contains significant information about the primes. As one example, dropping the sum term in (5) yields the inequality $\vartheta(x) \log x \leq 2x \log x + O(x)$, which shows the true order of magnitude of $\vartheta(x)$ and hence (by summation by parts) that of $\pi(x)$.

Next, let us set

$$\limsup_{x \rightarrow \infty} \vartheta(x)/x = A \quad \text{and} \quad \liminf_{x \rightarrow \infty} \vartheta(x)/x = a.$$

Now (5) and the familiar prime number estimate

$$(6) \quad \sum_{p \leq x} \frac{\log p}{p} \sim \log x$$

give the interesting and useful relation

$$(7) \quad A + a = 2.$$

Here is a sketch of how (7) is proved. In place of (5) write

$$(5') \quad \frac{\vartheta(x)}{x} + \frac{1}{x \log x} \sum_{p \leq x} \vartheta(x/p) \log p = 2 + o(1),$$

and insert the upper bound $\vartheta(x/p) \leq \{A + o(1)\}x/p$ into (5'). Approximating the resulting sum by (6), we find that

$$\frac{\vartheta(x)}{x} + A \geq 2 + o(1)$$

holds for all large values of x . Now let $x \rightarrow \infty$ along a sequence on which $\vartheta(x)/x \rightarrow a$, and we see that $a + A \geq 2$. Next, use the bound $\vartheta(x/p) \geq (a + o(1))x/p$ in (5') and proceed analogously to show that $A + a \leq 2$. Together, the inequalities give (7).

As a third example of the utility of (5), write it with argument $x + cx$ in place of x and then subtract (5) from it. Using $\log(x + cx) = \log x + O(1)$ and the Chebyshev bound $\vartheta(x) = O(x)$, we find that

$$(8) \quad \left\{ \vartheta(x + cx) - \vartheta(x) \right\} \log x \\ + \sum_{p \leq x+cx} \left\{ \vartheta\left(\frac{x+cx}{p}\right) - \vartheta\left(\frac{x}{p}\right) \right\} \log p = 2cx \log x + O(x).$$

If we drop the sum, we obtain the short interval estimate

$$(9) \quad \vartheta(x + cx) - \vartheta(x) \leq 2cx + O(x/\log x).$$

(We can assume here that $c \in (0, 1)$, but by Chebyshev's bounds (4), the last inequality holds for larger c as well.)

Upon seeing Selberg's formula, Erdős asserted that it could yield an extension of the Bertrand Postulate to intervals $[x, x + cx]$ for arbitrarily small $c > 0$ for all sufficiently large x . More colorfully, his claim is that the ratio of consecutive primes approaches 1 at infinity. In fact, Erdős had proved (see [10]) rather more than a short interval Bertrand result, namely, an inequality in the opposite direction from (9):

Theorem 2. For $c > 0$ (no matter how small),

$$(10) \quad \vartheta(x + cx) - \vartheta(x) \gg_c x.$$

Equivalently, there exist positive numbers $\delta = \delta(c)$ and $X = X(c)$ such that for all $x > X$,

$$\pi(x + cx) - \pi(x) > \delta x / \log x.$$

This was the first elementary lower estimate of the number of primes in such short intervals. (Differencing Chebyshev's bounds yields only

$$(11) \quad \liminf_{x \rightarrow \infty} x^{-1} \{ \vartheta((1.2 + \eta)x) - \vartheta(x) \} > \eta b > 0.92\eta$$

for any $\eta > 0$.) Theorem 2 was used by Selberg in his first elementary proof of the PNT and was the heart of the proof Erdős published in [10]. (Selberg subsequently found another argument, which he used in [14].) For an insightful analysis of these proofs see the Mathematical Reviews article of Ingham [13].

Erdős' path to the PNT uses a subtle contradiction argument, made the more challenging by typos and some errors and omissions. We present here his proof of this key relation, both for its interest and to make available a more readable account. A mark “(†)” appears at points where material changes have been made. We shall assume the Selberg Formula (5) and its two consequences, (7) and (9).

The argument is an elaboration of ingredients in the proof of (7): if $\vartheta(x)/x$ is too small somewhere, then it has to be too large elsewhere. We henceforth assume that (10) does not hold, specifically, that there is a number $C > 0$ and a sequence of values $x \rightarrow \infty$ along which

$$(12) \quad \vartheta(x + Cx) - \vartheta(x) = o(x).$$

We know from (11) that $C \leq 0.2$. We further assume that C is so near the supremum of numbers for which the preceding o -relation holds that it satisfies the following technical condition: for a number $t > 0$ that is small in comparison to C we have

$$(13) \quad \vartheta((x + Cx)(1 + t)) - \vartheta(x) > 2\eta x$$

for some fixed number $\eta > 0$ and all sufficiently large x .

If we combine (12) with the Selberg Formula (8), it follows that

$$(14) \quad \sum_{p \leq x + Cx} \left\{ \vartheta\left(\frac{x + Cx}{p}\right) - \vartheta\left(\frac{x}{p}\right) \right\} \log p \sim 2Cx \log x$$

holds along the same x sequence. The first step in the argument is to show that the individual summands in (14) are large for “most” primes p . This result will be used later to show that $\vartheta(x)/x$ would then be larger than Chebyshev’s upper bound (4) allows.

Lemma 1. *Assume (12). Then for all primes $p \leq x$, except possibly for a subset $\mathcal{I} = \mathcal{I}(x) \subset [1, x]$ for which*

$$(15) \quad \sum_{p \in \mathcal{I}} \frac{\log p}{p} = o(\log x),$$

we have

$$(16) \quad \vartheta((x + Cx)/p) - \vartheta(x/p) = 2Cx/p + o(x/p).$$

Proof. First, note that (9) insures that (16) holds for *all* $p \leq x$ with “ \leq ” in place of “ $=$ ”. Suppose, by way of a contradiction, that strict inequality holds in (16) for a “substantial” set of primes. That is, there exist positive constants b_1 and b_2 with the following property: for any large x for which (12) holds, there exists a set of primes $\mathcal{J} = \mathcal{J}(x) \subset [1, x]$ for which

$$\sum_{p \in \mathcal{J}} \frac{\log p}{p} \geq b_1 \log x$$

(i.e., \mathcal{J} is substantial) and for each prime $p \in \mathcal{J}$

$$\vartheta((x + Cx)/p) - \vartheta(x/p) < (2C - b_2)x/p.$$

Let S denote the left side of (14), and write S as a sum over \mathcal{J} plus one over $\mathcal{J}' := \{p \in [1, x]\} \setminus \mathcal{J}$. Using the hypothesis, the upper estimate (9), and the sum over primes (6), we find

$$\begin{aligned} S &\leq \sum_{p \in \mathcal{J}} (2C - b_2)(x/p) \log p + \sum_{p \in \mathcal{J}'} 2C(x/p) \log p + o(x \log x) \\ &\leq (2C - b_1 b_2)x \log x + o(x \log x). \end{aligned}$$

This contradicts (14), and thus (16) must hold for all primes p outside a small set. ■

(†) We should say a word about asymptotic or o -relations involving x/p . Suppose a function $f(y)$ in our analysis is within a preassigned number $\varepsilon > 0$ of its limit (in an appropriate sense) whenever $y \geq K$ for some suitably large number K . How do we proceed when $x/p < K$, which puts $f(x/p)$ outside its “good” range? In this case we have

$$\sum_{x/K < p \leq x} \frac{\log p}{p} = o(\log x),$$

and such primes can simply be adjoined to the set \mathcal{I} of the last lemma.

The primes in $[1, x]$ satisfying (16) we shall call *good primes* and the remaining ones *bad primes* (with the understanding that goodness or badness depends on x (†) and the tolerance level implicit in $o(x/p)$). We shall prove that a sequence of good primes $p_1 < p_2 < \dots < p_k$ exists satisfying the conditions

$$(17) \quad (\dagger) \quad p_1 = o(x),$$

$$(18) \quad 10p_1 < p_k < 100p_1,$$

and

$$(19) \quad (1 + C)(1 + t)^2 p_i > p_{i+1} > (1 + t)p_i, \quad i = 1, 2, \dots, k - 1,$$

where t is the fixed positive number that was selected for (13). For primes p_i satisfying (18) and (19) we must have $(1 + t)^{k-1} p_1 < p_k < 100p_1$; it follows that $k < k_0$, where $k_0 = k_0(t)$.

Assuming for the moment that a sequence of primes satisfying (18) and (19) exists, we now prove Theorem 2. At this point, Erdős makes a lower estimate of

$$\sum_{i=1}^{k-1} \left\{ \vartheta \left(\frac{x}{p_i} (1 + C) \right) - \vartheta \left(\frac{x}{p_{i+1}} \right) \right\}.$$

These terms can represent overlapping intervals, so it is not clear how to use his estimate. (†) Instead, we show that $\vartheta(x/p_i) - \vartheta(x/p_{i+1})$ is large for each i ; an estimate with these summands is clearly valid and will lead to a value for $\vartheta(x/p_1)$ that is too large.

Consider two intervals

$$(20) \quad \left[\frac{x}{p_{i+1}}, \frac{x}{p_{i+1}}(1+C) \right], \quad \left[\frac{x}{p_i}, \frac{x}{p_i}(1+C) \right].$$

Suppose first that the intervals overlap. Then, using also (19),

$$\frac{x}{p_{i+1}}(1+t) < \frac{x}{p_i} \leq \frac{x}{p_{i+1}}(1+C).$$

We claim that

$$(21) \quad \vartheta\left(\frac{x}{p_i}\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) = 2\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right) + o\left(\frac{x}{p_i}\right).$$

By (9), the last relation holds with “ \leq .” If the inequality were strict, there would be a number $c_1 > 0$ such that

$$\vartheta\left(\frac{x}{p_i}\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) < (2 - c_1)\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right).$$

But since p_{i+1} is a good prime, (16) holds with $p = p_{i+1}$; therefore, in the remaining portion of the interval $[x/p_{i+1}, (1+C)x/p_{i+1}]$ we have

$$\begin{aligned} \vartheta\left(\frac{x}{p_{i+1}}(1+C)\right) - \vartheta\left(\frac{x}{p_i}\right) &> 2C\frac{x}{p_{i+1}} + o\left(\frac{x}{p_{i+1}}\right) - (2 - c_1)\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right) \\ &= 2\left\{\frac{x}{p_{i+1}}(1+C) - \frac{x}{p_i}\right\} + c_1\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right) + o\left(\frac{x}{p_{i+1}}\right). \end{aligned}$$

By the last inequality in (19) and a small manipulation,

$$c_1\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right) > \frac{c_1 t}{C-t}\left\{\frac{x}{p_{i+1}}(1+C) - \frac{x}{p_i}\right\}.$$

It follows that

$$\vartheta\left(\frac{x}{p_{i+1}}(1+C)\right) - \vartheta\left(\frac{x}{p_i}\right) > (2 + c_2)\left\{\frac{x}{p_{i+1}}(1+C) - \frac{x}{p_i}\right\},$$

with $c_2 = c_1 t / (C - t) > 0$, in violation of (9). Thus (21) holds if the intervals overlap.

On the other hand, if the intervals (20) do not overlap, we have

$$\vartheta\left(\frac{x}{p_i}\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) \geq \vartheta\left(\frac{x}{p_{i+1}}(1 + C)\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) = \frac{(2C + o(1))x}{p_{i+1}},$$

with the equality arising from the goodness of p_{i+1} . We claim that $2Cx/p_{i+1}$ is not far from $2(x/p_i - x/p_{i+1})$. Indeed, by the inequality $p_{i+1} < (1 + C)(1 + t)^2 p_i$, from (19), we have

$$1.9\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right) < 1.9\left\{(1 + C)(1 + t)^2 - 1\right\} \frac{x}{p_{i+1}} < \frac{1.95Cx}{p_{i+1}}$$

for any $C > 0$ and sufficiently small t .

In either case we have

$$\vartheta\left(\frac{x}{p_i}\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) > 1.9\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right)$$

for $1 \leq i \leq k - 1$. Now add together these inequalities and recall that $p_k > 10p_1$. On a sequence $x \rightarrow \infty$ we obtain

$$\vartheta\left(\frac{x}{p_1}\right) \geq \vartheta\left(\frac{x}{p_1}\right) - \vartheta\left(\frac{x}{p_k}\right) > 1.9\left(\frac{x}{p_1} - \frac{x}{p_k}\right) > 1.7x/p_1.$$

(†) By (17), x/p_1 is unbounded as $x \rightarrow \infty$, so the preceding ϑ inequality violates the Chebyshev bound $\limsup \vartheta(x)/x < 1.11$. This concludes the proof of the theorem under the assumption that there is a sequence of good primes satisfying (17), (18), and (19).

Now we show that such good primes exist. Consider the intervals

$$I_r := (B^{2r}, B^{2r+1}), \quad r = 0, 1, \dots, \left\lfloor \frac{\log x}{2 \log B} \right\rfloor - 1,$$

where B is a fixed, sufficiently large number. Clearly, all the intervals I_r lie in $(0, x)$.

(†) First we show that for “nearly all” (i.e. with the exception of at most $o(\log x)$) indices r , the interval I_r contains at least one good prime lying in $I'_r := (B^{2r}, B^{2r+1/3})$. From Chebyshev’s bounds (4) we have, for each r ,

$$\sum_{p \in I'_r} \frac{\log p}{p} > \frac{\vartheta(B^{2r+1/3}) - \vartheta(B^{2r})}{B^{2r+1/3}}$$

$$> \frac{0.92B^{2r+1/3} - 1.11B^{2r}}{B^{2r+1/3}} > 0.92 - \frac{1.11}{100^{1/3}} =: c_1,$$

provided that $B > 100$, say. If for some positive c_2 , there were $c_2 \log x$ intervals $I'_r \subset [1, x]$ without good primes, then we would have

$$\sum_{p \text{ bad}} \frac{\log p}{p} > c_1 c_2 \log x,$$

which contradicts (15).

Next, we show that (18) and (19) hold for nearly all intervals I_r . Let $p_1^{(r)}$ be the smallest good prime in I'_r (assuming such exists), and suppose that a sequence $p_1^{(r)}, p_2^{(r)}, \dots, p_i^{(r)} \in I_r$ satisfying (19) exists with $p_i^{(r)}(1+t)^2(1+C) < B^{2r+1}$. Further, suppose that the sequence terminates too soon in the sense that no $p_{i+1}^{(r)}$ can be found in I_r , because all the primes in

$$J_i^{(r)} := [p_i^{(r)}(1+t), p_i^{(r)}(1+t)^2(1+C)]$$

are bad. In this case we say that the interval I_r has an *obstruction*.

We have

$$(22) \quad (1+C)(1+t)^2 < 2,$$

since, as we noted above, $C \leq 0.20$ and we can take t small. Now

$$p_i^{(r)}(1+t)^2(1+C) < 2B^{2r+1} < B^{2r+2},$$

so the intervals $J_{i_1}^{(r_1)}, J_{i_2}^{(r_2)}, \dots$ do not overlap. Also, we have from (13) that

$$\sum_{p \in J_i^{(r)}} \log p > 2\eta p_i^{(r)}(1+t) > \eta p_i^{(r)}(1+t)^2(1+C),$$

and thus

$$(23) \quad \sum_{p \in J_i^{(r)}} \frac{\log p}{p} > \eta.$$

(†) Suppose that there were at least $c_3 \log x$ intervals I_r having an obstruction, for some $c_3 > 0$. In this case, it follows from (23) that

$$\sum_{p \text{ bad}} \frac{\log p}{p} > c_3 \eta \log x,$$

contradicting (15). Thus, nearly every interval I_r contains a sequence of good primes without an obstruction.

(†) We can achieve the second inequality in (18) for every interval I_r that contains good primes simply by trimming from the sequence any good primes $p \geq 100p_1$. (The only role the inequality serves is to insure that, for a fixed choice of the parameter t , the number of primes in our sequence, is bounded.)

(‡) For the first inequality of (18), suppose that I_r has at least one good prime satisfying $p_1 < B^{2r+1/3}$ but that p_k , the largest prime of the sequence, satisfies

$$p_k \leq 10p_1 < 10B^{2r+1/3} < (1/2)B^{2r+1}$$

(for the last inequality, recall $B > 100$). Further, by (22), we have

$$(1 + C)(1 + t)^2 p_k < B^{2r+1}.$$

Since there is no good prime in $[(1 + t)p_k, (1 + C)(1 + t)^2 p_k]$ despite there being room for one, I_r has an obstruction. The collection of intervals having obstructions is sparse, as we have seen, so (18) holds for nearly all intervals I_r .

Finally, we establish (17), which guarantees a large argument x/p_1 in our Chebyshev-type bound, justifying the punchline in the proof of Theorem 2. There are $\lfloor \log x / (2 \log B) \rfloor$ intervals I_r , of which at most $o(\log x)$ either lie near x or lack a sequence of primes satisfying (18) and (19). Thus there exists an index $r^* < \log x / (5 \log B)$ for which I_{r^*} contains a good sequence, and, for large x , the first prime in the sequence is smaller than \sqrt{x} . This completes the proof of Theorem 2.

We conclude this section with some general remarks. When elementary proofs of the PNT first were discovered, they generated much interest and excitement for at least two reasons: they upset long-held notions about “equivalence” and “depth” of propositions, and they provided hope for new insights about the distribution of primes.

Before there were such proofs, two propositions about primes were called *equivalent* if each could be deduced from the other by reasonably direct real variable arguments. For example, it was said that “convergence of $\sum_{n \leq x} \mu(n)/n$ is equivalent to the PNT.” On the other hand, the PNT would have been regarded as “deeper” than the Selberg formula

$$\vartheta(x) \log x + \sum_{p \leq x} \vartheta(x/p) \log p = (2 + o(1))x \log x$$

because the formula follows easily from $\vartheta(x) \sim x$, while no real variable proof of the opposite implication was then known. Admittedly, depth and equivalence were not precise notions, though they did appear plausible. However, after the discovery of elementary proofs of the PNT, such statements are reserved for informal use only.

Today, there are at least three distinct elementary paths to the PNT: that of Selberg-Erdős, the partial sieving method of H. Daboussi, and the large sieve method of A. J. Hildebrand, none of which is particularly easy. PNT error terms deduced from Selberg-type formulas, besides being difficult to establish, are not as good as those found by analytic methods. As Carl Pomerance suggested in his Postscript to the Graham-Spencer article [12], perhaps the greatest legacy of the elementary proof of the PNT is the motivation it provided for developing methods in combinatorial number theory.

4. DISTRIBUTION PROBLEMS FOR EULER'S FUNCTION

Among the results in Erdős' article [9] was the asymptotic formula

$$\#\{m : \varphi(m) \leq y\} \sim cy, \quad y \rightarrow \infty,$$

for some positive constant c , found by him and Paul Turán. (The version of this formula given in Mathematical Reviews contains an obvious error.) Their proof involved analysis of the distribution function

$$D_\varphi(\alpha) := \lim_{x \rightarrow \infty} x^{-1} \#\{n \leq x : \varphi(n)/n \leq \alpha\},$$

which had been studied first by I. J. Schoenberg and was subsequently shown to be purely singular by Erdős. The Erdős-Turán argument did not yield a value for the constant c .

What it did produce was a cottage industry studying problems of this type. First, Robert Dressler showed in [7] that $c = \zeta(2)\zeta(3)/\zeta(6)$ by evaluating the residue of the generating function. Next, Paul Bateman [1] gave three proofs of the Erdős-Turán result, one of which had an error term similar to that of the PNT. In [2], I then treated a version of this problem for rectangles,

$$\#\{n \leq x : \varphi(n) \leq y\} = xg(y/x) + E(y), \quad 1 \leq y < x,$$

with $E(y)$ again a PNT-type error term and $g(\alpha)$ a function on $[0, 1]$. The distribution function g is connected with Schoenberg's function by the differential equation

$$D_\varphi(\alpha) = g(\alpha) - \alpha g'(\alpha).$$

Differentiability questions for g caught Erdős' fancy, as he had long been interested in related questions for D_φ . This gave rise to more letters and several papers. One tidbit arising from our discussions is an identity that appeared as Problem 6363 in the American Mathematical Monthly 88 (1981): Prove that

$$D_\varphi(1/2) - D_\varphi(1/4) + D_\varphi(1/8) - D_\varphi(1/16) \pm \cdots = 1/2.$$

Another of our joint papers, [4], gave an abstraction of the preceding φ problems to ones for arithmetic functions having values uniformly distributed in $(0, \infty)$.

In [11], Erdős studied the modulus of continuity of the distribution function of $\sigma(n)/n$, where $\sigma(n)$ denotes the sum of positive divisors of n . Using his masterful technique of excluding sparse, inconvenient sets of integers and then making an elementary count of survivors having suitable divisibility properties, he showed, for every $a \geq 1$, that

$$\frac{1}{x} \# \left\{ n \leq x : a \leq \frac{\sigma(n)}{n} < a + \frac{1}{t} \right\} < \frac{c}{\log t},$$

an estimate that is best-possible apart from the value of c . The same argument establishes the corresponding result for $\varphi(n)/n$ as well. Correspondence with Erdős inspired Dennis Rhoads and me [6] to treat this problem via an analytic method, which gives essentially the same conclusion. The elementary method has the advantage of applying also for finite x , whereas that of [6] holds only as $x \rightarrow \infty$.

Erdős continued to have interest in the differential behavior of D_φ . One of his favorite distribution questions, at least in letters to me, was whether the derivative of D_φ is necessarily 0 at each point of differentiability. Sadly, this was a problem that outlasted him.

REFERENCES

- [1] P. T. Bateman, *The distribution of values of the Euler function*, Acta Arith. 21 (1972), 329–345.
- [2] H. G. Diamond, *The distribution of values of Euler's phi function* (in Analytic Number Theory, Proc. Sympos. Pure Math., Vol. XXIV, St. Louis Univ., St. Louis, Mo., 1972), pp. 63–75 Amer. Math. Soc., Providence, R.I., 1973.
- [3] ———, *Elementary methods in the study of the distribution of prime numbers*, Bull. Am. Math. Soc. (N.S.) 7 (1982), 553–589.

- [4] H. G. Diamond and P. Erdős, *Arithmetic functions whose values are uniformly distributed in $(0, \infty)$* (in Proceedings of the Queen's Number Theory Conference, 1979), 329–378, Queen's Papers in Pure and Appl. Math., 54, Queen's Univ., Kingston, Ont., 1980.
- [5] ———, *On sharp elementary prime number estimates*, L'Enseignement Math. 26 (1980), 313–321.
- [6] H. G. Diamond and D. Rhoads, *The modulus of continuity of the distribution function of $\varphi(n)/n$* (in Topics in Classical Number Theory, Vol. I, II, Budapest, 1981), 335–353, Colloq. Math. Soc. János Bolyai, 34, North-Holland, Amsterdam, 1984.
- [7] R. Dressler, *A density which counts multiplicity*, Pacific J. Math. 34 (1970), 371–378.
- [8] P. Erdős, *Beweis eines Satzes von Tschebyschef*, Acta Litt. Sci. Szeged 5 (1932), 194–198.
- [9] ———, *Some remarks on Euler's φ -function and some related problems*, Bull. Amer. Math. Soc. 51 (1945), 540–544.
- [10] ———, *On a new method in elementary number theory which leads to an elementary proof of the prime number theorem*, Proc. Natl. Acad. Sciences USA 35 (7) (1949), 374–384.
- [11] ———, *On the distribution of numbers of the form $\sigma(n)/n$ and on some related questions*, Pacific J. Math. 52 (1974), 59–65.
- [12] R. Graham and J. Spencer, *The elementary proof of the prime number theorem. With a note on the controversy by E. G. Straus and a postscript by Carl Pomerance*, Math. Intelligencer 31 (2009), 18–23.
- [13] A. E. Ingham, Review of *On a new method in elementary number theory which leads to an elementary proof of the prime number theorem*, Mathematical Reviews, MR0029411 (10,595c).
- [14] A. Selberg, *An elementary proof of the prime number theorem*, Ann. of Math. (2) 50 (1949), 305–313.

Harold G. Diamond

*Department of Mathematics,
University of Illinois,
1409 W. Green St.,
Urbana IL 61801,
USA*

e-mail: diamond@math.uiuc.edu

URL: <http://www.math.uiuc.edu/~diamond>

THE HISTORY OF DEGENERATE (BIPARTITE) EXTREMAL GRAPH PROBLEMS

ZOLTÁN FÜREDI and MIKLÓS SIMONOVITS*

This paper is a survey on Extremal Graph Theory, primarily focusing on the case when one of the excluded graphs is bipartite. On one hand we give an introduction to this field and also describe many important results, methods, problems, and constructions.

Contents:

1. Introduction	169
2. The General Theory, Classification	178
3. Excluding Complete Bipartite Graphs	190
4. Excluding Cycles: C_{2k}	204
5. Paths and Long Cycles	220
6. Excluding Trees	224
7. More Complex Excluded Subgraphs	227
8. Eigenvalues and Extremal Problems	235
9. Excluding Topological Subdivisions	236
10. Hypergraph Extremal Problems	237
11. Supersaturated Graphs	239
12. Ordered Structures	242
13. Applications in Geometry	245
14. Further Connections and Problems	247

1. INTRODUCTION

This survey describes the theory of *Degenerate Extremal Graph Problems*, the main results of the field, and its connection to the surrounding areas.

*Research supported in part by the Hungarian National Science Foundation OTKA 104343, and by the European Research Council Advanced Investigators Grant 267195 (ZF) and by the Hungarian National Science Foundation OTKA 101536, and by the European Research Council Advanced Investigators Grant 321104. (MS).

Extremal graph problems we consider here are often called Turán type extremal problems, because Turán's theorem and his questions were the most important roots of this huge area [242], [243].

Generally, we have a *Universe* of graphs, \mathbb{U} , where this universe may be the family of ordinary graphs, or digraphs, or hypergraphs, or ordered graphs, or bipartite graphs, etc and a property \mathcal{P} , saying, e.g., that $G \in \mathbb{U}$ does not contain some subgraphs $L \in \mathcal{L}$, or that it is Hamiltonian, or it is at most 3-chromatic, and we have some parameters on \mathbb{U} , say $v(G)$ and $e(G)$, the number of vertices and edges. Our aim is to maximize the second parameter under the condition that G has property \mathcal{P} and its first parameter is given.

We call such a problem *Turán type extremal problem* if we are given a family \mathcal{L} of graphs from our universe, G_n is a graph of n vertices, $e(G_n)$ denotes the number of edges of G_n and we try to maximize $e(G_n)$ under the condition that G_n contains no $L \in \mathcal{L}$, where “contains” means “not necessarily induced subgraph”. (Here graph may equally mean digraph, or multigraph, or hypergraph).

The maximum will be denoted by $\mathbf{ex}(n, \mathcal{L})$ and the graphs attaining this maximum without containing subgraphs from \mathcal{L} are called *extremal graphs*. The family of extremal graphs is denoted by $\mathbf{EX}(n, \mathcal{L})$ and $\mathbf{ex}(n, \mathcal{L})$ is called the *Turán number* of the family \mathcal{L} .

Speaking of $\mathbf{ex}(n, \mathcal{L})$ we shall always assume that $n \geq |V(L)|$, otherwise the problem is trivial.

Definition 1.1. If the Universe \mathbb{U} is the family of r -uniform hypergraphs¹, then we shall call the problem *degenerate* if the maximum,

$$\mathbf{ex}(n, \mathcal{L}) = o(n^r).$$

Otherwise we shall call it *non-degenerate*

Below we shall mention several open problems. Yet to get more problems, we refer the reader to the

Erdős homepage: www.renyi.hu/~p_erdos

where the papers of Erdős can be found [59]. Also, many open problems can be found in Chung-Graham [47].

¹ $r = 2$ included, moreover, mostly we think of $r = 2$.

1.1. Some central theorems of the field

We start with some typical theorems of the field and two conjectures. The aim of this “fast introduction” is to give a feeling for what are the crucial types of results here.

Theorem 1.2 (Kővári–T. Sós–Turán, [164]). *Let $K_{a,b}$ denote the complete bipartite graph with a and b vertices in its color-classes. Then*

$$\mathbf{ex}(n, K_{a,b}) \leq \frac{1}{2} \sqrt[a]{b-1} \cdot n^{2-(1/a)} + O(n).$$

We use this theorem with $a \leq b$, since that way we get a better estimate.

Theorem 1.3 (Kollár–Alon–Rónyai–Szabó [159], [11]). *If $b > (a-1)!$, then*

$$\mathbf{ex}(n, K_{a,b}) > c_a n^{2-(1/a)}.$$

Theorem 1.4 (Erdős, Bondy and Simonovits [32]).

$$\mathbf{ex}(n, C_{2k}) \leq 100kn^{1+(1/k)}.$$

Theorem 1.5 (Erdős–Simonovits, Cube Theorem [90]). *Let Q_8 denote the cube graph defined by the vertices and edges of a 3-dimensional cube. Then*

$$\mathbf{ex}(n, Q_8) = O(n^{8/5}).$$

Conjecture 1.6 (Erdős and Simonovits, Rational exponents). *For any finite family \mathcal{L} of graphs, if there is a bipartite $L \in \mathcal{L}$, then there exists a rational $\alpha \in [0, 1)$ and a $c > 0$ such that*

$$\frac{\mathbf{ex}(n, \mathcal{L})}{n^{1+\alpha}} \rightarrow c.$$

Theorem 1.7 (Füredi [111], [104]). *If $q \neq 1, 7, 9, 11, 13$, and $n = q^2 + q + 1$, then*

$$\mathbf{ex}(n, C_4) \leq \frac{1}{2}q(q+1)^2.$$

Moreover, if q is a power of a prime, then

$$\mathbf{ex}(n, C_4) = \frac{1}{2}q(q+1)^2.$$

Conjecture 1.8 (Erdős).²

$$\mathbf{ex}(n, \{C_3, C_4\}) = \frac{1}{2\sqrt{2}}n^{3/2} + o(n^{3/2}).$$

We close this part with a famous result of Ruzsa and Szemerédi:

²This conjecture is mentioned in [70] but it is definitely older, see e.g. Brown, [37].

Theorem 1.9 (Solution of the (6,3) problem, [210]). *If $\mathcal{H}_n^{(3)}$ is a 3-uniform hypergraph not containing 6 vertices determining (at least) 3 hyperedges, then this hypergraph has $o(n^2)$ hyperedges.*

The above theorems will be discussed in more details below.

1.2. The structure of this paper

The area is fairly involved. Figure 1 shows a complicated – but not com-

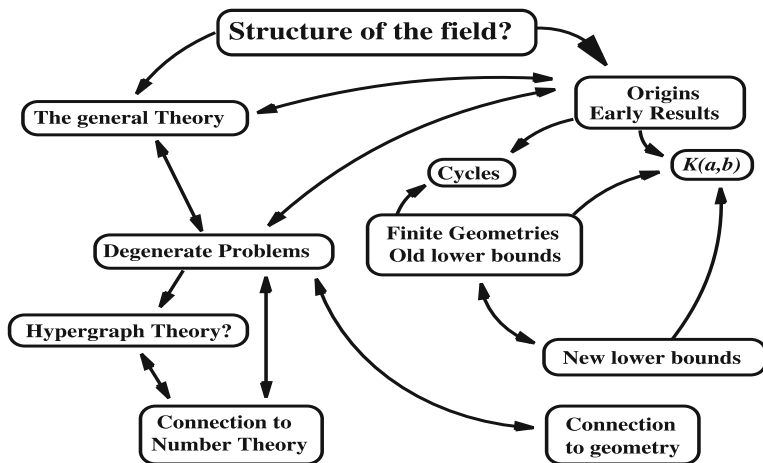


Fig. 1. Area Map

plete – map of the interactions of some subfields of the field discussed here. We start with describing the Extremal problems in general, then move to the Degenerate problems, also describing why they are important. Among the most important degenerate extremal problems we mention are the extremal problem of $K_{a,b}$, and also C_{2k} , where – to classify the extremal problems – we shall need the Random Graph Method to get a lower bound in the problem of C_{2k} . These results are enough to give a good *classification* of degenerate extremal graph problems.

1. The two lowest boxes of Figure 1 show that this whole area has a strong connection to Geometry and Number Theory. This will be explained in Sections 13, 1.5, and 14.2.
2. The origins of this field are
 - (a) an early, singular result of Mantel,
 - (b) the multiplicative Sidon Problem (see Section 1.5)
 - (c) Turán’s theorem and his systematic approach to the field.

So the origins come – in some sense – from Number Theory, are strongly connected to Finite Geometry, and in this way also to ordinary geometry (Turán’s theorem comes from the Erdős–Szekeres version of Ramsey Theorem, which they invented to solve the Esther Klein problem from Geometry.)

3. To understand the field we start with a very short description of the general theory, and then – skipping most of the hypergraph theory – we move to the main area of this paper: to the questions where we consider ordinary extremal graphs, and exclude some bipartite L : therefore, by Theorem 1.2, we have $\mathbf{ex}(n, \mathcal{L}) = O(n^{2-c})$.
4. One important phenomenon is that many extremal graph problems can be “reduced” to some *degenerate* extremal graph problems that we also call sometimes *bipartite extremal problems*.
5. The upper bounds in the simpler cases are obtained by some double counting, Jensen type inequalities, or applying some supersaturated graph theorems.³
6. There are also much more complicated cases, where the above simple approach is not enough, we need some finer arguments. Perhaps the first such case was treated by Füredi: Section 7.3 and [107]. Also such an approach is the application of the general Dependent Random Choice Method, (see the survey of Fox and Sudakov [101]).
7. The lower bounds are sometimes provided by random graphs (see Section 2.5) but these are often too weak. So we often use some finite geometric constructions, (see Section 3.2) or their generalizations – coming from commutative algebras (see Sections 3.6, 4.9, 8), etc., and they occasionally provide matching upper and lower bounds. Again, there is an important general method with many important results, which we shall call the Lazebnik–Ustimenko method but will treat only very superficially in Section 3.6.

1.3. Extremal problems

We shall almost entirely restrict ourselves to Turán type extremal problems for ordinary simple graphs, i.e. loops, multiple edges are excluded.

To show the relation of the areas described here, we start with a list of some related areas.

³Lovász and Szegedy had a beautiful conjecture, which we formulate here only in a restricted form: Any (valid) extremal theorem can be proven by applying the Cauchy–Schwarz inequality finitely many times. This conjecture was killed in this strong form – by Hatami and Norine [142] – but proven in a weaker, “approximation-form”.

1. Ramsey Theory

- Problems not connected to density problems; in some sense these are the real Ramsey Problems
- Problems connected to density problems, i.e. cases, where we do not really use the Ramsey Condition, only that some color class is large.

2. Ordinary extremal graph theory

- Excluding bipartite graphs (degenerate problems)
- Excluding topological subgraphs (very degenerate extremal problems)
- Matrix problems, ordered and not ordered;
- Non-degenerate case, and its relation to degenerate problems

3. Theory of extremal digraph problems

4. Ramsey–Turán Problems

5. Connection to Random Graphs

6. Hypergraph extremal problems

7. Connection of Number Theoretical problems to Extremal Graph Theory

8. Continuous problems

9. Applications

There are several surveys on these fields, see e.g., T. Sós [232], Füredi [108], [110], Simonovits [224], [228], [227], [222], Simonovits and Sós [230], [155]. Perhaps the first survey on this topic was Vera Sós' paper [232], discussing connections between extremal graph problems, finite geometries, block designs, etc. and, perhaps, the nearest to this survey is [225], Bollobás [28], Sidorenko [217], [101], and also some books, e.g., Bollobás [26]. Of course, a lot of information is hidden in the papers of Erdős, among others, in [70], [73], [76].

So, here we shall concentrate on Case 2, but to position this area we shall start with some related fields, among others, with the general asymptotic in Case 2.

Problem 1.10 (General Host-graphs). In a more general setting we have a sequence (H_n) of “host” graphs and the question is, how many edges can a subgraph $G_n \subset H_n$ have under the condition that it does not contain any forbidden subgraph $L \in \mathcal{L}$. The maximum will be denoted by $\text{ex}(H_n, \mathcal{L})$.

For $H_n = K_n$ we get back the ordinary extremal graph problems. There are several further important subcases of this question:

- (a) when $H_n = K_{a,b}$ for $a \approx n/2$;
- (b) when the host-graph is the d -dimensional cube, $n = 2^d$; see Section 14.3.
- (c) when H_n is a random graph on n vertices, see e.g. [209].

Notation. Given some graphs $G_n, T_{n,p}, T_k, H_\nu, \dots$ the (first) subscript will almost always denote the number of vertices.⁴ So K_p is the complete graph on p vertices, P_k the path on k vertices, C_k is the cycle of k vertices, while $\mathcal{C}_{\geq k}$ will denote the family of cycles of length at least k . $\delta(x)$ denotes the degree of the vertex x .

The complete bipartite graph $K_{a,b}$ with a vertices in its first class and b in its 2nd class will be crucial in this paper. Often, we shall denote it by $K(a, b)$, and its p -partite generalization by $K_p(a_1, \dots, a_p)$. If $\sum a_i = n$ and $|n_i - n_j| \leq 1$, then $K_p(a_1, \dots, a_p)$ is the Turán graph $T_{n,p}$ on n vertices and p classes.

Given two graphs U and W , their product graph is the one obtained from vertex-disjoint copies of these two graphs by joining each vertex of U to each vertex of W . This will be denoted by $U \otimes W$.⁵

Given a graph H , $v(H)$ is its number of vertices, $e(H)$ its number of edges and $\chi(H)$ its chromatic number, $d_{\min}(G)$ and $d_{\max}(G)$ denote the minimum and maximum degrees of G , respectively.

We shall write $f(x) \approx g(x)$ if $f(x)/g(x) \rightarrow 1$. Occasionally $[n]$ denote the set of first n integers, $[n] := \{1, 2, \dots, n\}$.

The Overlap. Some twenty years ago Simonovits wrote a survey [227] on the influence of Paul Erdős in the areas described above, Many-many features of these areas changed drastically since that. Jarik Nešetřil and Ron Graham were the editors of that survey-volume, and now they decided to republish it. Fortunately, the authors had the option to slightly rewrite their original papers. Simonovits has rewritten his original paper [228], basically keeping everything he could but *indicating* many new developments, and adding remarks and many new references to it.

To make *this* paper readable and self-contained, we shall touch on some basic areas also treated there, or in other survey papers of ours, Here, however, we shall explain many-many results and phenomena just mentioned in other survey papers.

⁴One important exception is the complete bipartite graph $K(a, b) = K_{a,b}$, see below. Another exception is, when we list some excluded subgraphs, like L_1, \dots, L_ν .

⁵This product is often called also the joint of the two graphs.

Remark 1.11. There is also a third new survey to be mentioned here: Simonovits gave a lecture at the conference on Turán’s 100th anniversary, in 2011. His lecture tried to cover the whole influence of Paul Turán in Discrete Mathematics. In the volume of this conference Simonovits wrote a survey [229] covering his lecture, except that

- the area called Statistical Group Theory is discussed in a survey of Pálffy and Szalay [200] and
- some parts of the applications of Extremal Graph Theory, primarily in Probability Theory are covered by Katona [153], in the same volume.

1.4. Other types of extremal graph problems

Above we still tried to maximize the number of edges, hyperedges, etc. More generally, instead of maximizing $e(G_n)$, we may maximize something else:

1. *Min-degree problems* (or Dirac type problems): How large min-degree can G_n have without containing subgraphs from \mathcal{L} .
2. *Median problems* which will be called here Loebel–Komlós–Sós type problems: Given a graph G_n , which m and d ensure that if G_n has at least m vertices of degree $\geq d$, then G_n contains some $L \in \mathcal{L}$.
3. *Eigenvalue-extremal problems*⁶: maximize the maximum eigenvalue $\lambda(G_n)$ under the condition that G_n does not contain any $L \in \mathcal{L}$. (These are sharper forms of some extremal results, since the maximum eigenvalue

$$\lambda(G_n) \geq \frac{2e(G_n)}{n},$$

see Section 8.)

4. *Subgraph count inequalities*, which assert that if G_n contains many copies of some subgraphs L_1, \dots, L_λ , then we have at least one (or maybe “many”) subgraphs L .
5. *Diameter-extremal problems*. Here we mention just a subcase: if

$$\text{diam}(G_n) \leq d \quad \text{and} \quad d_{\max}(G_n) < M,$$

at least how many edges must G_n have. The Erdős–Rényi paper [85] is of importance here and also some related papers, like Füredi [105], [109].

6. *Combined extremal problems*: There are many–many further types of extremal problems. Here we mention, as an example, the results

⁶As usual, given a graph G_n , an $n \times n$ matrix is associated to it, having n eigenvalues. The largest and the second largest is what we are mostly interested in.

of Balister, Bollobás, Riordan and Schelp [17], where an odd cycle is excluded, and at the same time an upper bound is fixed on the degrees and the number of edges are to be maximized.

The approach 4 is very popular in the theory of Graph Limits [177]. We mention a breakthrough in this area in connection with Erdős' combinatorial problems, of type 4. A famous conjecture of Erdős was

Conjecture 1.12 (Erdős [74]). *A K_3 -free G_n contains at most $\left(\frac{n}{5}\right)^5$ copies of C_5 's.*

The motivation of this is that the blowup⁷ C_5 , i.e. $C_5[n/5]$ has no triangles and has $\left(\frac{n}{5}\right)^5$ copies of C_5 . Erdős conjectured that no triangle-free G_n can have more C_5 's than this. The first "approximation" was due to Ervin Győri:

Theorem 1.13 (Győri [133]). *A K_3 -free G_n contains at most $1.03\left(\frac{n}{5}\right)^5$ C_5 's.*

Next Füredi improved the constant to 1.001 (unpublished) [114], and finally independently Grzesik [124], and Hatami, Hladký, Král, Norine, and Razborov [141] proved the conjecture.

1.5. Historical remarks

Erdős in 1938 [60] considered the following "multiplicative Sidon Problem"⁸.

Problem 1.14. How many integers, $a_1, \dots, a_m \in [1, n]$ can we find so that $a_i a_j = a_k a_\ell$ does not hold for any i, j, k, ℓ , unless $\{i, j\} = \{k, \ell\}$.

To get an upper bound in Problem 1.14, Erdős proved

Theorem 1.15. *Let $G[n, n]$ be a bipartite graph with n vertices in both classes. If it does not contain C_4 , then $e(G[n, n]) < 3n\sqrt{n}$,*

Much later this problem was asked in a more general setting: find an upper bound on $e(G[n, n])$ if $K_{a,b} \not\subset G[n, n]$. Zarankiewicz [254] posed the following question:

⁷Given a graph H , its blowup version $H[t]$ is defined as follows: we replace each vertex x of H by t independent new vertices and we join two new vertices coming from distinct vertices x, y iff xy was an edge of H .

⁸For a longer description of the number theoretical parts see [228]. Erdős also refers to his "blindness" overlooking the general problem in [70].

Problem 1.16 (Zarankiewicz problem). Determine the largest integer $Z(m, n, a, b)$ for which there is an $m \times n$ 0-1 matrix containing $Z(m, n, a, b)$ 1's without an $a \times b$ submatrix consisting entirely of 1's.

Hartman, Mycielski and Ryll-Nardzewski [139] gave upper and lower bounds for the case $a = b = 2$, weaker than the Erdős–Klein⁹ result, and Kővári, T. Sós and Turán (see Theorem 1.2) provided a more general upper bound. We shall discuss these problems and results in details in Sections 2.4 and 3.2.

While exact values of $Z(m, n, a, b)$ are known for infinitely many parameter values, mostly only asymptotic bounds are known in the general case. Even $Z(m, n, 2, 2)$ is not sufficiently well known.

2. THE GENERAL THEORY, CLASSIFICATION

In many ordinary extremal problems the minimum chromatic number plays a decisive role. The **subchromatic number** $p(\mathcal{L})$ of \mathcal{L} is defined by

$$(2.1) \quad p(\mathcal{L}) = \min\{\chi(L) : L \in \mathcal{L}\} - 1.$$

Recall that the Turán graph $T_{n,p}$ is the largest graph on n vertices and p classes.

Claim 2.1.

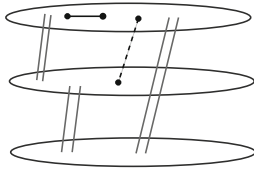
$$(2.2) \quad \mathbf{ex}(n, \mathcal{L}) \geq e(T_{n,p}) = \left(1 - \frac{1}{p}\right) \binom{n}{2} + o(n^2).$$

Indeed, $T_{n,p}$ does not contain any $L \in \mathcal{L}$. An easy consequence of the Erdős–Stone theorem [95] provides the asymptotic value of $\mathbf{ex}(n, \mathcal{L})$, at least if $p(\mathcal{L}) > 1$.

Theorem 2.2 (Erdős–Simonovits [89]). *If \mathcal{L} is a family of graphs with subchromatic number $p > 0$, then*

$$\mathbf{ex}(n, \mathcal{L}) = \left(1 - \frac{1}{p}\right) \binom{n}{2} + o(n^2).$$

⁹In [70] Erdős (again) attributes the finite geometric construction to Eszter (Esther) Klein.



This means that $\mathbf{ex}(n, \mathcal{L})$ depends only very loosely on \mathcal{L} ; up to an error term of order $o(n^2)$; it is already determined by $p(\mathcal{L})$.¹⁰ The question is whether the structure of the extremal graphs is also almost determined by $p(\mathcal{L})$, and (therefore) it must be very similar to that of $T_{n,p}$.¹¹ The answer

is **YES**. This is expressed by the following results of Erdős and Simonovits [66], [67], [218]:

Theorem 2.3 (The Asymptotic Structure Theorem). *Let \mathcal{L} be a family of forbidden graphs with subchromatic number p . If $S_n \in \mathbf{EX}(n, \mathcal{L})$, (i.e., S_n is extremal for \mathcal{L}), then it can be obtained from $T_{n,p}$ by deleting and adding $o(n^2)$ edges. Furthermore, if \mathcal{L} is finite, then the minimum degree*

$$d_{\min}(S_n) = \left(1 - \frac{1}{p}\right)n + o(n).$$

Further, the almost-extremal graphs are similar to $T_{n,p}$.

Theorem 2.4 (The First Stability Theorem). *Let \mathcal{L} be a family of forbidden graphs with subchromatic number p . For every $\varepsilon > 0$, there exist a $\delta > 0$ and an n_ε such that, if G_n contains no $L \in \mathcal{L}$, and if, for $n > n_\varepsilon$,*

$$(2.3) \quad e(G_n) > \mathbf{ex}(n, \mathcal{L}) - \delta n^2,$$

then G_n can be obtained from $T_{n,p}$ by changing¹² at most εn^2 edges.

Remark 2.5. For ordinary graphs ($r = 2$) we often call the degenerate extremal graph problems *bipartite extremal problems*. This is the case when \mathcal{L} contains some bipartite graphs. There is a slight problem here: we shall also consider the case when not only some $L \in \mathcal{L}$ is bipartite but $\chi(G_n) = 2$ is as well.

2.1. The importance of the Degenerate Case

There are several results showing that if we know sufficiently well the extremal graphs for the degenerate cases, then we can also reduce the non-degenerate cases to these problems.

¹¹Actually, this was the original question; Theorem 2.2 was a partial answer to it.

¹²deleting and adding

Exact Turán numbers, product conjecture. We start with an illustration. Let $O_6 = K(2, 2, 2)$ be the octahedron graph. Erdős and Simonovits proved that

Theorem 2.6 (Octahedron Theorem [91]). *If S_n is an extremal graph for the octahedron O_6 for n sufficiently large, then there exist extremal graphs G_1 and G_2 for the circuit C_4 and the path P_3 such that $S_n = G_1 \otimes G_2$ and $|V(G_i)| = \frac{1}{2}n + o(n)$, $i = 1, 2$.*

If G_1 does not contain C_4 and G_2 does not contain P_3 , then $G_1 \otimes G_2$ does not contain O_6 . Thus, if we replace G_1 by any H_1 in $\mathbf{EX}(v(G_1), C_4)$ and G_2 by any H_2 in $\mathbf{EX}(v(G_2), P_3)$, then $H_1 \otimes H_2$ is also extremal for O_6 .

More generally,

Theorem 2.7 (Erdős–Simonovits [91]). *Let L be a complete $(p+1)$ -partite graph, $L := K(a, b, r_3, r_4, \dots, r_{p+1})$, where $r_{p+1} \geq r_p \geq \dots \geq r_3 \geq b \geq a$ and $a = 2, 3$. There exists an $n_0 = n_0(a, b, \dots, r_{p+1})$ such that if $n > n_0$ and $S_n \in \mathbf{EX}(n, L)$, then $S_n = U_1 \otimes U_2 \otimes \dots \otimes U_p$, where*

1. $v(U_i) = n/p + o(n)$, for $i = 1, \dots, p$.
2. U_1 is extremal for $K_{a,b}$
3. $U_2, U_3, \dots, U_p \in \mathbf{EX}(n, K(1, r_3))$.

It follows that this theorem is indeed a reduction theorem.

Conjecture 2.8 (The Product Conjecture, Simonovits). *Assume that $p(\mathcal{L}) = \min_{L \in \mathcal{L}} \chi(L) - 1 > 1$. If for some constants $c > 0$ and $\varepsilon \in (0, 1)$*

$$(2.4) \quad \mathbf{ex}(n, \mathcal{L}) > e(T_{n,p}) + cn^{1+\varepsilon},$$

then there exist p forbidden families \mathcal{M}_i , with

$$p(\mathcal{M}_i) = 1 \quad \text{and} \quad \max_{M \in \mathcal{M}_i} v(M) \leq \max_{L \in \mathcal{L}} v(L),$$

such that for any $S_n \in \mathbf{EX}(n, \mathcal{L})$, $S_n = G_1 \otimes \dots \otimes G_p$, where G_i are extremal for \mathcal{M}_i .

This means that the extremal graphs S_n are products of extremal graphs for some degenerate extremal problems (for \mathcal{M}_i), and therefore we may reduce the general case to degenerate extremal problems.

Remarks 2.9. (a) If we allow infinite families \mathcal{L} , then one can easily find counterexamples to this conjecture.

(b) If we allow linear error-terms, i.e. do not assume (2.4), then one can also find counterexamples, using a general theorem of Simonovits [221]; however, this is not trivial at all, see [223].

(c) A weakening of the above conjecture would be the following: for arbitrary large n , in Conjecture 2.8 there are several extremal graphs, and for each $n > n_{\mathcal{L}}$, some of them are of product form, (but maybe not all of them) and the families \mathcal{M}_i also may depend on n a little.

Further sources to read: Griggs, Simonovits, and Thomas [127], Simonovits, [226].

2.2. The asymmetric case of Excluded Bipartite graphs

The degenerate extremal graph problems have three different forms:

Problem 2.10 (Three versions). (a) Ordinary extremal graph problems, where some bipartite or non-bipartite sample graphs are excluded, and we try to maximize $e(G_n)$ under this conditions.

(b) The *bipartite case*, where the host graph is $K(m, n)$ and we maximize $e(G_{n+m})$ under the conditions that $G_{n+m} \subseteq K(m, n)$ and G_{n+m} contains no $L \in \mathcal{L}$. (Here we may assume that all $L \in \mathcal{L}$ are bipartite.) In this case we often use the notation $\mathbf{ex}(m, n, \mathcal{L})$. If $m \leq n$ but $m > cn$ for some constant $c > 0$, then the answer to this problems and to the problem of $\mathbf{ex}(n, \mathcal{L})$ are the same, up to a constant. If, however, we assume that n is much larger than m , then some surprising new phenomena occur, see Section 14.2.

(c) The *asymmetric case*. Color the vertices of the sample graphs L in RED-BLUE and exclude only those $G_n \subseteq K(m, n)$ where the RED vertices of some $L \in \mathcal{L}$ are in the FIRST class of $K(m, n)$: maximize $e(G_{n+m})$ over the remaining graphs $G_{n+m} \subseteq K(m, n)$.

Denote the maximum number of edges in this third case by $\mathbf{ex}^*(m, n, \mathcal{L})$.

Remark 2.11. We have seen Zarankiewicz' problem (i.e. Problem 1.16). That corresponds to an asymmetric graph problem, (c). If we exclude in an $m \times n$ matrix both a $a \times b$ and an $b \times a$ submatrices of 1's, that will correspond to a bipartite graph problem, (b).

Conjecture 2.12 (Erdős, Simonovits [225]).

$$\mathbf{ex}^*(n, n, \mathcal{L}) = O(\mathbf{ex}(n, \mathcal{L})).$$

The simplest case when we cannot prove this is $L = K(4, 5)$.

Remark 2.13 (Matrix problems). Case (c) has also a popular matrix form where we consider 0-1 matrices and consider an $m \times n$ matrix not containing a submatrix A . The question is: how many 1-s can be in such a matrix. This problem has (at least) two forms: the unordered and ordered one. We return to the Ordered Case in Subsection 12.4.

2.3. Reductions: Host graphs

The following simple but important observation shows that there is not much difference between considering any graph as a “host” graph or only bipartite graphs.

Lemma 2.14 (Erdős’ bipartite subgraph lemma). *Every graph G_n contains a bipartite subgraph H_n with $e(H_n) \geq \frac{1}{2}e(G_n)$.*

This lemma shows that there is not much difference between considering K_{2n} or $K_{n,n}$ as a host graph.

Corollary 2.15. *If $\mathbf{ex}_B(n, \mathcal{L})$ denotes the maximum number of edges in an \mathcal{L} -free bipartite graph, then*

$$\mathbf{ex}_B(n, \mathcal{L}) \leq \mathbf{ex}(n, \mathcal{L}) \leq 2 \mathbf{ex}_B(n, \mathcal{L}).$$

Assume now that we wish to have an upper bound on $\mathbf{ex}(m, n, \mathcal{L})$, where $n \gg m$. One way to get such an upper bound is to partition the n vertices into subsets of size $\approx m$. If, e.g., we know that $\mathbf{ex}(m, m, \mathcal{L}) \leq cm^{1+\gamma}$, then we obtain that

$$(2.5) \quad \mathbf{ex}(m, n, \mathcal{L}) \leq \frac{n}{m} \cdot \mathbf{ex}(m, m, \mathcal{L}) \leq cnm^\gamma.$$

This often helps, however, occasionally it is too weak. Erdős formulated

Conjecture 2.16. *If $n > m^2$ then $\mathbf{ex}(m, n, C_6) = O(n)$.*

Later this conjecture was made more precise, by Erdős, A. Sárközy and T. Sós, and proved by Györi, see Section 14.2 and [134].

We start with a trivial lemma.

Lemma 2.17. *Let d be the average degree in G_n , i.e. $d := 2e(G_n)/n$. Then G_n contains a G_m with $d_{\min}(G_m) \geq d/2$.*

To solve the cube-problem, Erdős and Simonovits used two reductions. The first one was a reduction to bipartite graphs, see Section 2.14. The other one eliminates the degrees are much higher than the average.

Definition 2.18 (Δ -almost-regularity). G is Δ -almost-regular if $d_{\max}(G) < \Delta \cdot d_{\min}(G)$.

Theorem 2.19 (Δ -almost-regularization [90]). Let $e(G_n) > n^{1+\alpha}$, and $\Delta = 20 \cdot 2^{(1/\alpha)^2}$. Then there is a Δ -almost-regular $G_m \subset G_n$ for which

$$e(G_m) > \frac{2}{5}m^{1+\alpha}, \quad \text{where } m > n^{\alpha \frac{1-\alpha}{1+\alpha}},$$

unless n is too small.

This means that whenever we wish to prove that $\mathbf{ex}(n, \mathcal{L}) = O(n^{1+\alpha})$, we may restrict ourselves to bipartite Δ -almost-regular graphs.

It would be interesting to understand the limitations of this lemma better. The next remark and problem are in this direction.

Remark 2.20. By the method of random graphs one can show [90] that for every Δ and n and $\varepsilon > 0$, there is G_n with $e(G_n) = \lfloor n^{3/2} \rfloor$ which does not have a Δ -almost-regular subgraph G_m with $e(G_m) > \varepsilon \sqrt{nm}$.

Problem 2.21 (Erdős–Simnovits [90]). Is it true that for every Δ there exists an $\varepsilon > 0$ such that every G_n , with $e(G_n) = \lfloor n \log n \rfloor$, contains a Δ -almost-regular subgraph G_m , with $e(G_m) > \varepsilon m \log m$ where $m \rightarrow \infty$ when $n \rightarrow \infty$?

2.4. Excluding complete bipartite graphs

Certain questions from topology (actually, Kuratowski theorem on planar graphs) led to Zarankiewicz problem [254]. After some weaker results Kővári, T. Sós and Turán proved the following theorem, already mentioned in Section 1.1.

Theorem 2.22 (Kővári–T. Sós–Turán, [164]). Let $K_{a,b}$ denote the complete bipartite graph with a and b vertices in its color-classes. Then

$$(2.6) \quad \mathbf{ex}(n, K_{a,b}) \leq \frac{1}{2} \sqrt{b-1} \cdot n^{2-(1/a)} + \frac{a-1}{2}n.$$

Remarks 2.23. (a) If $a \neq b$ then (2.6) is better if we apply it with $a < b$.

(b) We know from Theorem 2.2 that $\mathbf{ex}(n, \mathcal{L}) = o(n^2)$ if and only if \mathcal{L} contains a bipartite L . Actually Claim 2.1 and Theorem 2.22 show that if $\mathbf{ex}(n, \mathcal{L}) = o(n^2)$ then $\mathbf{ex}(n, \mathcal{L}) = O(n^{2-c})$, for some constant $c > 0$.

Conjecture 2.24 ([164], see also e.g. [70]). *The upper bound in Theorem 1.2 is sharp:*

$$\mathbf{ex}(n, K_{a,b}) > c_{a,b} n^{2-(1/a)}.$$

Sketch of proof of Theorem 2.22. The number of a -stars $K_{a,1}$ in a graph G_n is $\sum \binom{d_i}{a}$ where d_1, \dots, d_n are the degrees in G_n . If G_n contains no $K_{a,b}$ then at most $b-1$ of these a -stars can share the same set of endpoints. We obtain

$$(2.7) \quad \sum \binom{d_i}{a} = \text{the number of } a\text{-stars} \leq (b-1) \binom{n}{a}.$$

Extending $\binom{n}{a}$ to all $x > 0$ by

$$\binom{x}{a} := \begin{cases} \frac{x(x-1)\dots(x-a+1)}{a!} & \text{for } x \geq a-1, \\ 0 & \text{otherwise} \end{cases}$$

we have a convex function. Then Jensen's Inequality implies that, the left hand side in (2.7) is at least $n \binom{2e(G)/n}{a}$, and the result follows by an easy calculation. ■

Remark 2.25. Slightly changing the above proof we get analogous upper bounds on $e(G_n)$ in all three cases of Problem 2.10.

We shall return to these questions in Sections 3.1, 3.4 where we shall discuss some improvements of the upper bound and also some lower bounds.

Further sources to read: Guy [128], Znám: [257], [256], Guy–Znám [129].

2.5. Probabilistic lower bound

The theory of random graphs is an interesting, important, and rapidly developing subject. The reader wishing to learn more about it should either read the original papers of Erdős, e.g., [61], [62], Erdős and Rényi, e.g., [84], or some books, e.g., Bollobás, [27], Janson, Łuczak and Ruciński, [149], Molloy and Reed [192].

Theorem 2.26 (Erdős–Rényi First Moment method). *Let $\mathcal{L} = \{L_1, \dots, L_t\}$ be a family of graphs, and let*

$$(2.8) \quad c = \max_j \min_{H \subseteq L_j} \frac{v(H)}{e(H)}, \quad \gamma = \max_j \min_{H \subseteq L_j} \frac{v(H) - 2}{e(H) - 1},$$

where the minimum is taken only for subgraphs H where the denominator is positive.

(a) Let G_n be a graph of order n chosen uniformly, at random, from graphs with E_n edges. For every $\varepsilon > 0$ there exists a $\delta > 0$ such that if $E_n < \delta n^{2-c}$, then the probability that G_n contains at least one $L \in \mathcal{L}$ is at most ε .

(b) If we know only $E_n < \varepsilon n^{2-\gamma}$, then the probability that G_n contains at least $\frac{1}{2}E_n$ copies of $L \in \mathcal{L}$ is at most ε .

This implies that

$$(2.9) \quad \text{ex}(n, \mathcal{L}) > c_{\mathcal{L}} n^{2-\gamma} \geq c_{\mathcal{L}} n^{2-c}$$

with $c \geq \gamma > 0$ defined above.

Remarks 2.27. (a) A graph L is called balanced if the minimum in (2.8), for c , is achieved for $H = L$. Erdős and Rényi formulated their result containing Theorem 2.26(a) only for balanced graphs L . The part we use is trivial from their proof.

(b) Later Bollobás extended the Erdős–Rényi theorem to arbitrary \mathcal{L} .

(c) Györi, Rothschild and Ruciński achieved the generality by embedding any graph into a balanced graph [136].

Corollary 2.28. *If a finite \mathcal{L} contains no trees,¹³ then for some $c_{\mathcal{L}} > 0$, $\text{ex}(n, \mathcal{L}) \geq c_{\mathcal{L}} n^{1+c}$.*

Mostly the weaker Theorem 2.26(a) implies Corollary 2.28: it does, whenever \mathcal{L} is finite and each $L \in \mathcal{L}$ contains at least two cycles in the same component. However, for cycles we need the stronger Theorem 2.26(b).

For example, for $L = K_{a,b}$ we have $c = a^{-1} + b^{-1}$. Then, for c_0 sufficiently small, the probability that a graph G_n with $c_0 n^{2-c}$ edges does not contain $K_{a,b}$ is positive. Hence

$$\text{ex}(n, K_{a,b}) \geq c_0 n^{2-(1/a)-(1/b)}.$$

Comparing this with the Kővári–T. Sós–Turán theorem (Theorem 2.22), we see that the exponent is sharp there, in some sense, if a is fixed while $b \rightarrow \infty$.

¹³neither forests

Proof of Theorem 2.26. Consider the random graph G_n with n labeled vertices, in which each edge occurs independently, with the same probability p . For each L_j , choose a subgraph H_j which attains the inner minimum for γ , in (2.8). Let $h_j := v(H_j)$, $e_j := e(H_j)$, and let α_j denote the number of copies of H_j in K_{h_j} , and β_j denote the expected number of copies of H_j in G_n .

Clearly, K_n contains $\alpha_j \binom{n}{h_j}$ copies of H_j . For each copy H of H_j , define a random “indicator” variable $k_H = k_H(G_n) = 1$ if $H \subseteq G_n$, and 0 otherwise. Since the number of copies of H_j in G_n is just $\sum_{H \subseteq K_n} k_H$, therefore, if \mathbb{E} denote the expected value, then

$$\beta_j = \sum_{H \subseteq K_n} \mathbb{E}(k_H) = \alpha_j \binom{n}{h_j} p^{e_j}.$$

Summing over j and taking $p = c_1 n^{-c}$, (for some $c_1 \in (0, 1)$) we get

$$\sum_j \beta_j \leq t \max_j \alpha_j \binom{n}{h_j} p^{e_j} \leq t \max c_1 n^{h_j - c e_j} = t c_1 n^{2-c}.$$

Now let $\eta(G_n) = e(G_n) - \sum_j \beta_j$. Then, for c_1 sufficiently small, the expected value is

$$\mathbb{E}(\eta(G_n)) > \frac{1}{2} \binom{n}{2} p > \frac{1}{5} c_1 n^{2-c}.$$

Hence there exists a G_n with $\eta(G_n) > \frac{1}{5} c_1 n^{2-c}$. Delete an edge from each H_j in this G_n . The resulting graph contains no L_j , and has at least $\frac{1}{5} \binom{n}{2} p \geq \frac{1}{11} c_1 n^{2-c}$ edges, completing the proof. ■

Remarks 2.29 (How did the probabilistic methods start?). Mostly we write that applications of the Random Graphs (probabilistic method) started when Erdős (disproving a conjecture of Turán on the Ramsey Numbers) proved the existence of graphs G_n without complete subgraphs of order $2 \log n$ and without independent sets of size $2 \log n$.

1. Erdős himself remarks (e.g., in [64]) that perhaps Szele was the first who applied this method in Graph Theory. (Erdős – in his birthday volume [76] – also mentions an even earlier application of J. Erőd but we did not succeed in locating that source.)
2. Perhaps the earliest case of applying probabilistic methods was that of Paul Turán’s proof of the Hardy-Ramanujan Theorem [241], where – reading the paper – it is obvious that Turán gave a probabilistic proof

of a beautiful and important theorem, using the Chebishev inequality. However, either Turán did not realize that this is an application of the probabilistic method or he did not wish to burden the reader with that.

3. An important application of the probabilistic methods was that of Claude Shannon, when he constructed random codes.

Applying Theorem 2.26 to some families of cycles we obtain

Corollary 2.30. *For some constant $c_m > 0$,*

$$\mathbf{ex}(n, \{C_3, \dots, C_m\}) \geq c_m n^{1+\frac{1}{m-1}}.$$

Erdős' even cycles theorem asserts that $\mathbf{ex}(n, C_{2t}) = O(n^{1+(1/t)})$, and this upper bound is probably sharp.¹⁴ The random method (that is, Theorem 2.26) yields a lower bound of $cn^{1+\frac{1}{2t-1}}$, a weaker result. Simonovits thinks that it is unlikely that Theorem 2.26 ever yields a sharp bound for a finite family.¹⁵

Corollary 2.30 is used in the next section to prove that $\mathbf{ex}(n, \mathcal{L}) = O(n)$ if and only if contains a tree or forest.

2.6. Classification of extremal problems

The extremal graph problems can be classified in several ways. Here we shall speak of (a) *non-degenerate*, (b) *degenerate* and (c) *linear* extremal problems.

For Case (a) Theorem 2.3 provides an appropriately good description of the situation. In Case (b) $p(\mathcal{L}) = 1$. Here the “main term” disappears, $(1 - \frac{1}{p}) = 0$; therefore “the error terms dominate”. Case (c) will be discussed here shortly and in Sections 6 and 9 in more details.

The classification immediately follows from the following theorems:

Theorem 2.31. *$\mathbf{ex}(n, \mathcal{L}) = o(n^2)$ if and only if \mathcal{L} contains a bipartite graph. Actually, if \mathcal{L} contains a bipartite graph then $\mathbf{ex}(n, \mathcal{L}) = O(n^{2-c})$ for, e.g., $c = 2/v(L)$ for any bipartite $L \in \mathcal{L}$. If \mathcal{L} does not contain bipartite graphs, then $\mathbf{ex}(n, \mathcal{L}) \geq \left\lfloor \frac{n^2}{4} \right\rfloor$.*

¹⁴The reference is missing here, since Erdős did formulate this theorem but never have published a proof of it, as far as we know.

¹⁵Some related results of G. Margulis, and A. Lubotzky, R. Phillips and P. Sarnak will be discussed in Section 4.9.

Theorem 2.32. For finite \mathcal{L} , $\mathbf{ex}(n, \mathcal{L}) = O(n)$ if and only if \mathcal{L} contains a tree, or a forest. If $L \in \mathcal{L}$ is a tree or a forest, then, for $v(L) \geq 3$,

$$(2.10) \quad \mathbf{ex}(n, \mathcal{L}) < (v(L) - 2)n.$$

Theorem 2.33 (Erdős [61, 62]). If \mathcal{L} is finite and no $L \in \mathcal{L}$ is a tree, then $\mathbf{ex}(n, \mathcal{L}) > n^{1+c_{\mathcal{L}}}$ for some $c_{\mathcal{L}} > 0$.

Theorem 2.34 (Erdős [62], Bondy and Simonovits [32]). Given an integer k , for some constants $c_k, \tilde{c}_k > 0$,

$$(2.11) \quad c_k n^{1+\frac{1}{2k-1}} < \mathbf{ex}(n, \{C_3, \dots, C_{2k}\}) \leq \mathbf{ex}(n, C_{2k}) \leq \tilde{c}_k n^{1+\frac{1}{k}}.$$

Proof of Theorems 2.31, 2.32, and 2.33. If there is a bipartite $L \in \mathcal{L}$, then Theorem 2.22 implies the sharper upper bound of Theorem 2.31. Indeed, for $v = v(L)$, by $L \subseteq K([v/2], v)$, we have,

$$\mathbf{ex}(n, \mathcal{L}) \leq \mathbf{ex}(n, L) \leq \mathbf{ex}(n, K([v/2], v)) < \frac{1}{2} \sqrt{2v} \cdot n^{2-(2/v(L))} = O(n^{2-c}).$$

If the minimum chromatic number $p = p(\mathcal{L}) \geq 3$, then $T_{n,p}$ contains no forbidden $L \in \mathcal{L}$. Therefore

$$\mathbf{ex}(n, \mathcal{L}) > e(T_{n,2}) \geq e(T_{n,p}) = \left(1 - \frac{1}{p}\right) \binom{n}{2} + O(n).$$

Actually, $e(T_{n,2}) = \left\lfloor \frac{n^2}{4} \right\rfloor$. This completes the proof of Theorem 2.31. ■

It is easy to show that if G_n has minimum degree at least $r - 1$, then it contains every tree T_r (by induction on r). An induction on n yields (2.10), implying half of Theorem 2.32, when \mathcal{L} contains a tree (or a forest). If \mathcal{L} is finite and contains no trees, i.e., all the forbidden graphs contain some cycles, then we use Theorem 2.34, or simply Corollary 2.28, proved by probabilistic methods.¹⁶

Remark 2.35 (Infinite families). For infinite families the situation is different: if e.g. \mathcal{C} is the family of all cycles, then $\mathbf{ex}(n, \mathcal{C}) = n - 1$: all graphs but the forests are excluded. There are many further families without trees where the extremal number is linear, see Section 9.

¹⁶There are also deterministic proofs of Corollary 2.28, e.g., via the Margulis–Lubotzky–Phillips–Sarnak construction of Ramanujan graphs, see Construction 4.43.

Proof of Theorem 2.34. The lower bound comes from a random graph argument of Erdős. Concentrate on the upper bound. If we are not interested in the value of the constant, then we can basically use the following argument: Take a graph G_n with $cn^{1+\alpha}$ edges. Delete its minimum degree vertex, then the minimum degree vertex in the remaining graph, etc. At the end we get a G_m with minimum degree at least c_1m^α . In the obtained graph G_m fix a vertex x and denote by S_j the set of vertices at distance j from x . If $\text{girth}(G_n) > 2k$, – as we assumed – then *basically* $|S_j| > d_{\min}(G_m) \cdot |S_{j-1}|$. Hence $m > |S_k| > c_1^k m^{\alpha k}$. So $\alpha \leq 1/k$. ■

Assume for a second that G_n itself is asymptotically regular:

$$\frac{d_{\min}(G_n)}{d_{\max}(G_n)} \rightarrow 1.$$

Then the previous argument asserts that $d := d_{\min}(G_n) < n^{1/k}$. Therefore

$$e(G_n) \leq \left(\frac{1}{2} + o(1)\right) nd \approx \frac{1}{2} n^{1+\frac{1}{k}}.$$

We shall return to the case of excluded trees, namely, to the Erdős–Sós conjecture on the extremal number of trees, and to the related Komlós–Sós conjecture in Section 6. One final question could be if $\text{ex}(n, \mathcal{L})$ can be sublinear. This is answered by the following trivial result.

Theorem 2.36. *If \mathcal{L} is finite and $\text{ex}(n, \mathcal{L}) < \lfloor n/2 \rfloor$, then $\text{ex}(n, \mathcal{L}) = O(1)$.*

Proof. Consider $n/2$ independent edges: this must contain an $L_1 \in \mathcal{L}$. Hence, there is an $L_1 \in \mathcal{L}$ contained in the union of t independent edges, for some t . Also, there exists an $L_2 \subseteq K(1, n-1)$. Hence an extremal graph S_n has bounded degrees and bounded number of independent edges. This proves 2.36. ■

Theorem 2.36 easily extends to hypergraphs.

2.7. General conjectures on bipartite graphs

We have already formulated Conjecture 1.6 on the rational exponents. We have to remark that for hypergraphs this does not hold: the Behrend construction [21] is used to get lower bounds in the Ruzsa–Szemerédi Theorem, (Thm 1.9), showing that there is no rational exponent in that case. Yet, Erdős and Simonovits conjectured that for ordinary graphs there is. One could also conjecture the inverse extremal problem:

Conjecture 2.37. For every rational $\alpha \in (0, 1)$ there is a finite \mathcal{L} for which $c_1 n^{1+\alpha} < \mathbf{ex}(n, \mathcal{L}) < c_2 n^{1+\alpha}$, for some constants $c_1, c_2 > 0$.

The third conjecture to be mentioned here is on “compactness” [93]:

Conjecture 2.38. For every finite \mathcal{L} there is an $L \in \mathcal{L}$ for which $\mathbf{ex}(n, \mathcal{L}) > c \cdot \mathbf{ex}(n, L)$, for some constants $c_{\mathcal{L}} > 0$.

3. EXCLUDING COMPLETE BIPARTITE GRAPHS

3.1. Bipartite C_4 -free graphs and the Zarankiewicz problem

Turán type extremal results (and Ramsey results as well) can often be applied in Mathematics, even outside of Combinatorics. Turán himself explained this applicability by the fact that – in his opinion – the extremal graph results were generalizations of the Pigeon Hole Principle.

Recall that $Z(m, n, a, b)$ denotes the maximum number of 1’s in an $m \times n$ matrix not containing an $a \times b$ minor consisting exclusively of 1’s. In 1951 Zarankiewicz [254] posed the problem of determining $Z(n, n, 3, 3)$ for $n \leq 6$, and the general problem has also become known as *the problem of Zarankiewicz*.¹⁷ Obviously, $Z(m, n, 1, b) = m(b - 1)$ (for $n \geq b - 1$). Observe that $Z(m, n, a, b) = \mathbf{ex}^*(m, n, K_{a,b})$ (where $\mathbf{ex}^*(m, n, \mathcal{L})$ was defined following Remark 2.10.) Considering the adjacency matrix of a $K_{a,b}$ -free graph on n vertices we get $2\mathbf{ex}(n, K_{a,b}) \leq Z(n, n, a, b)$. We will use this upper bound many times.

We will see that the easy upper bound in Theorem 2.22 is pretty close to the truth for $a \leq 2$. Actually, Kővári, T. Sós and Turán [164] proved an upper bound for the Zarankiewicz function

$$(3.1) \quad Z(m, n, a, b) \leq \sqrt[a]{b-1} \cdot mn^{1-(1/a)} + (a-1)n$$

which was slightly improved by Znám [257], [256], (he halved the last term to $(a-1)n/2$ in the case of $m = n$) and Guy [128].

A bipartite graph $G[M, N]$ where $|M| = m$, $|N| = n$ is C_4 -free if its “bipartite” $m \times n$ adjacency matrix contains no 2×2 full 1 submatrix.¹⁸

¹⁷In Graph Theory two problems are connected to Zarankiewicz’ name: the extremal problem for matrices that we shall discuss here and the Crossing Number conjecture which is not our topic. Actually, the crossing number problem comes from Paul Turán, see [244].

¹⁸Here the “bipartite adjacency matrix” $A = (a_{ij})_{m \times n}$ is defined for a bipartite graph $G[U, V]$ and $a_{ij} = 1$ if $u_j \in U$ is joined to $v_j \in V$, otherwise $a_{ij} = 0$.

In other terminology, the hypergraph defined by the rows of this matrix is linear, and their hyperedges pairwise meet in at most one element. There is an important class of such hypergraphs, the Steiner k -systems $S(n, k, 2)$. A family \mathcal{S} of k -subsets of an n -set N is a Steiner k -system if every pair of elements is covered exactly once. For such an \mathcal{S} , clearly, $|\mathcal{S}| = m = \binom{n}{2} / \binom{k}{2}$. Such families are known to exist for $(m, n, k) = (q^2 + q + 1, q^2 + q + 1, q + 1)$ (called finite projective planes of order q), and $(m, n, k) = (q^2 + q, q^2, q)$ (affine planes) whenever q is a power of a prime. Also for any given k there exists an $n_0(k)$ such that $S(n, k, 2)$ exists for all *admissible* $n > n_0(k)$, i.e., when $(n - 1)/(k - 1)$ and $n(n - 1)/k(k - 1)$ are integers (Wilson's existence theorem [248]).

Kővári, T. Sós and Turán [164] proved that

Theorem 3.1. $Z(n, n, 2, 2) = (1 + o(1))n^{3/2}$, and

$$(3.2) \quad Z(n, n, 2, 2) < [n^{3/2}] + 2n.$$

Further, if p is a prime, then

$$Z(p^2 + p, p^2, 2, 2) = p^3 + p^2.$$

Reiman [206] returned to this topic, (see also [207]), slightly improving (3.2)

Theorem 3.2 (Reiman [206]).

$$(3.3) \quad Z(m, n, 2, 2) \leq \frac{1}{2} \left(m + \sqrt{m^2 + 4mn(n - 1)} \right).$$

For large $m, n \rightarrow \infty$, and $m = o(n^2)$, this yields

$$Z(m, n, 2, 2) \leq \left(\frac{1}{2} + o(1) \right) n\sqrt{m}.$$

Further, for $m = n$, we get

$$(3.4) \quad Z(n, n, 2, 2) \leq \frac{1}{2}n (1 + \sqrt{4n - 3}) \approx n\sqrt{n}.$$

Reiman also provides infinitely many graphs, using Finite Geometries, showing the sharpness of (3.3) and (3.4). We have equality when $m = n(n - 1)/k(k - 1)$ and a Steiner system $S(n, k, 2)$ exists. Thus he determined the case

$$(3.5) \quad Z(n, n, 2, 2) = \frac{1}{2}n (1 + \sqrt{4n - 3}) = (q^2 + q + 1)(q + 1)$$

for $m = n = q^2 + q + 1$ when a projective plane of order q exists. Actually, in [207], Reiman also speaks about Zarankiewicz-extremal graphs connected to incidence-graphs of higher dimensional finite geometries.

Since Reiman's theorem the theory of finite geometries developed tremendously. We cite here a recent result whose proof used the most modern tools and stability results.

Theorem 1 (Damásdi, Héger, and Szőnyi [57]). *Let $q \geq 15$, and $c \leq q/2$. Then*

$$Z(q^2 + q + 1 - c, q^2 + q + 1, 2, 2) \leq (q^2 + q + 1 - c)(q + 1).$$

Equality holds if and only if a projective plane of order q exists. Moreover, graphs giving equality are subgraphs of the bipartite incidence graph of a projective plane of order q obtained by omitting c rows of its incidence matrix.

They proved many more exact results when a projective plane of order q exists. The extremal configurations are submatrices of the incidence matrix of a projective plane.

$$Z(q^2 + c, q^2 + q, 2, 2) = q^2(q + 1) + cq \quad (0 \leq c \leq q + 1),$$

$$Z(q^2 - q + c, q^2 + q - 1, 2, 2) = (q^2 - q)(q + 1) + cq \quad (0 \leq c \leq 2q),$$

$$Z(q^2 - 2q + 1 + c, q^2 + q - 2, 2, 2) = (q^2 - 2q + 1)(q + 1) + cq$$

$$(0 \leq c \leq 3(q - 1)).$$

These refer to bipartite host graphs. As we will see later, such exact results are rare for the general (non-bipartite) case. To estimate $\mathbf{ex}(n, C_4)$ seems to be harder, because the corresponding 0-1 matrices, the incidence matrix of a graph, should be symmetric.

3.2. Finite Geometries and the C_4 -free graphs

The method of finite geometric constructions is very important and powerful in combinatorics. In particular, it is often the best way to obtain lower bounds. It is for this reason that we include this section.

We give several constructions: the first two show that the Kővári–T. Sós–Turán theorem (Theorem 2.22) is sharp for both $K_{2,2}$ and $K_{3,3}$.

Remark 3.3. When we write that an upper bound is sharp, mostly we mean that it is sharp up to a multiplicative constant: it yields the correct exponent. There are a few exceptions, where sharpness means that the ratio of the upper and lower bounds tends to 1. This is the case for $C_4 = K_{2,2}$ and we have this also for $K_{3,3}$. Here, however, the matching upper bound for Construction 3.20 below is given not by Theorem 2.22 but by the Füredi improvement [112].

Perhaps the application of finite geometries in Extremal Graph Theory started in the Erdős paper, with the construction of Eszter Klein [60], to prove the sharpness of Theorem 1.15. The expression “Finite Geometry” was not mentioned there. We skip the description of this whole story, since it was described in several places, e.g., [227], [228].

Much later, Erdős and Rényi [85] used finite geometry for a diameter-extremal problem. This is a very large area, connected to our problems, yet we have to skip it. The interested reader is referred to [85], (translated into English in [208]).

Sharp extremal graph results were obtained by Reiman [206] and a Polarity Graph was used in [86] and [36] to give asymptotically sharp lower bound on $\mathbf{ex}(n, C_4)$. This lower bound can also be found in [85], **implicitly**: Erdős and Rényi considered the diameter-extremal problem, and do mention the properties of this graph.

The real breakthrough came by the Erdős–Rényi–T. Sós paper [86], (sharp lower bound for C_4) and by the Brown paper [36], providing asymptotically sharp lower bounds for $\mathbf{ex}(n, C_4)$ and for $\mathbf{ex}(n, K_{3,3})$. (See Remark 3.3.)

We know from Theorem 2.22 that $\mathbf{ex}(n, C_4) \leq \frac{1}{2}n^{3/2} + o(n^{3/2})$, but is this result sharp? In analyzing the proof, we realize that if it is sharp (that is, if there are infinitely many graphs G_n not containing C_4 and having $\approx \frac{1}{2}n\sqrt{n}$ edges), then almost all degrees are $\approx \sqrt{n}$ and almost every pair of vertices must have a common neighbor (and no pair has two). This suggests that the neighborhoods $N(x)$ behave much like the lines in a projective plane, in that the following statement “almost” holds: any two vertices lie in a common set, and any two sets intersect in one vertex.

Theorem 3.4 (Erdős–Rényi–T. Sós [86], and Brown [36], see also [164]).

$$\mathbf{ex}(n, C_4) = \frac{1}{2}n^{3/2} + O(n^{3/2-c}).$$

For the lower bound for $\mathbf{ex}(n, C_4)$ we use the following

Construction 3.5. Let p be a prime, $n = p^2 - 1$. Construct a graph as follows: the vertices are the $p^2 - 1$ non-zero pairs (x, y) of residues (modulo p), and (x, y) is joined to (a, b) by an edge if $ax + by = 1$. (This graph may contain loops, but we simply delete them.)

With $n = p^2 - 1$, the resulting graph H_n has the necessary properties to show the sharpness of Theorem 2.22 for C_4 :

(a) for a given pair (a, b) , mostly there are p solutions to $ax + by = 1$, so that, even after the loops are deleted, there are at least $\frac{1}{2}(p^2 - 1)(p - 1)$ edges in H_n and hence $e(H_n) > \frac{1}{2}n^{3/2} - n$;

(b) if H_n had a 4-cycle with vertices (a, b) , (u, v) , (a', b') and (u', v') , then the two equations $ax + by = 1$ and $a'x + b'y = 1$ would have two solutions, which is impossible. Since the primes are “dense” among the integers, this completes the proof of the sharpness of Theorem 2.22 for $a = b = 2$.

Remark 3.6. An alternative possibility is to use the much more symmetric polarity graph of the projective plane (we explain this in the next section): here we used the Affine Geometric Variant because *here* we did not wish to use anything from Projective Geometry.

3.3. Excluding C_4 : Exact results

The polarity graph¹⁹, used in [85], was also used in [86] and [36] to prove that

$$(3.6) \quad \text{ex}(n, C_4) \geq \frac{1}{2}q(q+1)^2, \quad \text{for } n = q^2 + q + 1.$$

if q is a prime power.

Construction 3.7 (The Polarity Graph from the finite field). Assume that q is a prime power. Consider the Finite Field $GF(q)$. The vertices of our graph are the equivalence classes of the non-zero triples $(a, b, c) \in GF(q)^3$ where two of them, (a, b, c) and (a', b', c') are considered the same if $(a', b', c') = \lambda(a, b, c)$ for some $\lambda \neq 0$. There are $(q^3 - 1)/(q - 1) = q^2 + q + 1$ such classes. Further, the equivalence class of (a, b, c) is connected by an edge to the class of (x, y, z) if $ax + by + cz = 0$. Finally, we delete the $q + 1$ loops, i.e. those edges, where $a^2 + b^2 + c^2 = 0$. This graph is C_4 -free and it has $\frac{1}{2}(n(q + 1) - (q + 1))$ edges.

¹⁹These C_4 -free graphs were studied earlier in finite geometry. The bipartite point-line incidence graph appeared in Levi’s book (1942) and polarity graphs (modulo loops) obtained from *Levi graph* had been described already by Artzy (1956). For more details and references see Bondy [30].

In general, a polarity corresponds to a *symmetric* incidence matrix of a finite plane of size $(q^2 + q + 1) \times (q^2 + q + 1)$. According to a theorem of Baer [13] such a matrix has at least $q + 1$ non-zero elements in its diagonal. Therefore using the polarity graph we cannot avoid losing on loops. This way Erdős, Rényi and Sós [86] showed that indeed $\text{ex}(q^2 + q + 1, C_4) < \frac{1}{2}n(q + 1)$. Yet one could hope to get a better construction. Erdős conjectured [66], [71] that there are no better constructions, that is, (3.6) is sharp if $n = q^2 + q + 1$, (q is a prime power).

Füredi settled this conjecture in the following sense: First he proved [103] that if $q = 2^k$, then Erdős' conjecture holds. Next he settled the case $q \geq q_0$. Later he found a much shorter proof of the weaker assertion that the Polarity graphs are extremal; however, this shorter version did not give the extremal structure. So Füredi published the shorter version, while the longer version can be found on his homepage.

Theorem 3.8 (Füredi [111], [104]). *If $q \neq 1, 7, 9, 11, 13$ and $n = q^2 + q + 1$, then $\text{ex}(n, C_4) \leq \frac{1}{2}q(q + 1)^2$ and for $q > 13$ the extremal graphs are obtained from a polarity of a finite projective plane. Hence if $q > 13$ is a prime power, then $\text{ex}(n, C_4) = \frac{1}{2}q(q + 1)^2$.*

The second part of this result probably holds for $q \in \{7, 9, 11, 13\}$, too.

Recently a new sharp construction has been found for $n = q^2 + q$.

Theorem 3.9 (Firke, Kosek, Nash and Williford [102]). *Suppose that q is even, $q > q_0$. Then*

$$\text{ex}(q^2 + q, C_4) \leq \frac{1}{2}q(q + 1)^2 - q.$$

Consequently, if $q > q_0$, $q = 2^k$ and $n = q^2 + q$ then $\text{ex}(n, C_4) = q(q + 1)^2 - q$.

They also announced that in a forthcoming paper they show that for all but finitely many even q , any $S_n \in \mathbf{EX}(q^2 + q, C_4)$ is derived from an orthogonal polarity graph by removing a vertex of minimum degree (the 1-vertex-truncated Polarity graph, see Construction 3.7). This result shows a kind of stability of the Polarity graph. More generally, McCuaig (private communication, 1985) **conjectured** that each extremal graph is a subgraph of some polarity graph. So this is true for infinitely many cases, but one of the present authors strongly disagrees and he believes just the opposite that for e.g., $n = q^2 + q + 2$ maybe the extremal graphs are obtained by *adding* an extra vertex and some edges to a polarity graph.

Remark 3.10. W. McCuaig calculated $\text{ex}(n, C_4)$ for $n \leq 21$ (unpublished letter, 1985). Clapham, Flockart and Sheehan determined the corresponding

extremal graphs [48], and Yuansheng and Rowlinson [252], – using computers, – extended these results to $n \leq 31$. (They also determined the graphs in $\mathbf{EX}(n, C_6)$ for $n \leq 26$, [253].) Garnick, Kwong, Lazebnik, and Nieuwejaar [122], [123] determined the values of $\mathbf{ex}(n, \{C_3, C_4\})$ for all $n \leq 30$.

3.4. Excluding $K(2, t + 1)$, $t > 1$

A slightly sharper form of the upper bound (3.1) was presented by Hyltén-Cavallius [146]

$$(3.7) \quad Z(m, n, 2, k) \leq \frac{1}{2}n + \left\{ (k-1)nm(m-1) + \frac{1}{4}n^2 \right\}^{1/2}.$$

Obviously, for fixed k and large values of n, m , if $n = o(m^2)$, then the right hand side of (3.7) is $\approx \sqrt{k-1}m\sqrt{n}$. Using again the observation $2\mathbf{ex}(n, K_{2,t+1}) \leq Z(n, n, 2, t+1)$ one obtains the upper bound

$$(3.8) \quad \mathbf{ex}(n, K_{2,t+1}) \leq \frac{1}{2}n\sqrt{tn - t + 1/4} + (n/4).$$

The following theorem shows that the above (easy) upper bound is the best possible asymptotically.

Theorem 3.11 (Füredi [113]). *For any fixed $t \geq 1$*

$$(3.9) \quad \mathbf{ex}(n, K_{2,t+1}) = \frac{1}{2}\sqrt{tn}^{3/2} + O(n^{4/3}).$$

To prove this Theorem one needs an appropriate lower bound, a construction. Let q be a prime power such that $(q-1)/t$ is an integer. We will construct a $K_{2,t+1}$ -free graph G on $n = (q^2 - 1)/t$ vertices such that every vertex has degree q or $q-1$. We will explain this below (Construction 3.15). Then G has more than $(1/2)\sqrt{tn}^{3/2} - (n/2)$ edges. The gap between the lower and upper bounds is only $O(\sqrt{n})$ for $n = (q^2 - 1)/t$. The lower bound for the Turán number for all n then follows from the fact that such prime powers form a dense subsequence among the integers. This means that for every sufficiently large n there exists a prime q satisfying $q \equiv 1 \pmod{t}$ and $\sqrt{nt} - n^{1/3} < q < \sqrt{nt}$ (see [145]).

Construction 3.15 below is inspired by constructions of Hyltén-Cavallius and Mörs given for Zarankiewicz's problem $Z(n, n, 2, t+1)$.

Theorem 3.12 (Hyltén-Cavallius [146]). $Z(n, n, 2, 3) = \sqrt{2}n^{3/2} + o(n^{3/2})$. Also

$$\sqrt{\lfloor k/2 \rfloor} \leq \liminf_{n \rightarrow \infty} \frac{Z(n, n, 2, k)}{n^{3/2}}.$$

Theorem 3.13 (Mörs [194]). For all $t \geq 1$,

$$\frac{Z(n, n, 2, t + 1)}{n^{3/2}} \rightarrow \sqrt{t}, \quad \text{as } n \rightarrow \infty.$$

The topic was so short of constructions that, as a first step, P. Erdős [66, 69] even proposed the problem whether $\lim_t(\liminf_n \mathbf{ex}(n, K_{2,t+1})n^{-3/2})$ goes to ∞ as $t \rightarrow \infty$.

Remark 3.14. Here we see three distinct quantities, exactly as it is described in Problem 2.10. $Z(m, n, 2, t + 1) = \mathbf{ex}^*(m, n, K_{2,t+1})$, estimated from below by Mörs, by a construction, and $\mathbf{ex}(m, n, K_{2,t+1})$ estimated by Füredi by the same construction. Füredi showed that the matrix of Mörs contains neither a $(t + 1) \times 2$ submatrix, nor a $2 \times (t + 1)$ submatrix of 1’s; finally, Füredi, slightly changing the definitions in Mörs’s construction extended this “asymmetric matrix” result to the symmetric case and provided a non-bipartite graph, proving (3.9).

Construction 3.15. Let $GF(q)$ be the q -element finite field, and let $h \in GF(q)$ be an element of order t . This means, that $h^t = 1$ and the set $H = \{1, h, h^2, \dots, h^{t-1}\}$ form a t -element subgroup of $GF(q) \setminus \{0\}$. For $q \equiv 1 \pmod t$ such an element $h \in GF(q)$ always exists.

We say that $(a, b) \in GF(q) \times GF(q)$, $(a, b) \neq (0, 0)$ is equivalent to (a', b') , in notation $(a, b) \sim (a', b')$, if there exists some $h^\alpha \in H$ such that $a' = h^\alpha a$ and $b' = h^\alpha b$. The elements of the vertex set V of G are the t -element equivalence classes of $GF(q) \times GF(q) \setminus (0, 0)$. The class represented by (a, b) is denoted by $\langle a, b \rangle$. Two (distinct) classes $\langle a, b \rangle$ and $\langle x, y \rangle$ are joined by an edge in G if $ax + by \in H$. This relation is symmetric, and $ax + by \in H$, $(a, b) \sim (a', b')$, and $(x, y) \sim (x', y')$ imply $a'x' + b'y' \in H$. So this definition is compatible with the equivalence classes.

For any given $(a, b) \in GF(q) \times GF(q) \setminus (0, 0)$ (say, $b \neq 0$) and for any given x and h^α , the equation $ax + by = h^\alpha$ has a unique solution in y . This implies that there are exactly tq solutions (x, y) with $ax + by \in H$. The solutions come in equivalence classes, so there are exactly q classes $\langle x, y \rangle$. One of these classes might coincide with $\langle a, b \rangle$ so the degree of the vertex $\langle a, b \rangle$ in G is either q or $q - 1$.

We claim that G is $K_{2,t+1}$ -free. First we show, that for $(a, b), (a', b') \in GF(q) \times GF(q) \setminus (0, 0)$, $(a, b) \not\sim (a', b')$ the equation system

$$(3.10) \quad ax + by = h^\alpha \quad \text{and} \quad a'x + b'y = h^\beta$$

has at most one solution $(x, y) \in GF(q) \times GF(q) \setminus (0, 0)$. Indeed, the solution is unique if the determinant $\det \begin{pmatrix} a & b \\ a' & b' \end{pmatrix}$ is not 0. Otherwise, there exists a c such that $a = a'c$ and $b = b'c$. If there exists a solution of (2) at all, then multiplying the second equation by c and subtracting it from the first one we get on the right hand side $h^\alpha - ch^\beta = 0$. Thus $c \in H$, contradicting the fact that (a, b) and (a', b') are not equivalent.

Finally, there are t^2 possibilities for $0 \leq \alpha, \beta < t$ in (3.10). The set of solutions again form t -element equivalence classes, so there are at most t equivalence classes $\langle x, y \rangle$ joint simultaneously to $\langle a, b \rangle$ and $\langle a', b' \rangle$. ■

Since then, there have been two additional almost optimal constructions, strongly related to the Construction 3.15 above.

Construction 3.16 (Lazebnik, Mubayi [167]). Let $GF(q)^*$ be the finite field of order q without the zero element. Suppose $q \equiv 1 \pmod{t}$ and let H be the t -element multiplicative subgroup of $GF(q)^*$. Define the graph G^\times as follows. Let $V(G^\times) = (GF(q)^*/H) \times GF(q)$. For $\langle a \rangle, \langle b \rangle \in (GF(q)^*/H)$ and $x, y \in GF(q)$, make $(\langle a \rangle, x)$ adjacent to $(\langle b \rangle, y)$ if $x + y \in \langle ab \rangle$.

This graph (after deleting the eventual loops) is $K_{2,t+1}$ -free and every vertex has degree $q - 1$ or $q - 2$. Actually, Construction 3.16 differs from Construction 3.15 only in that its vertex set is smaller and instead of using the rule that $\langle a, b \rangle$ is adjacent to $\langle x, y \rangle$ if $ax + by \in H$ they use the rule $ay + bx \in H$. This change allows them to generalize it to multipartite hypergraphs.

The following example works only if t is a power of a prime, and $t|q$.

Construction 3.17 (Lenz, Mubayi [173]). Suppose that t divides q and let H be an additive subgroup of $GF(q)$ of order t . Define the graph G^+ as follows. Let $V(G^+) = (GF(q)/H) \times GF(q)^*$. We will write elements of $GF(q)/H$ as $\langle a \rangle$. It is the additive coset of H generated by a , $\langle a \rangle = \{h + a : h \in H\}$. For $\langle a \rangle, \langle b \rangle \in (GF(q)/H)$ and $x, y \in GF(q)^*$, make $(\langle a \rangle, x)$ adjacent to $(\langle b \rangle, y)$ if $xy \in \langle a + b \rangle$. (This, in fact, means that there exists an $h \in H$ such that $xy = a + b + h$).

3.5. Excluding $K(3, 3)$, and improving the upper bound

The main result of this section is the description of the asymptotically sharp value of $\mathbf{ex}(n, K_{3,3})$.

Theorem 3.18 (Brown [36] and Füredi [112]).

$$\mathbf{ex}(n, K_{3,3}) = \frac{1}{2}n^{5/3} + O(n^{(5/3)-c}) \quad \text{for some } c > 0.$$

The lower bound can be obtained from Brown’s example (discussed below as Construction 3.20) who gave a $(p^2 - p)$ -regular $K_{3,3}$ -free graph on p^3 vertices for each prime p of the form $4k - 1$.

Improving the upper bound in Theorem 2.22 Füredi showed that Brown’s example is asymptotically optimal.

Theorem 3.19 (Füredi [112]). *For all $m \geq a, n \geq b, b \geq a \geq 2$ we have*

$$(3.11) \quad Z(m, n, a, b) \leq (b - a + 1)^{1/a} mn^{1-(1/a)} + (a - 1)n^{2-(2/a)} + (a - 2)m.$$

For fixed $a, b \geq 2$ and $n, m \rightarrow \infty$ the first term is the largest one for $n = O(m^{a/(a-1)})$. This upper bound is asymptotically optimal for $a = 2$ and for $a = b = 3$ ($m = n$). We obtain

$$(3.12) \quad \mathbf{ex}(n, K_{3,3}) \leq \frac{1}{2}Z(n, n, 3, 3) \leq \frac{1}{2}n^{5/3} + n^{4/3} + \frac{1}{2}n.$$

Alon, Rónyai and Szabó [11] gave an example (discussed as Construction 3.25) showing that

$$(3.13) \quad \mathbf{ex}(n, K_{3,3}) \geq \frac{1}{2}n^{5/3} + \frac{1}{3}n^{4/3} - C.$$

for some absolute constant $C > 0$ for every n of the form $n = p^3 - p^2$, p is a prime. Their example shows that the upper bound (3.12) (and (3.11)) is so tight that that we cannot leave out the second order term. It would be interesting to see whether (3.11) is tight for other values of a and b , too.

The first step of the proof of Theorem 3.19 is that given a $K_{a,b}$ -free graph G , we apply the original bound (3.1) to the bipartite subgraphs $G[N(x), V \setminus N(x)]$ generated by the neighborhood of a vertex x and its complement.

When Brown gave his construction, the matching upper bound of Theorem 3.19 was not known yet. He wrote that even the existence of $\lim_{n \rightarrow \infty} \mathbf{ex}(n, K_{3,3})/n^{5/3}$ was unknown.

Construction 3.20. Let p be an odd prime $n = p^3$ and $d \in GF(p)$, $d \neq 0$ a quadratic residue if p is of the form $4k - 1$ and d be a non-residue otherwise. Construct a graph B_n whose vertices are the triples (x, y, z) of residue classes (modulo p) and whose edges join vertices (x, y, z) and (x', y', z') if

$$(3.14) \quad (x - x')^2 + (y - y')^2 + (z - z')^2 = d.$$

It is easy to see that the graph B_n has $\frac{1}{2}n^{5/3} + O(n^{4/3})$ edges. Given a vertex (x', y', z') , the equation (3.14) has $p^2 - p$ solutions by a theorem of Lebesgue. Thus (x', y', z') has this many neighbors.

We claim that B_n does not contain $K_{3,3}$. The geometric idea behind Construction 3.5 (concerning C_4 -free graphs) was to join a point of the finite plane to the points of its “polar” (with respect to the unit circle), and then to use the fact that two lines intersect in at most one point. In contrast, the Brown construction uses the fact that, if points of the Euclidean space \mathbb{E}^3 at distance 1 are joined, then the resulting infinite graph G does not contain $K_{3,3}$. This is easily seen as follows: suppose G does contain $K_{3,3}$. Then the three points of one color class cannot be collinear since no point is equidistant from three collinear points. On the other hand, only two points are equidistant from three points on a circle, and so $K_{3,3}$ cannot occur. There is one problem with this “proof”: in finite fields $\sum_i x_i^2 = 0$ can occur even if not all x_i 's are 0's. Therefore in finite geometries, in some cases, a sphere can contain a whole line. So here the geometric language must be translated into the language of analytic geometry, and the right hand side of (3.14) (that is d) must be chosen appropriately.

Theorem 3.21 (Nikiforov, [198]). *For $b \geq a \geq 2$ let $k \in [0, a - 2]$ be an integer. Then*

$$Z(m, n, a, b) \leq (b - k - 1)^{1/a} mn^{1-(1/a)} + (a - 1)n^{1+(k/a)} + km.$$

For $k = 0$ we get back Theorem 2.22, and substituting $k = a - 2$ we obtain (3.11). Nikiforov remarks that letting k run from 0 to $a - 2$, we may get the best results for various values of k as the relation of m and n varies, but we still have no constructions to substantiate this. Nikiforov also proves results on the spectral radius.

3.6. Further applications of Algebraic Methods

Most of the constructions providing sufficiently good lower bounds for Bipartite Extremal Graph Problems are coming either from Geometry or from Algebra²⁰. In all these cases the vertices of the graph-construction are “coordinatized” and two vertices are joined if some (usually polynomial) equations are satisfied.²¹

Actually, this motivated Conjecture 1.6 or its weakening: If we use a typical finite geometric construction, then there is a d -dimensional space, where each vertex is joined to a t -dimensional subspace. Hence $n = p^d$, the degrees are around $n^{t/d}$, so the construction has around $n^{1+(t/d)}$ edges. The conjecture suggests that there are always such almost extremal constructions.²²

The most important question in this part is if one can find constructions²³ to provide lower bounds where the exponents match the exponents in the upper bounds. Here we shall discuss when do we know the sharpness of the Kővári–T. Sós–Turán upper bound, $\mathbf{ex}(n, K_{a,b}) = O(n^{2-(1/a)})$.

As we have mentioned in Section 1.1, Kollár, Rónyai and T. Szabó [159] gave a construction which was improved by Alon, Rónyai and Szabó [11] (Constructions 3.23 and 3.25 below). The basic idea of their proofs was – at least in our interpretation – the same as that of William G. Brown; however, much more advanced. In the three dimensional Euclidean space \mathbb{E}^3 the Unit Distance Graph contains no $K_{3,3}$. If we change the underlying field to a finite field $GF(q)$ (as Brown did in Construction 3.20) then we obtain a finite graph having $n = q^3$ vertices. The neighborhood of each vertex will have $\approx q^2$ neighbors, and therefore $\approx \frac{1}{2}n^{5/3}$ edges. Now comes the crucial part: despite the fact, that this is highly nontrivial, we could say, that – because of the geometric reason, – this graph contains no $K_{3,3}$ proving the sharpness of (2.6).²⁴

If we wish to extend the above construction to get lower bounds for $\mathbf{ex}(n, K_{a,a})$ and we mechanically try to use unit balls in the a -dimensional

²⁰The Random Graph Methods are very nice but mostly they are too weak to provide sufficiently sharp lower bounds.

²¹Some of the constructions may seem number theoretic.

²²Here we have to make some remarks about our “Conjectures”: Many of them have the feature that it is not that interesting if they are true or false: in proving any alternative, we get new, important knowledge about our topics. The first such “Conjecture” was that of Turán on “Diagonal” Ramsey Numbers, that lead to the Erdős Random Graph Approach, see Remark 2.29.

²³Or “random constructions”.

²⁴Actually, as we have already discussed this in Subsection 3.20, the Will Brown’s lower bound also proves this sharpness, only, the lower bound of [11] is a little better.

space $GF(q)^a$ then several problems occur. We would need that any a of them intersect in at most $a - 1$ points. Then we would be home.

In \mathbb{E}^4 we can choose two orthogonal circles of radii $\frac{1}{\sqrt{2}}$, e.g.,

$$\{(x_1, x_2, 0, 0) : x_1^2 + x_2^2 = 1/2\} \quad \text{and} \quad \{(0, 0, x_3, x_4) : x_3^2 + x_4^2 = 1/2\},$$

then each point on the first one has distance 1 from each point in the second one. Hence the “Unit Distance Graph” contains $K(\infty, \infty)$. (Similarly, the “Unit Distance Graph” of $GF(q)^4$ contains a $K_{q,q}$.) So everything seems (!) to break down? Not quite, by the Kollár–Rónyai–Szabó construction. Instead of the ‘Euclidean metric’ they use a so-called *norm* in the space $GF(q^a)$. Two vectors \mathbf{x} and \mathbf{y} are connected if the norm of their sum is 1; $N(\mathbf{x} + \mathbf{y}) = 1$. (In this context there is not much difference between connecting them this way or take a bipartite graph and connecting the vertices in it if $N(\mathbf{x} - \mathbf{y}) = 1$).

Theorem 3.22 (Kollár, Rónyai, and T. Szabó [159] for $b > a!$, Alon, Rónyai, and Szabó [11] for $b > (a - 1)!$). *There exists a $c_a > 0$ such that for $b > (a - 1)!$ we have*

$$\text{ex}(n, K_{a,b}) > c_a n^{2-(1/a)}.$$

Below we provide the Kollár–Rónyai–Szabó construction and a short verification. The norm of an element $\mathbf{x} \in GF(q^a)$ is defined as

$$N(\mathbf{x}) := \mathbf{x} \cdot \mathbf{x}^q \cdots \mathbf{x}^{q^{a-1}}.$$

Construction 3.23 (Kollár–Rónyai–T. Szabó [159], the Norm Graph). The vertices of $G(q, a)$ are the elements $\mathbf{x} \in GF(q^a)$. The elements \mathbf{x} and \mathbf{y} are joined if $N(\mathbf{x} + \mathbf{y}) = 1$.

We claim that $G(q, a)$ is $K_{a,b}$ -free where $b = a! + 1$. If we have a $K_{b,a} \subseteq G(q, a)$, then fixing – as parameters – the a vertices $\mathbf{y}_1, \dots, \mathbf{y}_a$, we get a equations of the form $N(\mathbf{x} + \mathbf{y}_i) = 1$ with b solutions $\mathbf{x} \in \{\mathbf{x}_1, \dots, \mathbf{x}_b\}$. Then we can use the following result from Algebraic Geometry with $t = a$.

Lemma 3.24. *Let K be a field and $\alpha_{i,j}, \beta_i \in K$ for $1 \leq i, j \leq t$ such that $\alpha_{i_1,j} \neq \alpha_{i_2,j}$ if $i_1 \neq i_2$. Then the system of equation*

$$\begin{aligned} (x_1 - \alpha_{1,1})(x_2 - \alpha_{1,2}) \cdots (x_t - \alpha_{1,t}) &= \beta_1 \\ (x_2 - \alpha_{2,1})(x_2 - \alpha_{2,2}) \cdots (x_t - \alpha_{2,t}) &= \beta_2 \\ &\vdots \\ (x_t - \alpha_{t,1})(x_2 - \alpha_{t,2}) \cdots (x_t - \alpha_{t,t}) &= \beta_t \end{aligned}$$

has at most $t!$ solutions $(x_1, x_2, \dots, x_t) \in K^t$. ■

Construction 3.25 (Alon–Rónyai–T. Szabó [11]). The vertices of the graph $H(q, a)$ are the elements $(x, X) \in GF(q)^* \times GF(q^{a-1})$ and (x, X) and (y, Y) are joined if $N(X + Y) = xy$.

Here the norm $N(X)$ is defined in $GF(q^{a-1})$ and so it is $X \cdot X^q \cdots X^{q^{a-2}}$. The graph $H(q, a)$ has $(q - 1)q^{a-1}$ vertices, it is $q^{a-1} - 1$ regular, and contains no $K_{a,b}$ with $b = (a - 1)! + 1$. To show this we use Lemma 3.24 with $t = a - 1$ only.

Theorem 3.26 (Ball and Pepe [19]). *The Alon–Rónyai–T. Szabó graph $H(q, 4)$ does not contain $K_{5,5}$. Hence $\mathbf{ex}(n, K_{5,5}) \geq (\frac{1}{2} + o(1))n^{7/4}$.*

This is better than the earlier lower bounds of $\mathbf{ex}(n, K_{a,b})$ for $a = 5$, $5 \leq b \leq 12$, and $a = 6$, $6 \leq b \leq 8$.

Recently, Blagojević, Bukh, and Karasev [24] gave a new algebraic construction to provide lower bounds on $Z(m, n, a, b)$ matching the (3.1) upper bound. Their example is weaker than the Kollár–Rónyai–Szabó in the sense that it only works for $b > (a^2(a + 1))^a$. On the other hand, they give new insight about the limits of the Algebraic Geometric method on which constructions may and which may not work.

We close this section mentioning that Noga Alon has a survey paper in the Handbook of Combinatorics [7] providing ample information on the topics treated here (i.e., applications of algebra in combinatorics).

3.7. The coefficient in the Kővári–T. Sós–Turán bound

Alon, Rónyai and Szabó [11] observed that their Construction 3.25 can be factored with a t -element subgroup $H \subset GF(q)^*$ (when t divides $q - 1$) in the same way as it was done in Construction 3.15. Namely, the vertex set of the new graph $H^t(q, a)$ are the elements $(x, X) \in GF(q)^*/H \times GF(q^{a-1})$ and (x, X) and (y, Y) are joined if $N(X + Y)x^{-1}y^{-1} \in H$. Then the graph $H^t(q, a)$ has $n = (q - 1)q^{a-1}/t$ vertices, its degrees are about q^{a-1} , and it contains no $K_{a,b}$ for $b = (a - 1)!t^{a-1} + 1$. Let $q \rightarrow \infty$. Then also $n \rightarrow \infty$, and we get that for these fixed values of a and b one gets

$$\mathbf{ex}(n, K_{a,b}) \geq (1 - o(1)) \frac{\sqrt[a]{b-1}}{2 \sqrt[a]{(a-1)!}} n^{2-(1/a)}.$$

This shows that the order of magnitude of the coefficient in the KST bound (3.1) should be indeed $\sqrt[a]{b-1}$.

Montágh [193] found a clever factorization of the Brown graph (using the spherical symmetry of the balls) thus proving the same result with even a slightly better bound than the bound of Alon, Rónyai and Szabó, for $a = 3$.

3.8. Excluding large complete subgraphs

The following theorem was discovered many times because its connections with Computer Science problems: given a graph G with n vertices, there exists a decomposition of its edges into complete balanced bipartite graphs K_{a_i, a_i} having altogether $O(n^2/\log n)$ vertices, $\sum_i a_i = O(n^2/\log n)$. Lately, Mubayi and Gy. Turán [195] gave a polynomial algorithms finding such a subgraph partition efficiently. Strictly speaking, this is not a Turán type problem but their result implies, e.g., that there exists a polynomial algorithm to find a $K_{a,a}$ in a graph of $n^2/4$ edges of size $a = \Theta(\log n)$. The bound $O(\log n)$ is the best possible (shown by the random graph).

It is also not very difficult to show that usually the random graph gives the correct order of the Turán number $\mathbf{ex}(n, K_{a,a})$ for $n, a \rightarrow \infty$ simultaneously.

The case when a, b are very large i.e. $a + b = \Omega(n)$ was considered by Griggs, Quyang, and Ho [126], [125]. In this case $Z(m, n, a, b)$ is almost mn so they considered the dual question. Let us mention only one result of this type by Balbuena, García-Vázquez, Marcote, and Valenzuela, who have more papers on this topic.

Theorem 3.27 (see [14], [16] and the references there). $Z(m, n; a, b) = mn - (m + n - a - b + 1)$ if $\max\{m, n\} \leq a + b - 1$.

There is another direction of research, when the ratio of m and n is extreme. Here we only mention a classical result, that it is easy to solve the case when n is very large compared to m .

Theorem 3.28 (Čulík [55]).

$$Z(m, n, a, b) = (a - 1)n + (b - 1) \binom{m}{a} \quad \text{for } n \geq (b - 1) \binom{m}{a}.$$

4. EXCLUDING CYCLES: C_{2k}

To start with, Bondy wrote a long chapter in the Handbook of Combinatorics [30] and also a very nice survey on Erdős and the cycles of graphs [31].

Let \mathcal{C} be a (finite or infinite) set of cycles. The study of $\mathbf{ex}(n, \mathcal{C})$ is especially interesting if \mathcal{C} has a member of even length. However, constructions of dense graphs without some given even cycles is usually very difficult; the examples use polarities of finite geometries (generalized polygons [171]), or Ramanujan graphs [190], [181] or some other families of polynomials [170].

An odd cycle, C_{2k+1} is chromatically critical. Hence a theorem of Simonovits [221] implies that $\text{ex}(n, C_{2k+1}) = \lfloor \frac{1}{4}n^2 \rfloor$ for $n > n_k$ and the only extremal graph is $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

In this Section we concentrate on even cycles C_{2k} .

4.1. Girth and Turán numbers, upper bounds

What is $\text{ex}(n, \{C_3, C_4, \dots, C_{g-1}\})$, the maximum number of edges in a graph with n vertices and girth g ? This problem can be considered in a dual form, what is the least number of vertices $n = n(d, g)$ in a graph of girth g and an average degree at least d ? If we replace ‘average’ with ‘minimum’ δ then a simple argument gives the so-called *Moore bound* for odd girth:

$$(4.1) \quad |V(G)| = n \geq n_0(\delta, 2k + 1) := 1 + \delta \sum_{0 \leq i \leq k-1} (\delta - 1)^i.$$

Alon, Hoory and Linal [8] showed that (4.1) holds for the average degree, too. Rearranging we have $d_{\text{ave}} < n^{1/k} + 1$, in other words

Theorem 4.1 (Upper bound when the girth is odd).

$$(4.2) \quad \text{ex}(n, \{C_3, C_4, \dots, C_{2k}\}) < \frac{1}{2}n^{1+(1/k)} + \frac{1}{2}n.$$

To prove an upper bound $n^{1+(1/k)}$ is trivial by induction on n . Then (4.2) was improved but with a larger linear additive term.

Theorem 4.2 (Lam and Versträete [166], Excluding only even cycles).

$$(4.3) \quad \text{ex}(n, \{C_4, C_6, \dots, C_{2k}\}) < \frac{1}{2}n^{1+(1/k)} + 2^{k^2}n.$$

They also note that for $k = 2, 3, 5$ the n -vertex polarity graphs of generalized $(k + 1)$ -gons (defined by Lazebnik, Ustimenko and Woldar [171] described below as Construction 4.27) have $\frac{1}{2}n^{1+(1/k)} + O(n)$ edges and have no even cycles of length at most $2k$.

Corollary 4.3 ([166] and [171]) Even girth is 6, 8 or 12). In case of $2k \in \{4, 6, 10\}$ we have

$$(4.4) \quad \text{ex}(n, \{C_4, C_6, \dots, C_{2k}\}) = (1 + o(1))\frac{1}{2}n^{1+(1/k)}.$$

On the other hand, Füredi, Naor and Verstaëte [117] showed that if we exclude only C_{2k} , then $\mathbf{ex}(n, C_6) > 0.53n^{4/3}$ (see below as Construction 4.29) and Lazebnik, Ustimenko, and Woldar [169], showed that $\mathbf{ex}(n, C_{10}) > 0.579n^{6/5}$ (see below as Construction 4.28).

Concerning the Moore bound for even girth we have

$$(4.5) \quad |V(G)| = n \geq n_0(\delta, 2k + 2) := 2 \sum_{0 \leq i \leq k} (\delta - 1)^i.$$

Alon, Hoory and Linial [8] showed that (4.5) holds for the average degree, too. Rearranging, we have $d_{\text{ave}} < (n/2)^{1/k} + 1$, in other words

Theorem 4.4 (Upper bound when the girth is even).

$$(4.6) \quad \mathbf{ex}(n, \{C_3, C_4, \dots, C_{2k+1}\}) < \frac{1}{2^{1+(1/k)}} n^{1+(1/k)} + \frac{1}{2}n.$$

This upper bound with a weaker error term was also proved earlier by Erdős and Simonovits [93].

Note that because of Theorems 3.2, 4.21, and 4.23 one can easily show that asymptotic bound holds in (4.6) for $2k = 4, 6, 10$. The other cases are unsolved.

Theorem 4.5. For $2k = 4, 6$ and 10 as $n \rightarrow \infty$ we have

$$(4.7) \quad \mathbf{ex}(n, \{C_3, C_4, \dots, C_{2k+1}\}) = (1 + o(1)) \frac{1}{2^{1+(1/k)}} n^{1+(1/k)}.$$

Moreover, infinitely many exact values are obtained for $2k = 4, 6, 10$: for $n = 2(q^k + q^{k-1} + \dots + q + 1)$,

$$(4.8) \quad \mathbf{ex}(n, \{C_3, C_4, \dots, C_{2k+1}\}) = (q + 1)(q^k + q^{k-1} + \dots + q + 1)$$

whenever q is a power of a prime.

4.2. Excluding a single C_{2k} , upper bounds

Concerning our central problem, Erdős showed that excluding just one even cycle has essentially the same effect as excluding all smaller cycles as well. This is far from trivial! Erdős never published a proof of his result.

Theorem 4.6 (Erdős, The Even Cycle Theorem).

$$(4.9) \quad \text{ex}(n, C_{2k}) = O(n^{1+(1/k)}).$$

The first proof was published by Bondy and Simonovits in the following stronger form.

Theorem 4.7 (Bondy and Simonovits [32]). *Let G_n be a graph with e edges, and let t satisfy $2 \leq t \leq e/(100n)$ and $tn^{1/t} \leq e/(10n)$. Then G_n contains a C_{2t} .*

In some sense, this is a “pancyclic theorem”: there is a meta-principle, that if some reasonable conditions ensure the existence of a Hamiltonian cycle, then they ensure the existence of all shorter cycles. Here we go the other direction: if we ensure the existence of a C_{2k} , then we ensure the existence of all longer cycles, up to a natural limit, with the natural parity.

Corollary 4.8. *If G_n has at least $100kn^{1+(1/k)}$ edges, then it contains a C_{2t} , for every $t \in [k, kn^{1/k}]$.*

The Erdős–Bondy–Simonovits upper bound together with earlier known constructions imply that the exponent $1 + (1/k)$ is sharp for C_4 (see, e.g., Theorem 3.4), C_6 , and C_{10} (Theorems 4.22 and 4.24 below).

Corollary 4.9 (The only known exact exponents for single cycles).

$$\text{ex}(n, C_4) = \Theta(n^{3/2}), \quad \text{ex}(n, C_6) = \Theta(n^{4/3}), \quad \text{ex}(n, C_{10}) = \Theta(n^{6/5}).$$

The upper end of the interval in Corollary 4.8 is also sharp, apart from the constant 100 take the disjoint union of complete graphs of order $200kn^{1/k}$. We made the following conjecture:

Conjecture 4.10 (Erdős–Simonovits). $\text{ex}(n, C_{2k}) \geq c_k n^{1+(1/k)}$. Moreover,

$$\frac{\text{ex}(n, C_{2k})}{n^{1+(1/k)}}$$

converges to a positive limit.

It is only known for C_4 . A weakening of this conjecture would be the following: Let $\Theta_{k,\ell}$ denote the graph of order $2 + (k - 1)\ell$ in which two vertices are joined by ℓ paths of length k .

Conjecture 4.11 (Simonovits). *For each k there is an $\ell = \ell(k)$ for which $\text{ex}(n, \Theta_{k,\ell}) \geq c_k n^{1+(1/k)}$.*

Perhaps the first very annoying unsolved problem on this area is

Conjecture 4.12. $\text{ex}(n, C_8) \geq c_4 n^{5/4}$.

Returning to the Turán number of C_{2k} , the multiplicative constant of the upper bound in the Bondy–Simonovits theorem was improved by Verstraëte [246] from 100 to 8. The best known upper bound today is that of Oleg Pikhurko:

Theorem 4.13 (Pikhurko, [203]).

$$\text{ex}(n, C_{2k}) \leq (k-1)n^{1+(1/k)} + 16(k-1)n.$$

Historical Remark 1.

(a) Pikhurko, in his very nice paper [203] gives a short description of the whole story.

(b) Pikhurko mentions that the Bondy–Simonovits proof gives a constant 20: originally it was stated as 100. It would be extremely interesting if the upper bound $k-1+o(1)$ for $\text{ex}(n, C_{2k})/n^{1+(1/k)}$ could be improved to $o(k)$.

4.3. Eliminating short cycles, a promising attempt

It was relatively easy to prove the upper bound (4.2) for the number of edges for a graph G_n with girth exceeding $2k$, $e(G_n) = O(n^{1+(1/k)})$. Suppose that G has no C_{2k} . Erdős bipartite subgraph lemma 2.14 states that there is a bipartite subgraph H with $e(H) \geq \frac{1}{2}e(G)$. This way we have eliminated all the odd cycles $C_3, C_5, \dots, C_{2k-1}$ from G . It is a natural to ask whether one can eliminate other short cycles, thus obtaining an easy proof for the Erdős–Bondy–Simonovits upper bound, (4.9).

Problem 4.14. Is it true that there exists a constant $\alpha_{2k} > 0$ such that each C_{2k} -free G_n contains an H_n with $\text{girth}(H_n) > 2k$ and $e(H_n) > \alpha_{2k}e(G_n)$?

The answer is still unknown. The first step was done by E. Györi. The following lemma implies that α_6 exists and it is at least $1/4$.

Lemma 4.15 (Györi [134]). *If G_n is bipartite and it does not contain any C_6 , then it contains an H_n with*

$$e(H_n) \geq \frac{1}{2}e(G_n) + 1,$$

not containing C_4 's either (for $e(G) \geq 2$). This is sharp only for $G_n = K_{2,n-2}$.

We mention two generalizations.

Theorem 4.16 (Füredi, Naor and Verstraete [117]). *Let G be a hexagon-free graph. Then there exists a subgraph of G of girth at least five, containing at least half the edges of G .*

Furthermore, equality holds if and only if G is a union of edge-disjoint complete graphs of order four or five. We got $\alpha_6 = 1/2$.

Theorem 4.17 (Getting rid of C_4 's, Kühn and Osthus [165]). *Every bipartite C_{2k} -free graph G contains a C_4 -free subgraph H with $e(H) \geq e(G)/(k-1)$.*

The factor $1/(k-1)$ is best possible, as the example $K_{k-1, n-k+1}$ shows.

These theorems settle some special cases (namely $\mathcal{L} = \{C_4, C_{2k}\}$) of the following compactness conjecture of Erdős and Simonovits.

Conjecture 4.18 (Compactness. Erdős–Simonovits [93]). *For every finite family of graphs \mathcal{L} (containing bipartite members as well) there exists an $L_0 \in \mathcal{L}$ for which $\mathbf{ex}(n, \mathcal{L}) = O(\mathbf{ex}(n, L_0))$.*

The following result of Kühn and Osthus makes a little step toward solving Problem 4.14 and Conjecture 4.18.

Theorem 4.19 ([165]). *Let $g \geq 4$ be an even integer and let $\ell(g) =: \prod_{1 \leq i \leq g/2} i$. Suppose that $k-1$ is divisible by $\ell(g)$ and G_n is a C_{2k} -free graph. Then G_n contains an H_n with $\mathbf{girth}(H_n) > g$ such that $e(H_n) \geq e(G_n)/2(4k)^{(g-2)/2}$.*

In other words, for some very special values of k 's a C_{2k} -free graph contains a subgraph having a positive fraction of the edges and of girth at least $\Omega(\log k / \log \log k)$.

4.4. A lower bound for C_6 : The Benson Construction

In the preceding section, we asserted that the Erdős theorem on even circuits is sharp for C_4 , C_6 and C_{10} (and is conjectured to be sharp in all cases). For C_4 , the sharpness follows from Construction 3.5. For C_6 , it can be deduced from the Benson construction [22] which we explain below. Note that (about the same time) Singleton [231] described the same graph but his definition was much more complicated.

The points of the d -dimensional finite projective geometry $PG(d, q)$ are the equivalence classes of the nonzero vectors of $GF(q)^{d+1}$ where \mathbf{x} and \mathbf{y} are equivalent if there is a $\gamma \in GF(q)^*$ such that $\mathbf{x} = \gamma\mathbf{y}$. There are $(q^{d+1} -$

$1)/(q-1)$ such classes. Then the i -dimensional subplanes are generated by the $(i+1)$ -dimensional subspaces of the vector space $GF(q)^{d+1}$.

Let

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Clearly, A is non-singular. Define the surface S by the equation $\mathbf{x}A\mathbf{x}^T = 0$,

$$S := \{\mathbf{x} \in PG(4, q) : \mathbf{x}A\mathbf{x}^T = 0\}$$

Construction 4.20 (Benson's C_6 -free bipartite graph). Let \mathcal{L} be the set of lines of $PG(4, q)$ contained entirely in S . The vertex set of the bipartite graph B_q is $S \cup \mathcal{L}$, and $\mathbf{x} \in S$ is joined to $L \in \mathcal{L}$ if $\mathbf{x} \in L$.

Theorem 4.21 (Benson [22], Singleton [231]). B_q is a $(q+1)$ -regular, bipartite, girth 8 graph with $2(q^3 + q^2 + q + 1)$ vertices.

Corollary 4.22. $\text{ex}(n, C_6) \geq (1 + o(1))(n/2)^{4/3}$.

First, we can see that S does not contain a full 2-dimensional projective plane. We can use the fact that for \mathbf{x} and \mathbf{y} on S , the line \mathbf{xy} consists of the points $\mathbf{z} = a\mathbf{x} + (1-a)\mathbf{y}$, and lies entirely in S if both $\mathbf{y}A\mathbf{y}^T = 0$ and $\mathbf{x}A\mathbf{y}^T = 0$.

Second, the number of lines from \mathcal{L} containing a given point $\mathbf{x} \in S$ is $q+1$. Since the number of points on a line is $q+1$ we immediately get that $|S| = |\mathcal{L}|$.

Furthermore, B_q contains no cycles of length 3, 4, 5 or 7. (For the odd cases this is because it is bipartite, and the existence of a 4-cycle would imply that two points of S are on two distinct lines.) Now suppose that B_q contains a 6-cycle $v_1w_1v_2w_2v_3w_3v_1$. Then S must contain the three lines v_1v_2 , v_2v_3 , and v_3v_1 , and so it must contain the plane $\langle v_1v_2v_3 \rangle$. But this is impossible. If we apply a coordinate transformation T with v_1 , v_2 and v_3 as the first three base vectors, we get the matrix

$$\begin{pmatrix} 0 & 0 & 0 & ? & ? \\ 0 & 0 & 0 & ? & ? \\ 0 & 0 & 0 & ? & ? \\ ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? \end{pmatrix}$$

since $v_i Av_j^T = 0$. But then A' cannot be regular, contradicting the regularity of A . Hence B_q cannot contain C_6 either.

All these imply that $|S| = q^3 + q^2 + q + 1$ and that every $\mathbf{x} \in S$, $L \in \mathcal{L}$ if $\mathbf{x} \notin L$ then there exists a unique line $L' \in \mathcal{L}$ such that $\mathbf{x} \in L'$ and $L \cap L' \neq \emptyset$.

In concluding this section, we note that finite geometry constructions can also be used in hypergraph extremal problems (see [39], [40] and [220]).

4.5. Girth 12 graphs by Benson and by Wenger

Theorem 4.23 (Benson [22]). *Let q be an odd prime power. There is a $(q + 1)$ -regular, bipartite, girth 12 graph B_q^* with $2(q^5 + q^4 + q^3 + q^2 + q + 1)$ vertices.*

Corollary 4.24. $\text{ex}(n, C_{10}) \geq (1 + o(1))(n/2)^{6/5}$.

One half of the vertex set of B_q^* are the points of the quadric Q_6 in $PG(6, q)$ defined by $x_0^2 + x_1x_{-1} + x_2x_{-2} + x_3x_{-3} = 0$. Its size is exactly $(q^6 - 1)/(q - 1)$. Then we select a set of lines \mathcal{L} contained entirely in Q_6 and covering each point of Q_6 exactly $q + 1$ times. The family \mathcal{L} is selected as follows: If $\mathbf{x} \in Q_6$ and $\mathbf{x}, \mathbf{y} \in L \in \mathcal{L}$ then \mathbf{x} and \mathbf{y} must satisfy the following six bilinear equations:

$$x_0y_i - x_iy_0 + x_{-j}y_{-k} - x_{-k}y_{-j} = 0$$

where (i, j, k) is a cyclic permutation of $(1, 2, 3)$ or $(-1, -2, -3)$.

Construction 4.25. The bipartite graph B_q^* is defined, as before, by the incidences $\mathbf{x} \in L$.

Now consider the much simpler example of Wenger.

Construction 4.26 (Wenger [247]). Let p be a prime, $k = 2, 3$ or 5 . $H_k(p)$ is defined as a bipartite graph with two vertex classes \mathbf{A} and \mathbf{B} , where $|\mathbf{A}| = |\mathbf{B}| = p^k$ and the vertices of \mathbf{A} are k -tuples $\mathbf{a} = (a_0, a_1, \dots, a_{k-1}) \in GF(p)^k$ and same for $\mathbf{b} = (b_0, b_1, \dots, b_{k-1}) \in \mathbf{B}$. The vertices \mathbf{a} and \mathbf{b} are joined if

$$b_j \equiv a_j + a_{j+1} \cdot b_{k-1} \pmod{p} \text{ for } j = 0, 1, \dots, k - 2.$$

One can see that for every $\mathbf{a} \in \mathbf{A}$ each b_{k-1} determines exactly one $\mathbf{b} \in \mathbf{B}$ joined to it. This easily implies that $G[\mathbf{A}, \mathbf{B}]$ is p -regular, with $n = 2p^k$ vertices and $p^{k+1} = (n/2)^{1+(1/k)}$ edges.

Wenger gives an elegant proof of that $H_2(p)$ has no C_4 , $H_3(p)$ has no C_4 , nor C_6 . Finally, $H_5(p)$ contains no C_4, C_6 or C_{10} , however, it has many C_8 's.

4.6. Short cycles, C_6 and C_{10}

The densest constructions of $2k$ -cycle-free graphs for certain small values of k arise from the existence of rank two geometries called *generalized d -gons*. These may be defined as rank two geometries whose bipartite incidence graphs are regular graphs of diameter d and girth $2d$. These are known to exist when d is three, four or six. This is the background of the above Constructions 4.20 and 4.25.

Construction 4.27 (Lazebnik, Ustimenko and Woldar [171]). One can use the existence of polarities of the generalized $(k+1)$ -gons to obtain dense $2k$ -cycle-free graphs when $k \in \{2, 3, 5\}$. In particular, for these k 's

$$(4.10) \quad \mathbf{ex}(n, C_{2k}) \geq \frac{1}{2}n^{1+(1/k)} + O(n)$$

for infinitely many n .

In [93], Erdős and Simonovits formulated the following conjecture. For fixed k and $n \rightarrow \infty$, $\mathbf{ex}(n, C_{2k}) = \frac{1}{2}n^{1+(1/k)} + o(n^{1+(1/k)})$. This holds for C_4 (Theorem 3.4), but was disproved first for C_{10} , then for C_6 by the following two examples.

Construction 4.28 (Lazebnik, Ustimenko and Woldar [171]). Consider a bipartite graph $G[A, B]$ of girth exceeding $2k$. Replace each vertex of A by $k-1$ new vertices with the same neighborhood. Then the new graph $G[(k-1)A, B]$ is still C_{2k} -free. In particular, starting with the girth 12 bipartite graph of Theorem 4.23 (here $k=5$) one gets a graph of about $5q^5$ vertices and about $4q^6$ edges, implying

$$(4.11) \quad \mathbf{ex}(n, C_{10}) \geq 4(n/5)^{6/5} > 0.5798n^{6/5}$$

for infinitely many n .

Since the C_6 -free graph of Construction 4.27 does not have C_3 and C_4 either, doubling a random subset appropriately, one obtains a denser C_6 -free graph:

Theorem 4.29 (Füredi, Naor and Versträete [117]). *For infinitely many n ,*

$$\mathbf{ex}(n, C_6) > \frac{3(\sqrt{5}-2)}{(\sqrt{5}-1)^{4/3}}n^{4/3} + O(n) > 0.5338n^{4/3}.$$

They also showed that

Theorem 4.30 (Füredi, Naor and Verstraëte [117]). $\mathbf{ex}(n, C_6) \leq \lambda n^{4/3} + O(n)$, where $\lambda \approx 0.6271$ is the real root of $16\lambda^3 - 4\lambda^2 + \lambda - 3 = 0$.

These theorems give the best known lower and upper bounds for $\mathbf{ex}(n, C_6)$. The proof of Theorem 4.30 requires a statement about hexagon-free bipartite graphs, which is interesting in its own right (see de Caen and Székely [44]). Let $\mathbf{ex}(m, n, C_6)$ be the maximum number of edges amongst all m by n bipartite hexagon-free graphs. Then

Theorem 4.31 (Füredi, Naor and Verstraëte [117]). *Let m, n be positive integers with $n \geq m$. Then*

$$\mathbf{ex}(m, n, C_6) < 2^{1/3}(mn)^{2/3} + 10n.$$

Furthermore, if $n = 2m$ then as n tends to infinity,

$$\mathbf{ex}(m, n, C_6) = \begin{cases} 2^{1/3}(mn)^{2/3} + O(n) & \text{for infinitely many } m \\ 2^{1/3}(mn)^{2/3} - o(n^{4/3}) & \text{for all } m. \end{cases}$$

The lower bound is given by the graph defined in Construction 4.28 starting with the Benson graph (Theorem 4.21, $k = 3$).

4.7. Bipartite hosts with extreme sides

We have already seen two such results concerning the Zarankiewicz number, by Reiman (Theorem 3.2) and Čulík (Theorem 3.28). András Sárközy and Vera Sós formulated the following conjecture²⁵

Conjecture 4.32.

$$\mathbf{ex}(m, n, C_6) < 2n + c(nm)^{2/3}.$$

A weaker version of this was proved by Gábor N. Sárközy, [212] and later Győri [134] proved a stronger

Theorem 4.33. *There exists a constant $c_k > 0$ for which if $G[A, B]$ is a bipartite graph with color classes A, B , and $|A| = m, |B| = n \geq m^2$, and*

$$e(G[A, B]) \geq (k - 1)n + c_k m^2,$$

then $G[A, B] \supset C_{2k}$.

²⁵A weaker version of this conjecture was formulated by Erdős several years earlier.

This means that for $n > m^2$ the extremal number becomes linear. For more recent results see, e.g., Balbuena, García-Vázquez, Marcote, and Valenzuela [15]. Later Györi [135] showed that $c_3 = 1/8$, proving

$$\mathbf{ex}(m, n, C_6) \leq 2n + \frac{1}{8}m^2,$$

for $n, m > 100$, $n \geq m^2/16$ and here equality holds if m is a multiple of 4.

4.8. The effect of odd cycles

Let \mathcal{L} be a set of graphs and let $\mathbf{ex}_{\text{bip}}(n, \mathcal{L})$ denote the *bipartite Turán number* of \mathcal{L} , the size of the largest \mathcal{L} -free bipartite graph on n vertices.

Theorem 4.34 (Erdős and Simonovits [93]).

$$\mathbf{ex}(n, \{C_4, C_5\}) = (1 + o(1))\mathbf{ex}_{\text{bip}}(n, C_4) = (1 + o(1))(n/2)^{3/2}.$$

They also **conjecture** that the same holds for $\{C_3, C_4\}$ (i.e., for the girth problem) but this is still unsolved. Then, they make the following bold conjecture.

Conjecture 4.35 (Erdős and Simonovits [93] on the effect of odd cycles). *Let $\mathcal{C}_{2\ell+1}^{\text{odd}}$ denote the set of odd cycles $\{C_3, C_5, \dots, C_{2\ell+1}\}$. For any family \mathcal{L} consisting of bipartite graphs there exists an odd integer $2\ell + 1$ such that $\mathbf{ex}(n, \mathcal{L} \cup \mathcal{C}_{2\ell+1}^{\text{odd}}) \approx \mathbf{ex}_{\text{bip}}(n, \mathcal{L})$.*

This conjecture was verified in a few cases by extending and sharpening Theorem 4.34 as follows.

Theorem 4.36 (Keevash, Sudakov and Verstraëte [157]). *Let $\mathcal{C}_{2k}^{\text{even}}$ denote the set of even cycles $\{C_4, C_6, \dots, C_{2k}\}$. Suppose that $2k \in \{4, 6, 10\}$ and suppose that $2\ell + 1 > 2k$. Then*

$$\mathbf{ex}(n, \mathcal{C}_{2k}^{\text{even}}, C_{2\ell+1}) = (1 + o(1))\mathbf{ex}_{\text{bip}}(n, \mathcal{C}_{2k}^{\text{even}}) \sim (n/2)^{1+(1/k)}.$$

They even proved a stability result (when $n \rightarrow \infty$) and, using it, an exact version: If $2k \in \{4, 6, 10\}$ and $2\ell + 1 \geq 5, 15$, or 23 , respectively, and $n = 2(q^k + q^{k-1} + \dots + q + 1)$ then for $n > n_{2\ell+1}$ we have

$$\mathbf{ex}(n, \mathcal{C}_{2k}^{\text{even}} \cup C_{2\ell+1}) \leq (q + 1)n$$

and here equality holds only if there is a generalized $(k + 1)$ -gon of order q .

In a more recent work Allen, Keevash, Sudakov and Verstraëte [5] verified the stronger form of the Erdős–Simonovits conjecture proving that for any fixed $2\ell + 1 \geq 5$ one has $\text{ex}(n, \{K_{2,t}, C_{2\ell+1}\}) \sim \text{ex}_{\text{bip}}(n, K_{2,t})$ and $\text{ex}(n, \{K_{3,3}, C_{2\ell+1}\}) \sim \text{ex}_{\text{bip}}(n, K_{3,3})$. They also show

$$\text{ex}(n, \{K_{2,t}, B_t, C_{2\ell+1}\}) \sim \text{ex}_{\text{bip}}(n, \{K_{2,t}, B_t\}) \sim (n/2)^{3/2}$$

for any fixed $t \geq 2$ and $2\ell + 1 \geq 9$, where B_t is a “book” of t C_4 ’s sharing an edge: it has $2t + 2$ vertices and $3t + 1$ edges. Their main tool is the smoothness of the corresponding Turán number’s and the sparse regularity lemma of A. Scott [213].

On the other hand, for any $t \geq 1$ and prime $q > 2^{t^4}$, they construct $(t + 2)$ -partite graphs $G_{q,t}$ with no triangle or $K_{2,2t+1}$ having $n = (t + 2)q^2$ vertices and $\binom{t+2}{2}q^2(q - 1)$ edges. This implies

$$(4.12) \quad \text{ex}(n, \{K_{2,2t+1}, C_3\}) \geq (1 + o(1)) \frac{t + 1}{\sqrt{t + 2}} n^{3/2}.$$

So, using $\text{ex}_{\text{bip}}(n, K_{2t+1}) \sim \sqrt{tn}^{3/2}$, which follows easily from (3.7) and (3.9), they obtain

$$(4.13) \quad \liminf_{n \rightarrow \infty} \frac{\text{ex}(n, \{K_{2,2t+1}, C_3\})}{\text{ex}_{\text{bip}}(n, K_{2,2t+1})} \geq \frac{t + 1}{\sqrt{t(t + 2)}} > 1.$$

In particular the ratio is $2/\sqrt{3} + o(1)$ for $K_{2,3}$. We explain their construction yielding (4.12) only for $t = 1$.

Construction 4.37 (Allen, Keevash, Sudakov and Verstraëte [5]). Let $q \equiv 2 \pmod{3}$ be a prime. Let G^q be a three-partite graph with parts A_1, A_2 and A_3 which are copies of $GF(q) \times GF(q)$. Join $(x_1, x_2) \in A_i$ to $(y_1, y_2) \in A_{i+1}$ if

$$(y_1, y_2) = (x_1, x_2) + (a, a^2)$$

for some $a \in GF(q)$, $a \neq 0$.

The obtained graph is $K_{2,3}$ and C_3 -free, and has $n = 3q^2$ vertices and $n^{3/2}/\sqrt{3}$ edges. This yields the ratio $2/\sqrt{3} + o(1)$ for $K_{2,3}$ in (4.13). They believe that Erdős’ Conjecture 1.8 is false:

Conjecture 4.38 ([5]).

$$\liminf_{n \rightarrow \infty} \frac{\text{ex}(n, \{C_3, C_4\})}{\text{ex}_{\text{bip}}(n, C_4)} > 1.$$

4.9. Large girth: Ramanujan graphs

Until this point we were fixing the excluded subgraphs. However, there is a subcategory of extremal graph problems, which we could also call “Parametrized Extremal Graph Problems”. Instead of defining them we give an almost trivial but important example: Horst Sachs and Erdős [87] reformulated the Moore bounds (4.1)–(4.5), in a slightly different form.

Theorem 4.39. *If the minimum degree of G_n , $d := d_{\min}(G_n) > 2$ then G_n contains a C_ℓ with*

$$(4.14) \quad \ell < \frac{2 \log n}{\log(d-1)}.$$

Here we arrived at an area where some constructions (for lower bounds) were needed, and the lower bounds were easily obtained by probabilistic arguments; however they were very difficult to obtain them in a constructive way. Instead of going into details, we mention a result of Margulis [188] that (4.14) is sharp up to a constant: there are – not too complicated – Cayley graphs of constant (even) degrees d and girth at least $c \log_{d-1} n$. Here – surprisingly, Margulis’ construction is better than the random graph and a construction of Imrich yields an even better constant c :

Theorem 4.40 (Imrich [147]). *For every integer $d > 2$ one can (effectively) construct infinitely many d -regular Cayley graphs X_n with*

$$\text{girth}(X_n) > 0.4801 \frac{\log n}{\log(d-1)} - 2.$$

The next step in this area was a much deeper and more important results of Margulis [190, 189], Lubotzky, Phillips and Sarnak, [181] on the Expander graphs, that are eigenvalue-extremal. In this sense the Margulis–Lubotzky–Phillips–Sarnak graph is very nice. There is only one problem with it. While defining these graphs is non-trivial, but not extremely complicated, to verify their extremal properties requires deep mathematical tools. Below we give a very compressed description of it.

Definition 4.41. Given a connected k -regular graph X , we denote by $\lambda(X)$ the largest of the absolute values of eigenvalues of the adjacency matrix of X , different from k . An n -vertex k -regular graph $X_{n,k}$ is a *Ramanujan graph* if $\lambda(X_{n,k}) \leq 2\sqrt{k-1}$.

Remark 4.42. In case of k -regular graphs, the largest absolute values of the eigenvalues is k . The bipartite graphs have the property that if λ_i is eigenvalue, then $-\lambda_i$ is also an eigenvalue. By the Alon–Boppana inequality, (see Proposition 4.2 of [181])

$$\liminf_{n \rightarrow \infty} \lambda(X_{n,k}) = 2\sqrt{k-1}.$$

Ramanujan graphs are important because they are expander graphs, which are extremely important in Theoretical Computer Science.

There are quite a few cases, where – instead of using “random graph constructions” one tries to use Cayley Graphs. *Cayley graphs* are graphs whose vertices are the elements of some group \mathcal{G} and the edges are the pairs $(g, \alpha_i g)$, where $g \in \mathcal{G}$ and $\alpha_1, \dots, \alpha_k$ are elements of \mathcal{G} . If we look for a digraph, then this is a correct definition. However, if we are looking for ordinary graphs, then we have to assume that $S := \{\alpha_1, \dots, \alpha_k\}$ is closed under taking the inverse: if $\alpha \in S$ then $\alpha^{-1} \in S$ as well. If we choose \mathcal{G} and S appropriately, then the obtained graph will provide us with nice constructions; mainly, because it behaves as if it were a random graph, or, occasionally, even better.

Construction 4.43 ([181]). Let p and q be unequal primes congruent to 1 mod 4. The Ramanujan graphs $X^{p,q}$ of [181] are $p+1$ -regular Cayley graphs²⁶ of the group $\mathbf{PSL}(2, \mathbb{Z}/q\mathbb{Z})$: $p+1$ generators of the group are fixed, which are obtained from the solutions of

$$(4.15) \quad p = a^2 + b^2 + c^2 + d^2, \quad \text{where } a > 0 \text{ is odd and } b, c, d \text{ are even.}$$

The number of solutions of (4.15) is connected to the famous Ramanujan conjecture, which is still open. However, good approximations are known, by Eichler and Igusa, enough for the purposes of [181]. Originally most of the authors were interested in the eigenvalue properties (spectral gap) of these graphs, that are also strongly connected to them being expander graphs (see Alon, [6], Alon–Milman [10]).

From here on, $X_{n,k} = X^{p,q}$ is a special sequence of Ramanujan graphs, which is non-bipartite if the Jacobi symbol $\left(\frac{q}{p}\right) = 1$; then it has $n = (q^3 - q)/2$ vertices.

Theorem 4.44. For $k = p+1$, $X^{p,q}$ is k -regular, its eigenvalues are $\lambda = \pm k$ or $|\lambda| \leq 2\sqrt{k-1}$.

²⁶There are two of them, a bipartite and a non-bipartite, we forget the bipartite one.

This property is optimal and leads to the best known explicit expander graphs. Alon turned the attention of the authors to that these graphs satisfy a number of extremal combinatorial properties.

Theorem 4.45 (Observation of Alon). *The girth of $X_{n,k}$ is asymptotically $\geq \frac{4}{3} \frac{\log n}{\log(k-1)}$.*

This gives larger girth than what was previously known by explicit or non-explicit constructions. Also, it is one of the “cleanest” way to define graphs with large girth and high chromatic number:

Theorem 4.46 ([181]). *If $X_{n,k}$ is a non-bipartite Ramanujan graph, then its independence number and chromatic number satisfy*

$$\alpha(X_{n,k}) \leq \frac{2\sqrt{k-1}}{k}n \quad \text{and} \quad \chi(X_{n,k}) \geq \frac{k}{2\sqrt{k-1}}.$$

For a more informative description of these and many other related areas see the survey of Alon in the Handbook [7].

4.10. The girth problem: the Lazebnik–Ustimenko approach

After 20 years Theorem 4.47 still yields the best known lower bound for the girth problem: Lazebnik, Ustimenko and Woldar’s work [170] gives a slight improvement (an $O(1)$ in the denominator of the exponent) to what we can get from the Ramanujan’ graphs.

Theorem 4.47 ([170]). $\text{ex}(n, \{C_3, C_4, \dots, C_{2k+1}\}) = \Omega(n \cdot n^{2/(3k-3+\varepsilon)})$ where $k \geq 2$ is fixed, $\varepsilon = 0$ if k is odd, $\varepsilon = 1$ if k is even and $n \rightarrow \infty$.

We have seen basically two approaches on how to *construct* graphs with high girth. One was the use of Finite Geometries, and the other the use of Cayley Graphs of some matrix groups (Ramanujan graphs). There is (at least) one further important approach to this question which we find in the works of Lazebnik and Ustimenko and later Lazebnik, Ustimenko and Woldar.

Remark 4.48 (History). In this survey many important areas had to be skipped. One of them is the family of Lazebnik–Ustimenko type algebraic constructions. This family of constructions is much more flexible than many earlier ones, and provides a lot of new constructions in extremal graph theory, in Ramsey type problems, for graphs and hypergraphs as well. The first results were achieved by Lazebnik and Ustimenko [168]. Lazebnik and his coworkers created a school in this area. The reader is referred here to [167].

The main feature of this approach can be described (perhaps slightly cheating) as follows. We take a set R (a finite or infinite ring or field), its d^{th} power, and a sequence of polynomials f_2, \dots, f_d . Define a bipartite graph, where the colour classes A and B consist of vectors (a_1, \dots, a_d) and (b_1, \dots, b_d) that are joined if

$$\begin{aligned} a_2 + b_2 &= f_2(a_1, b_1) \\ a_3 + b_3 &= f_3(a_1, b_1, a_2, b_2) \\ &\dots \\ a_d + b_d &= f_d(a_1, b_1, \dots, a_d, b_d). \end{aligned}$$

We may also identify A and B to get non-bipartite graphs as well. In general, either we get digraphs, or some symmetry conditions are assumed on the functions f_i , ensuring that if (a_1, \dots, a_d) is joined to (b_1, \dots, b_d) , then (b_1, \dots, b_d) and (a_1, \dots, a_d) are joined as well. Yet, it is not an easy area to describe it on a few pages: this is why we basically skip it. Perhaps the more interested reader should look at [172].

4.11. Cycle length distribution

As a measure of the density of the cycle lengths in a graph G , Erdős introduced the number $L(G)$, the sum of the reciprocals of the distinct cycle lengths of G . The following beautiful theorem, due to Gyárfás, Komlós and Szemerédi, proves a conjecture of Erdős and Hajnal, asserting that in some sense the complete graph or the complete bipartite graph are the densest concerning cycle lengths:

Theorem 4.49 ([131]). *There exists a positive constant $c > 0$ such that if $d_{\min}(G) \geq k$, then for the sum of the reciprocals of the cycle lengths ℓ_i of G we have*

$$L(G) = \sum \frac{1}{\ell_i} > c \log k.$$

The union of complete graphs K_{k+1} or bipartite graphs $K_{k,m}$ (where $m \geq k$) show that this lower bound is sharp.

Generalizing a theorem of Bondy and Vince [33], Gengua Fan proved several nice results on the distribution of cycle lengths. We mention only one of them.

Theorem 4.50 (G. Fan [96]). *Let xy be an edge in a 2-connected graph G , k be a positive integer and suppose that all the vertices of G but x and y have degrees at least $3k$. Then xy is contained in $k + 1$ cycles C^0, C^1, \dots, C^k , such that $k + 1 < |E(C^0)| < |E(C^1)| < \dots < |E(C^k)|$, $|E(C^i)| - |E(C^{i-1})| = 2$ for $i = 1, \dots, k - 1$ and $1 \leq |E(C^k)| - |E(C^{k-1})| \leq 2$.*

A related result concerning k odd cycle lengths can be found in Gyárfás [130].

Next we recall a conjecture of Burr and Erdős.

Conjecture 4.51 (Burr and Erdős). *For every odd integer $k > 0$, and every integer ℓ , there exists a c_k such that if $e(G_n) > c_k n$, then some $m \equiv \ell \pmod{k}$, we have $C_m \subseteq G_n$.*

This was proved by Bollobás [25] with $c_k \leq ((k+1)^k - 1)/k$. Häggkvist and Scott ([137], [138]) decreased c_k and extended the Bollobás result, proving that every graph G_n with minimum degree at least $300k^2$ contains k cycles of consecutive even lengths. Soon after, the right order of magnitude of c_k was established.

Theorem 4.52 (Verstraëte [246]). *Let G_n be a graph with $e(G_n) \geq 4kn$. Then there are cycles of k consecutive even lengths in G_n .*

We close this part with the following theorem:

Theorem 4.53 (Sudakov, Verstraëte [234]). *Let $\text{girth}(G_n) = g$ be fixed and $d = 2e(G_n)/n$. Let $\mathcal{C}(G)$ denote the set of cycle-lengths in G . Then $\mathcal{C}(G_n)$ contains at least $\Omega(d^{\lfloor (g-1)/2 \rfloor})$ consecutive even integers, as $d \rightarrow \infty$.*

5. PATHS AND LONG CYCLES

In this section we shall describe results connected with $\text{ex}(n, P_k)$, $\text{ex}(n, \mathcal{C}_{\geq k})$, (where the cycles of at least k vertices are excluded). This problem was proposed by Turán and the (asymptotic) answer were given by Erdős–Gallai.

5.1. Excluding long cycles

Theorem 5.1 (Erdős and Gallai [80]). *Let G_n be a graph with more than $\frac{1}{2}(k-1)(n-1)$ edges, $k \geq 3$. Then G_n contains a cycle of length at least k . This bound is the best possible if $n-1$ is divisible by $k-2$.*

A matching lower bound $\frac{1}{2}(k-1)n - O(k^2)$ can be obtained gluing together complete graphs of sizes at most $k-1$. If k is odd, then there are nearly extremal graphs having a completely different structure. Namely, one can take a complete bipartite graph with partite sets A and B of sizes $|A| = \frac{k-1}{2}$ and $|B| = n - \frac{k-1}{2}$ and add all edges in A , too.

The exact value was determined by Woodall [249] and independently and at the same time by Kopylov [162].

Theorem 5.2 ([162], [249]). *Let $n = m(k - 2) + r$, where $1 \leq r \leq k - 2$, $k \geq 3$, $m \geq 1$ integers. If*

$$e(G_n) > m \binom{k-1}{2} + \binom{r}{2},$$

then G_n contains a cycle of length at least k , and this bound is the best possible:

$$(5.1) \quad \mathbf{ex}(n, \mathcal{C}_{\geq k}) = \frac{1}{2}(k-1)n - \frac{1}{2}r(k-r).$$

Caccetta and Vijayan [43] gave an alternative proof of the result. We need a definition.

Construction 5.3. Let $H_{n,k,s}$ be an n -vertex graph consisting of a complete graph K_{k-s} on the set $A \cup B$, $|A| = k - 2s$, $|B| = s$ and a complete bipartite graph $K_{s,n-(k-s)}$ with parts B and C where A , B and C form a partition of $V(H)$ (hence $|C| = n - (k - s)$ and $n \geq k$, $(k - 1)/2 \geq s \geq 1$).

The graph H contains no cycle of size k or larger and for $s \geq 2$ it is 2-connected. Denote its size by $h(n, k, s)$.

They all ([162], [249], [43]) characterized the structure of the extremal graphs in Theorem 5.2. Namely either

- the blocks of G_n are m complete graphs K_{k-1} and a K_r , or
- k is odd, $r = (k + 1)/2$ or $(k - 1)/2$ and q of the blocks of G_n are K_{k-1} 's and a copy of a $H_{n-q(k-2),k,(k-1)/2}$.

The strongest result on the field is due to Kopylov who also investigated 2-connected graphs.

Theorem 5.4 (Kopylov [162]). *Suppose that $n \geq k \geq 5$ and the 2-connected graph G_n contains no cycles of length of k or larger. Then*

$$e(G_n) \leq \max\{h(n, k, 2), h(n, k, \lfloor \frac{1}{2}(k-1) \rfloor)\}$$

and this bound is the best possible.

Moreover, only the graphs $H_{n,k,s}$ could be extremal, $s \in \{2, \lfloor (k - 1)/2 \rfloor\}$.

This theorem was also conjectured by Woodall [249] and he also proved it for $n \geq (3k - 5)/2$. It was also reproved much later in [97].

5.2. Excluding P_k

One of the oldest problems is the question of determining $\text{ex}(n, P_k)$.

Theorem 5.5 (Erdős and Gallai [80]). *If G_n is a graph containing no P_k , ($k \geq 2$), then*

$$e(G_n) \leq \frac{k-2}{2}n$$

with equality if and only if $k-1$ divides n and all connected components of G are complete graphs on $k-1$ vertices.

Consider the n -vertex graph G_n which is the union of $\lfloor n/(k-1) \rfloor$ vertex-disjoint K_{k-1} and a K_r ($0 \leq r \leq k-2$). If T_k is any connected k -vertex graph, then $T_k \not\subseteq G_n$. Hence

$$(5.2) \quad \text{ex}(n, T_k) \geq \frac{k-2}{2}n - \frac{1}{8}k^2.$$

In particular,

$$(5.3) \quad \text{ex}(n, P_k) \geq \frac{k-2}{2}n - \frac{1}{8}k^2.$$

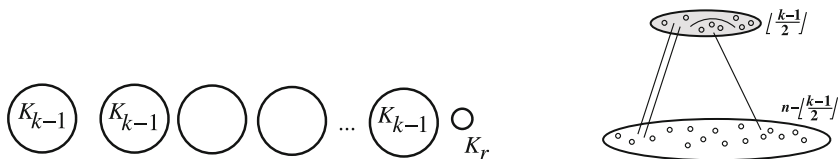


Fig. 2. Potential extremal graphs

If k is even, then there are nearly extremal graphs having a completely different structure. Namely, one can take a complete bipartite graph with partite sets A and B of sizes $|A| = \frac{k-2}{2}$ and $|B| = n - \frac{k-2}{2}$ and add all edges in A , too (Fig. 2). Faudree and Schelp [98] proved that the extremal graph for P_k can indeed be obtained in this way for all n and k . They needed this to prove some Ramsey theorems on paths. The variety of extremal graphs makes the solution difficult.

Theorem 5.6 (Faudree and Schelp [98] and independently Kopylov [162]). *Let $n \equiv r \pmod{k-1}$, $0 \leq r < k-1$, $k \geq 2$. Then*

$$(5.4) \quad \text{ex}(n, P_k) = \frac{1}{2}(k-2)n - \frac{1}{2}r(k-1-r).$$

Faudree and Schelp also described the extremal graphs which are either
 — vertex disjoint unions of m complete graphs K_{k-1} and a K_r , or
 — k is even and $r = k/2$ or $k/2 - 1$ and another extremal graphs can be obtained by taking a vertex disjoint union of t copies of K_{k-1} ($0 \leq t < m$) and a copy of $H(n - t(k - 1), k/2, k/2 - 1)$.

Theorem 5.7 (Kopylov [162]). *Let G_n be a connected graph containing no P_k , ($k \geq 4$) and $n \geq k$. Then*

$$e(G) \leq \max\{h(n, k - 1, 1), h(n, k - 1, \lfloor \frac{1}{2}(k - 2) \rfloor)\}$$

and this bound is the best possible.

Moreover, only the graphs $H_{n,k-1,s}$ could be extremal, $s \in \{1, \lfloor (k - 2)/2 \rfloor\}$.

Balister, Győri, Lehel and Schelp [18] also provided the extremal structures.

5.3. Proof ideas

Let $\mathbf{ex}_{\text{con}}(n, P_k)$ be the maximum number of edges in connected, n -vertex, P_k -free graphs, and let $\mathbf{ex}_{2\text{-con}}(n, \mathcal{C}_{\geq k})$ denote the maximum number of edges in 2-connected, n -vertex, $\mathcal{C}_{\geq k}$ -free graphs. Determining these functions give upper bounds for $\mathbf{ex}(n, P_k)$ and $\mathbf{ex}(n, \mathcal{C}_{\geq k})$.

Indeed, every P_k -free graph is a vertex disjoint union of P_k -free components, we have

$$\mathbf{ex}(n, P_k) = \max_{\sum n_i = n, n_i \geq 1} \sum \mathbf{ex}_{\text{con}}(n_i, P_k).$$

Similarly, a maximal $\mathcal{C}_{\geq k}$ -free graph is connected and every connected graph is a cactus-like union of 2-connected blocks (and edges) so we have

$$(5.5) \quad \mathbf{ex}(n, \mathcal{C}_{\geq k}) = \max_{\sum (n_i - 1) = n - 1, n_i \geq 2} \sum \mathbf{ex}_{2\text{-con}}(n_i, \mathcal{C}_{\geq k}),$$

where we define $\mathbf{ex}_{2\text{-con}}(2, \mathcal{C}_{\geq k}) = 1$.

Let G be a connected, n -vertex, P_k -free graph. Add a new vertex to it and join to all other vertices. We obtain G_{n+1} with $e(G_{n+1}) = e(G) + n$. This new graph has no cycle of length exceeding k and its connectivity is one larger than that of G_n . We obtain

$$(5.6) \quad \mathbf{ex}(n, P_k) + n \leq \mathbf{ex}(n + 1, \mathcal{C}_{\geq k+1})$$

and

$$(5.7) \quad \mathbf{ex}_{\text{con}}(n, P_k) + n \leq \mathbf{ex}_{2\text{-con}}(n + 1, \mathcal{C}_{\geq k+1}).$$

So Theorem 5.1 and (5.6) imply Theorem 5.5. Similarly, Theorem 5.2 and (5.6) imply Theorem 5.6.

The upper bounds for $\mathbf{ex}_{2\text{-con}}(n, \mathcal{C}_{\geq k+1})$ yield upper bounds for $\mathbf{ex}_{\text{con}}(n, P_k)$. (Actually, (5.7) and Theorem 5.4 lead to the solution of $\mathbf{ex}_{\text{con}}(n, P_k)$, Theorem 5.7).

Again Theorem 5.4 and (5.5) lead to Theorem 5.2 which is obviously stronger than Theorem 5.1.

Finally, the proof of Theorem 5.4 uses induction on n and k , by deleting small degree vertices, contracting edges, and finally applying Pósa's theorem on Hamiltonian graphs.

5.4. Generalizations

In a recent work Lidický, Hong Liu and Cory Palmer [174] determined the exact Turán number (and the unique extremal graph) when the forbidden graph L is a *linear forest*, each component is a path. They also considered *star-forests*.

Gyárfás, Rousseau, and Schelp [132] determined $\mathbf{ex}(K(m, n), P_k)$ for all m, n, k . Their formula and proof are rather involved, they distinguish 10 subcases.

6. EXCLUDING TREES

Here we shall discuss two extremal problems on trees: the Erdős–Sós conjecture and the Loeb–Komlós–Sós conjecture.

6.1. Erdős–Sós conjecture

We have already discussed the Erdős–Gallai theorems. Since the extremal numbers for P_k and for the star $K_{1, k-1}$ are roughly the same, this led Erdős and T. Sós to the following famous conjecture.

Conjecture 6.1 (Erdős–Sós [63]). *Let T_k be an arbitrarily fixed k -vertex tree. If a graph G_n contains no T_k , then*

$$(6.1) \quad e(G_n) \leq \frac{1}{2}(k-2)n.$$

As we have seen – by (5.2) – the disjoint union of complete graphs K_{k-1} shows that $\text{ex}(n, T_k) \geq \frac{1}{2}(k-2)n - \frac{1}{8}k^2$. Though several partial cases were settled, the upper bound was unknown until Ajtai, Komlós, Simonovits, and Szemerédi proved:

Theorem 6.2 (Main Theorem, Sharp [1, 2, 3]). *There exists an integer k_0 such that if $k > k_0$ and T_k is an arbitrarily fixed k -vertex tree, and the graph G_n contains no T_k , then*

$$(6.2) \quad e(G_n) \leq \frac{1}{2}(k-2)n.$$

Below we list a few subcases where this conjecture is verified, but we do not try to give a complete list.

Theorem 6.3 (Sidorenko [215]). *If T_k has a vertex x connected to at least $k/2$ vertices of degree 1 (i.e., leaves) then the Erdős–Sós conjecture holds for this T_k .*

Theorem 6.4 (McLennan [182]). *If the diameter of T_k is at most 4, then the Erdős–Sós conjecture holds for this T_k .*

Dobson (and coauthors) have several results in this area, under some strong condition of sparsity. We mention only the Brandt–Dobson theorem [34], or Sacle and Wozniak, [251], [211].

6.2. Sketch of the proof of Theorem 6.2

We are given a T_k , and a G_n violating (6.2). We wish to embed T_k into G_n ($T_k \hookrightarrow G_n$). The proof is very involved and will be given in three rather long papers. The following weakening plays a central role.

Theorem 6.5 (η -weakening [1]). *For any (small) constant $\eta > 0$ there exists a $k_0(\eta)$ such that for $n \geq k > k_0(\eta)$, if*

$$(6.3) \quad e(G_n) > \frac{1}{2}(k-2)n + \eta kn,$$

then each k -vertex tree T_k is contained in G_n .

(a) First, in [1] we prove this theorem. If, in addition, we assume that G_n is dense: for some $c > 0$, $k > cn$, then we can apply the Szemerédi Regularity Lemma [235]. The proof of this theorem follows basically the line which was later used to prove the Loeb Conjecture, by Ajtai, Komlós and

Szemerédi [4] and later by Yi Zhao [255]. Also it was used in the Komlós–Sós conjecture by Diana Piguet and Maya Stein [201], Cooley [54], Hladký and Piguet [144], in stronger and stronger form, and now the publication of that proof is almost finished by Hladký, Komlós, Simonovits, Stein, and Szemerédi [143].

(b) In the second part, [2] we prove several theorems asserting that under some very special conditions $T_k \subseteq G_n$. Some of these steps are “stability arguments”.

Analyzing the proof of Theorem 6.5, shows that either we can gain at some points, in some of the estimates ηkn edges, and therefore Theorem 6.5 (more precisely, its slightly modified proof) implies the sharp version, Theorem 6.2, or else G_n must have a very special structure: it contains a smaller copy of the conjectured extremal graphs: for some $m \approx k$,

(b₁) either it contains a G_m which is almost a K_m ;

(b₂) or a G_m which is almost a $K(m/2 - \varepsilon m, m + \delta m)$.

(c) In both cases, if many edges connect $G_n - G_m$ to G_m , then we can embed T_k into G_n , embedding a smaller part of T_k outside of G_m , a larger part in the dense G_m , concluding that $T_k \hookrightarrow G_n$.

(d) If, on the other hand, we have found such a “mini-almost-extremal” $G_m \subseteq G_n$, but $e(G_m, G_n - G_m)$ is “small”, then we prove that

$$e(G_n - G_m) > \frac{1}{2}(n - m)(k - 2).$$

Hence we may forget the larger G_n : replace it by the smaller $G_n - G_m$. (In other words, we can apply induction on n .)

(e) The real difficulty comes when we have sparse graphs: $e(G_n) = o(n^2)$. Then we partition $V(G_n)$ into three parts: \mathbb{C} contains the vertices of high degrees, \mathbb{B} contains a part of $V(G_n)$ not containing dense subgraphs, and therefore behaving in a pseudo-random way, and \mathbb{A} behaves very similarly to the graphs we have in the dense cases.

How do we handle the dense case? (i) Applying the Regularity Lemma to G_n , we get a so called Cluster Graph H_ν . If this cluster graph has an (almost)-1-factor, then we can relatively easily embed T_k into G_n , using the extra ηkn edges of (6.3).

(ii) Next we extend this case to a more general situation, when G_n contains a so called Generalized 1-factor. We can prove the η -weakening in this case as well.

(iii) If the Cluster Graph H_ν does not contain an almost-1-factor, then we apply the Gallai-Edmonds structure-theorem (on graphs without 1-factors) to H_ν . In this case we can either embed T_k into G_n directly, or

reduce this case to Case (ii) above. Case (iii) is a very important subcase, with 3-4 subsubcases (depending on, how do we count them). Some of them go back to Case (ii) and in some others we directly (pseudo-greedily) embed T_k into G_n .

6.3. Komlós–Sós conjecture on median degree

The Komlós–Sós conjecture was already formulated in Section 1.4. This is a generalization of the Loebbl conjecture:

Conjecture 6.6 (Loebl–Komlós–Sós Conjecture [79]). *If G_n has at least $n/2$ vertices of degree at least $k - 1$, then G_n contains all the k -vertex trees T_k .*

The authors of [143] plan to write up the sharp version as well, which asserts the following.

Theorem 6.7. *If k is sufficiently large, then the Loebbl–Komlós–Sós conjecture is true.*

Remarks 6.8. (a) The Loebbl conjecture originates from a paper of Erdős, Füredi, Loebbl, and T. Sós, on the discrepancy of trees [79].

(b) Pósa’s theorem on the existence of Hamiltonian cycles also is – in some sense – a theorem asserting that if G has many vertices of sufficiently high degree, then it is Hamiltonian. There were earlier cases, when Woodall [250], proved an Erdős–Gallai type theorem on cycles, using the condition that there are many vertices of high degree. Also, Erdős, Faudree, Schelp, and Simonovits – trying to prove some Ramsey type theorems, – found a similar statement [78], but not for all the trees, only for the paths, and they proved there an almost sharp theorem. Their sharp conjecture was later proved by Hao Li.

(c) There were many important steps to reach the theorem above. We should mention here Ajtai–Komlós–Szemerédi, [4], then Yi Zhao [255], next Piguet and Stein [201], [202], Cooley [54], Hladký and Piguet [144], and many others.

7. MORE COMPLEX EXCLUDED SUBGRAPHS

In this section we present three theorems, each leading to a reduction method to prove new results from old estimates. Still there is no general theory to determine the bipartite Turán numbers.

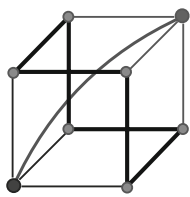
The three results we pick are the Erdős–Simonovits cube theorem (Theorem 7.1), $\mathbf{ex}(n, Q_8) = O(n^{8/5})$ which led to the Erdős–Simonovits reduction, the Faudree–Simonovits theorem (Theorem 7.8) concerning Theta graphs, a generalization of the Erdős–Bondy–Simonovits theorem, $\mathbf{ex}(n, C_{2k}) = O(n^{1+(1/k)})$, and Füredi’s theorem (Theorem 7.15) on two levels of the Boolean lattice which implies a general upper bound $\mathbf{ex}(n, L) = O(n^{2-(1/r)})$ for any graph L with vertices of degrees at most r on one side of L .

7.1. The Erdős–Simonovits Reduction and the Cube theorem

We have already mentioned Theorem 1.5, on the extremal number of the cube. Here we formulate a sharpening of it.

Theorem 7.1 ([90]). *Let Q_8 denote the graph determined by the 8 vertices and 12 edges of a cube, and Q_8^+ denote the graph obtained by joining two opposite vertices of this cube. Then*

$$\mathbf{ex}(n, Q_8) \leq \mathbf{ex}(n, Q_8^+) = O(n^{8/5}).$$



One reason why Erdős and Simonovits considered the extremal problem of the Cube graph was that this was one of Turán’s originally posed problems. The reason that Q_8^+ was also considered was that Erdős and Simonovits got it for free: their proof of Theorem 1.5 gave the same upper bound for Q_8^+ .

Let L be a bipartite graph with partite sets X and Y , and let $K_{t,t} * L$ denote the graph obtained by completely joining one partite set of $K_{t,t}$ to X and the other to Y .

Theorem 7.2 (Erdős and Simonovits Reduction Theorem [90]). *If L is a bipartite graph with $\mathbf{ex}(n, L) = O(n^{2-a})$, $a \leq 1$, and b is defined by $\frac{1}{b} = \frac{1}{a} + t$, then $\mathbf{ex}(n, K_{t,t} * L) = O(n^{2-b})$.*

The proof can go by induction on t and by counting the number of C_4 ’s.

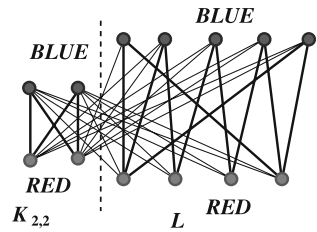
Let H be the graph obtained by deleting just three independent edges from $K_{4,4}$. Since $H = K_{1,1} * C_6$, Theorem 7.2 and $\mathbf{ex}(n, C_6) = O(n^{4/3})$ (Corollary 4.9) imply Theorem 7.1.

Since $\mathbf{ex}(n, L) = O(n)$ if L is a tree, so we have the following result:

Corollary 7.3. *For any tree L , $\mathbf{ex}(n, K_{t,t} * L) = O(n^{2-(1/(t+1))})$.*

This can be considered as a generalization of the Kővári–T. Sós–Turán theorem, since for $L = K_2$ we have $K_{t+1,t+1} = K_{t,t} * K_{1,1}$. Since $Q_8 - e$ is a subgraph of $K_{1,1} * P_6$, Corollary 7.3 implies

Corollary 7.4 (Erdős and Simonovits).
 $\text{ex}(n, Q_8 - e) = O(n^{3/2})$.



Further,

Theorem 7.5 (Erdős). *Delete an edge from $K_{a,a}$. For the resulting $L = K_{a,a} - e$ we have $\text{ex}(n, L) = O(n^{2-\frac{1}{a-1}})$.*

Indeed, for $a \geq 3$ the graph $K_{a,a} - e$ is a subgraph of $K_{a-2,a-2} * P_4$.

Since $K_{b,b} - K_{a,a}$ (for $b - 2 \geq a \geq 1$) can be written as $K_{b-a-1,b-a-1} * T$ where T is a double star, Corollary 7.3 also implies that $\text{ex}(n, K_{b,b} - K_{a,a}) = O(n^{2-(1/(b-a))})$. For this important case Füredi and West gave a sharper upper bound.

Theorem 7.6 ([121]). *For every $n \geq b > a$ we have*

$$\text{ex}(n, K_{b,b} - K_{a,a}) \leq \frac{1}{2}(b + a - 1)^{1/(b-a)} n^{2-(1/(b-a))} + \frac{1}{2}(b - a - 1)n.$$

In particular, it gives $\text{ex}(n, K_{3,3} - e) \leq \frac{1}{2}\sqrt{3}n^{3/2} + O(n)$. This was further improved by J. Shen [214] to

$$\text{ex}(n, K_{3,3} - e) \leq \frac{\sqrt{15}}{5}n^{3/2} + O(n).$$

He also showed that $\text{ex}(n, n, K_{3,3} - e) \leq (4/\sqrt{7})n^{3/2} + (n/2)$.

Pinchasi and Sharir extended the cube theorem, using a somewhat different proof:

Theorem 7.7 (Pinchasi and Sharir [204]). *A bipartite graph $G[A, B]$ with $|A| = m$ and $|B| = n$, not containing the cube Q has*

$$O(n^{4/5}m^{4/5} + mn^{1/2} + nm^{1/2})$$

edges.

Another, more explicit proof for Theorem 7.7 was presented in [115].

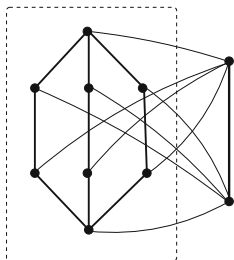
Historical remarks: (a) Erdős and Simonovits first proved the Cube theorem, using the Δ -almost-regularization.

(b) It seems that Theorem 7.2 covered all the cases known until that point.

(c) This (i.e. the Cube Recursion Theorem) was the first case, where one got an exponent, different from $2 - (1/a)$ and $1 + (1/a)$. Actually, Erdős thought earlier, that all the exponents must be of this form, (see [70]). This was disproved in their paper [90]: not by the cube, since there is no good lower bound for the cube: even $\mathbf{ex}(n, Q_8)/n^{3/2} \rightarrow \infty$ is not known. However, a more complicated example, for which the lower bound – using random graphs – was good enough, disproved Erdős’ conjecture. Actually, one thinks that each rational $\alpha \in (0, 1)$ is extremal exponent for some finite \mathcal{L}_α , see Conjecture 2.37.

To disprove the Erdős conjecture concerning the exponents are of the form $1 + (1/a)$ or $2 - (1/a)$ it is enough to notice that we have graphs H with

$$c_H n^{(8/5) - \varepsilon(H)} < \mathbf{ex}(n, H) = O(n^{8/5}), \quad \left(c_H > 0, \varepsilon(H) < \frac{1}{10} \right).$$



More generally, consider the graph $H(t, \ell)$ obtained by connecting a $\Theta(3, \ell)$ to $K(t, t)$, as described in Theorem 7.2. By the Theorem 2.26 (lower bound) and Theorem 7.2 and Theorem 7.8 (upper bound) we obtain

$$c_{\ell,t} n^{2 - \frac{2\ell+2t}{3\ell+t^2+2t(\ell+1)-1}} < \mathbf{ex}(n, H_{t,\ell}) \leq \tilde{c}_{\ell,t} n^{2 - \frac{2}{2t+3}}.$$

So all the numbers $2 - \frac{2}{2t+3}$ are points of accumulations of exponents, in this sense. Actually, applying this argument with $t = 1, \ell = 3$, we get a simple counterexample, with the upper bound $O(n^{8/5})$ and a lower bound $cn^{2-(8/17)}$ ($c > 0$).

(d) In [92], Erdős and Simonovits proved the Supersaturated graph theorem (see Section 11) corresponding to the cube, thus providing a second proof of the Cube Theorem, that needed “less regularization”.

7.2. Theta graphs and the Faudree–Simonovits reduction

There is an alternative proof for the Bondy–Simonovits Theorem in [99]. This proof enabled a generalization to Θ -graphs. Recall that $\Theta_{k,\ell}$ denotes the graph consisting of ℓ paths of length k with the same endpoints but no inner intersections. We have $v(\Theta_{k,\ell}) = 2 + (k - 1)\ell$ and $e(\Theta_{k,\ell}) = k\ell$.

Theorem 7.8 (Theta-graph, Faudree–Simonovits [99]). *For fixed k and $\ell \geq 2$ one has $\text{ex}(n, \Theta_{k,\ell}) = O(n^{1+(1/k)})$.*

This exponent is conjectured to be the best possible, see Conjecture 4.11.

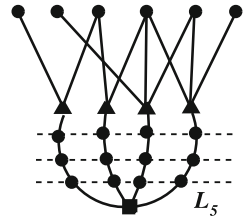
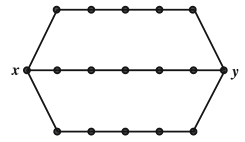
Applying the Erdős–Rényi Random Lower bound (Theorem 2.26) in its simpler form to $\Theta_{k,\ell}$ we get

$$\text{ex}(n, \Theta_{k,\ell}) > c_{k,\ell} n^{1+\frac{1}{k}-\frac{2}{k\ell}},$$

asymptotically matching the upper bound’s exponent.

The proof of Theorem 7.8 came from a “Recursion” theorem, asserting that if one knows good upper bounds for an L , and L^* is built from L in a simple way, then one has a good upper bound on $\text{ex}(n, L^*)$ as well.

Definition 7.9. Let L be a bipartite graph, with a fixed 2-colouring ψ in RED-BLUE with h RED vertices. Let $x \notin V(L)$ be a vertex from which h independent paths of $k - 1$ edges go the RED vertices of L , (these paths intersect only in x). Denote the obtained graph by $L_k(L, \psi)$.



Theorem 7.10 (Faudree–Simonovits Reduction, Trees [99]). *If L is a tree, then*

$$\text{ex}(n, L_k(L, \psi)) = O(n^{1+(1/k)}).$$

The Theta graph $\Theta_{k,\ell}$ is obtained from a star of ℓ edges. One has to be cautious with the next theorem, see Remark 7.12.

Theorem 7.11 (Faudree–Simonovits Reduction, General Case [99, 100]). *Let L be an arbitrary bipartite graph with a fixed coloring ψ and assume that*

$$(7.1) \quad \text{ex}^*(n, L) = O(n^{2-\alpha}).$$

Then for

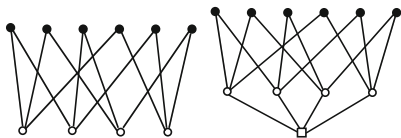
$$\beta = \frac{\alpha + \alpha^2 + \dots + \alpha^{k-2}}{1 + \alpha + \alpha^2 + \dots + \alpha^{k-2}}$$

we have

$$(7.2) \quad \mathbf{ex}(n, L_k(L, \psi)) \leq \mathbf{ex}^*(n, L_k(L, \psi)) = O(n^{2-\beta}).$$

Remark 7.12. Most probably, this recursion is never sharp but for trees. In its proof one has to apply standard arguments to subgraphs of $K(m, n)$ where $n \gg m$. We very seldom have matching lower and upper bounds in such cases.

7.3. A universal graph and dependent random choice



Erdős asked the following question: what are the extremal numbers for the two graphs on the left: The left one will be called M_{10} , the right one M_{11} and they are described as special cases of the following

Definition 7.13. Let k, r and t be given positive integers. $U(k, r, t)$ is obtained from the k vertices x_1, \dots, x_k by joining to each of the r -element subsets of $\{x_1, \dots, x_k\}$ t distinct vertices y_{i_1, \dots, i_r}^j . $U^+(k, r, t)$ is obtained from $U(k, r, t)$ by joining a new vertex w to all $x_h, h = 1, \dots, k$.

Problem 7.14 (Erdős). Determine (or estimate) $\mathbf{ex}(n, L)$ for $L := M_{10} = U(4, 2, 1)$ and $L := M_{11} = U^+(4, 2, 1)$.

One could ask for the motivation: why these graphs? Perhaps having obtained the cube theorem, we had good upper and lower bounds only in very special cases, when L contained some sample graphs – say a C_4 for which we have already provided sharp lower bounds. $U(4, 2, 1)$ clearly needed a new approach, and e.g. $U^+(k, 2, 1)$ contains many C_4 's but the earlier methods did not yield appropriate upper bounds. Füredi [107] answered this question proving that $\mathbf{ex}(n, U^+(k, 2, 1)) < k^{3/2}n^{3/2}$. More generally,

Theorem 7.15 (Füredi [107]). Let $U^+(k, r, t)$ be the universal bipartite graph from Definition 7.13. Then there exists a $c = c_r^{k,t} > 0$ such that

$$(7.3) \quad \mathbf{ex}(n, U^+(k, r, t)) < cn^{2-(1/r)}.$$

Concerning the Erdős question we have $\text{ex}(n, U^+(k, 2, 1)) < k^{3/2}n^{3/2}$ and more generally

$$(7.4) \quad \text{ex}(n, U^+(k, 2, t)) < n^{3/2} \cdot \sqrt{\frac{tk(k-1)^2 + 2(k-2)(k-1)}{8}} + n \frac{k-1}{4}.$$

Multiplying each vertex $(k-1)$ times in a C_4 -free graph we get a $U^+(k, 2, 1)$ -free graph which yields $\text{ex}(n, U^+(k, 2, 1)) \geq \Omega(k^{1/2}n^{3/2})$.

Erdős had the more general conjecture

Conjecture 7.16 (Erdős, [66], see also [92], [225]). *If every subgraph of the bipartite graph L has a vertex of degree at most r , then*

$$\text{ex}(n, L) = O(n^{2-(1/r)}).$$

The upper bound (7.3) for the universal graph immediately gives

Corollary 7.17. *If L is bipartite and has a 2-coloring where in the first color class all but one vertex is of degree at most r , then*

$$\text{ex}(n, L) = O(n^{2-(1/r)}).$$

Indeed, all such graphs can trivially be embedded into an appropriate $U^+(k, r, t)$.

Alon, Krivelevich, and Sudakov [9] gave a new probabilistic proof (for graphs where on one side all vertices are of degree at most r) with a better constant $c_r^{k,t}$. Their proof method became known as “dependent random choice”; for a survey see [101].

Lemma 7.18 (Dependent random choice, see, e.g., [101]). *Let k, t, r be positive integers. Let G_n be a graph with n vertices and average degree d , d be an integer. If there is a positive integer a such that*

$$\frac{d^a}{n^{a-1}} - \binom{n}{r} \left(\frac{t}{n}\right)^a \geq k,$$

then G_n contains a subset U of at least k vertices such that every r vertices in U have at least t common neighbors.

Note that in this lemma they do not claim that $U(k, r, t)$ is a subgraph. Nevertheless, using this lemma they improve the constant $c = c_r^{k,t}$ in (7.3) from $O((t+1)^{1/r}k^{2-(2/r)})$ to $c \leq 2^{-1+(2/t)}(t+1)^{1/r}k$.

As they mention at the end of [9], both proofs of Theorem 7.15 give a bit more (and thus Corollary 7.17 can be sharpened accordingly):

$$(7.5) \quad \mathbf{ex}(n, U^{+r}(k, r, t)) < cn^{2-(1/r)},$$

where the graph U^{+r} is obtained from $U^+(k, r, t)$ by replacing the vertex w in Definition 7.13 by an independent set of r vertices with the same neighbors, $x_1 \dots, x_k$.

However, the method of Dependent random choice gives more. Call a graph L_h on h vertices r -degenerate if it satisfies the condition of Conjecture 7.16. In other words, there is an ordering of its vertices x_1, \dots, x_h such that for every $1 \leq i \leq h$ the vertex x_i has at most r neighbors x_j with $j < i$.

Theorem 7.19 (Alon, Krivelevich, and Sudakov [9]). *If L is bipartite r -degenerate graph on h vertices, then for every $n \geq h$*

$$\mathbf{ex}(n, L) \leq h^{1/(2r)} n^{2-(1/4r)}.$$

Applying the above results with $r = 2$, $t = 1$ and $k = c\sqrt{n}$ to find a $U(k, 2, 1)$ one immediately obtains the following. Any graph on n vertices with $c_1 n^2$ edges contains a 1-subdivision of K_k with $k = c_2 \sqrt{n}$ for some positive c_2 depending on c_1 . This answers a question of Erdős [72]. The theorems of Bollobás and Thomason [29] and Komlós and Szemerédi [161] also imply the existence of such a large topological clique but their subgraph is not necessarily a 1-subdivision.

Given any graph L , let \bar{d} denote the $\max_{X \subseteq V(L)} \{2e_L(X)/|X|\}$, the maximum *local average degree*. Then L is $\lfloor \bar{d} \rfloor$ -degenerate. Hence the upper bound of Theorem 7.19 and the random method lower bound in (2.9) yield that

Corollary 7.20. *For every bipartite graph L ,*

$$(7.6) \quad \Omega(n^{2-c}) \leq \mathbf{ex}(n, L) \leq O(n^{2-(c/8)}),$$

where $c = 2/\bar{d}$, is the same as in (2.9).

8. EIGENVALUES AND EXTREMAL PROBLEMS

Let $A = A(G_n)$ be the adjacency matrix of G_n , and \mathbf{j} be the vector each entry of which is 1. Since

$$(8.1) \quad e(G_n) = \frac{1}{2} \mathbf{j} A \mathbf{j}^T$$

and, more generally,

$$(8.2) \quad w_k(G_n) = \frac{1}{2} \mathbf{j} A^k \mathbf{j}^T$$

counts the number of k -edge walks in G_n , therefore it is not so surprising that eigenvalues can be used in extremal graph problems. An easy to read source on spectra of graphs is Cvetkovič–Doob–Sachs [56].

Theorem 8.1 (Babai–Guiduli [12]). *Let $\Lambda(G) = \max |\lambda_i|$, where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $A(G_n)$. If $K_{a,b} \not\subseteq G_n$, (and $2 \leq a \leq b$) then*

$$(8.3) \quad \Lambda \leq \sqrt[a]{b-1} \cdot n^{1-(1/a)} + o(n^{1-(1/a)}).$$

Since trivially

$$(8.4) \quad 2e(G_n) \leq \Lambda n$$

the inequality (8.3) implies Theorem 2.22 apart from the $o()$ term.

Remark 8.2. For regular or almost regular graphs $\Lambda(G_n) \approx \frac{2e(G_n)}{n}$, and then the two estimates are basically equivalent. The constant in the above theorem is not sharp since – as we know from Theorem 3.19, – the constant can be improved.

We have already mentioned Nikiforov’s result (Theorem 3.21) on the Zarankiewicz problem. In fact, he proved [198] that for all $n \geq b \geq a \geq 2$ and a $K_{a,b}$ -free graph G_n we have

$$(8.5) \quad \Lambda(G_n) \leq (b-a+1)^{1/a} n^{1-(1/a)} + (a-1)n^{1-(2/a)} + (a-2).$$

This improves the coefficient in Theorem 8.1. It also implies Füredi’s bound (3.11) for the $\mathbf{ex}(n, K_{a,b})$ according to (8.4). For C_4 -free graphs he has $\Lambda^2 - \Lambda + 1 \leq n$.

Recall that $T_{n,k}$ denotes the Turán graph, the k -partite complete graph of maximum size. Given a K_{k+1} -free graph G_n Nikiforov [197] showed that $\Lambda(G_n) < \lambda(T_{n,k})$ unless $G = T_{n,k}$. For a recent reference of a generalization see Z. L. Nagy [196].

9. EXCLUDING TOPOLOGICAL SUBDIVISIONS

9.1. Large topological subgraphs

We have already mentioned that our classification does not hold for infinite families of excluded subgraphs. One important phenomenon is that $\mathbf{ex}(n, \mathcal{L})$ can be linear for infinite \mathcal{L} even if \mathcal{L} contains only cycles.²⁷ Here we consider a very central graph theoretical problem strongly connected to the 4-colour conjecture.

Definition 9.1. Given a graph H , its *subdivision* (or a topological H) is obtained from it by replacing each edge e of H by some paths P_e so that these paths do not have their inner (new) vertices in common.

Wagner asked if for any integer ℓ there exists a $k = k_\ell$ such that any G with chromatic number $\chi(G) > k_\ell$ must contain a topological subdivision of K_ℓ . This was proved by Gabor Dirac and H. Jung (independently). Answering a question of Dirac, Mader proved the following important result.

Theorem 9.2 (Mader, [183]). *If G_n is an n -vertex graph, and*

$$e(G_n) \geq n(\ell - 1)2^{\binom{\ell-1}{2}-1},$$

then G_n contains a subdivision of the complete ℓ -graph.

This statement is stronger than the original Wagner conjecture, since a graph with large chromatic number contains a subgraph with large minimum degree. Mader, and independently, Erdős and Hajnal conjectured that

Conjecture 9.3 (Mader, Erdős–Hajnal). *There exists a constant $c > 0$ such that if $e(G_n) > c\ell^2 n$, then G_n contains a topological K_ℓ .*

A slightly weaker form of this conjecture was proved by Komlós and Szemerédi, [160], then – by a different method – Bollobás and Thomason [29] proved this conjecture and, almost immediately after that, Komlós and Szemerédi [161] proved Conjecture 9.3 as well.

Theorem 9.4 (Bollobás–Thomason). *Every graph G_n of size at least $256\ell^2 n$ edges contains a topological complete subgraph of order ℓ .*

²⁷Here the simplest case is Theorems 5.1.

As to the small values of ℓ , Dirac conjectured that for $n \geq 3$ every G_n , with $e(G_n) \geq 3n - 5$ contains a topological K_5 . This improvement of the famous Kuratowski theorem was proved by Mader in [184] and the corresponding extremal graphs were characterized in [186]. The reader is recommended the excellent “featured review” of Carsten Thomassen on the paper of Mader [184], on the MathSciNet.

An excellent survey of Mader on this topic is [185].

9.2. Turán numbers of subdivided graphs

Let ε be a positive real, $0 < \varepsilon < 1$. Kostochka and Pyber [163] proved that every n -vertex graph G_n with at least $4^{t^2} n^{1+\varepsilon}$ edges contains a subdivision of K_t on at most $(7t^2 \ln t)/\varepsilon$ vertices, where $0 < \varepsilon < 1$. This (for $t = 5$) answers a question of Erdős about finding a non-planar subgraph of size $c(\varepsilon)$ in a graph with $n^{1+\varepsilon}$ edges.

Recently, T. Jiang [150] improved the Kostochka-Pyber upper bound to $O(t^2/\varepsilon)$. On the other hand, for each $0 < \varepsilon < 1$ and $n > n_0(\varepsilon)$ there are n -vertex graphs of girth at least $1/\varepsilon$ (see Corollary 2.30). In such a graph any subdivision of K_t must contain $\Omega(t^2/\varepsilon)$ vertices, so Jiang’s result is sharp.

Theorem 9.5 (Jiang and Seiver [151]). *Let L be a subdivision of another graph H . For each edge $xy \in E(H)$ let $\ell(x, y)$ denote the length of the path in L replacing the edge xy . Suppose that $\ell(x, y)$ is even for each edge of H , and let $\min\{\ell(x, y) : xy \in E(H)\} = 2m$. Then $\mathbf{ex}(n, L) = O(n^{1+(8/m)})$.*

The main tools in the proof are the Dependent Random Choice, Lemma 7.18, and the Erdős–Simonovits Δ -almost-regularization, Theorem 2.19.

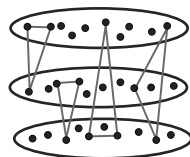
10. HYPERGRAPH EXTREMAL PROBLEMS

10.1. Positive Density problems

This is a short detour into Hypergraph Extremal Problems. Now our “Universe” is the class of r -uniform hypergraphs. Katona, Nemetz and Simonovits [154] showed (using a simple averaging) that

Theorem 10.1 (Katona, Nemetz and Simonovits [154]). *$\mathbf{ex}_r(n, \mathcal{L})/\binom{n}{r}$ is monotone decreasing, and therefore convergent.*

The hypergraph extremal problems are extremely hard. Even the simplest extension of Turán’s theorem is unsolved: Let $K_4^{(3)}$ be the three-uniform hypergraph with 4 vertices and 4 triples.



Construction 10.2 (Turán, the simplest case). We partition n vertices into three classes C_1, C_2, C_3 and we take all the triples of the form (x, y, z) , where

- (a) $x \in C_i, y \in C_i, z \in C_{i+1}$ (where the indices are taken mod 3);
- (b) the three vertices are in three different groups.

One can easily see that his construction contains no $K_4^{(3)}$.

Conjecture 10.3 (Turán). *Construction 10.2 is asymptotically extremal for $K_4^{(3)}$. (Perhaps it is extremal, not only asymptotically extremal, at least for $n > n_0$.)*

Here we cut it short and recommend the reader (among others) the survey of Füredi on hypergraph extremal problems [108], and also the papers of Füredi and Simonovits [120], Keevash and Sudakov [156], Füredi–Pikhurko–Simonovits [119], and the survey of Keevash [155].

10.2. Degenerate hypergraph problems

For r -uniform hypergraphs the r -partite graphs generalize the bipartite graphs. An important illustration of this is the one below, extending Theorem 2.31.

Theorem 10.4 (Degenerated hypergraph problems). *For an r -uniform extremal hypergraph problem of $\mathcal{L}^{(r)}$, $\text{ex}(n, \mathcal{L}^{(r)}) = o(n^r)$, if and only if there is an $L \in \mathcal{L}^{(r)}$ which can be r -vertex-colored so that each hyperedge of L gets r distinct colors.*

Theorem 10.4 is an easy corollary of the following theorem of Erdős, (which generalizes Theorem 2.22).

Theorem 10.5 (Erdős [65]). *Let $K^{(r)}(a_1, \dots, a_r)$ be the r -uniform hypergraph with r vertex-classes C_1, \dots, C_r , where $|C_i| = a_i$, and $a_1 = t$. Then*

$$\text{ex}^{(r)}(n, K^{(r)}(a_1, \dots, a_r)) = O(n^{r-(1/t^{r-1})}).$$

Extending some problems and results for ordinary graphs, Brown, Erdős and Sós started investigating the following

Problem 10.6 (Brown, Erdős, and Sós [39], [40]). Consider r -uniform hypergraphs for some fixed r , and denote by $\mathbb{H}_{k,\ell}^r$ the family of r -uniform k -vertex hypergraphs with ℓ hyperedges. Determine or estimate $f_r(n, k, \ell) := \mathbf{ex}(n, \mathbb{H}_{k,\ell}^r)$.

Brown, Erdős, and Sós proved many upper and lower bounds for special cases of Problem 10.6. We have already mentioned one of them: the $f_3(n, 6, 3)$ -problem.²⁸ It is easy to see that $f_3(n, 6, 3) < \frac{1}{6}n^2$. The real question was if $f_3(n, 6, 3) = o(n^2)$ or not. Ruzsa and Szemerédi [210] proved that the answer is YES. We formulated this in Theorem 1.9. This theorem became a very important one. We originate, among others, the “Removal Lemma” from here.

We shall return to this problem in the section on applications.

11. SUPERSATURATED GRAPHS

The theory of Supersaturated extremal problems is a very popular area today. Here we shall restrict ourselves to the supersaturated extremal graph problems related to bipartite excluded graphs, just mention a few further references, like Lovász and Simonovits [178], Razborov [179], Lovász [177], Reiher [205].

Given a graph G , denote by $N(G, F)$ the number of subgraphs of G isomorphic to F . Here we have to be slightly cautious: if F has non-trivial automorphisms, then we can count isomorphisms or isomorphic subgraphs, and the ratio of these two numbers equal to the automorphism number.

A theorem which asserts that a graph G_n contains very many graphs L from a family \mathcal{L} is called a **theorem on supersaturated graphs**. Such theorems are not only interesting in themselves, but also are often useful in establishing other extremal results. At this point it is worthwhile mentioning such a result for complete bipartite graphs, obtained by Erdős and Simonovits [94]:

Theorem 11.1 (Number of complete bipartite graphs). *For any integers a and b there exists a constant $c_{a,b} > 0$ such that if G_n is a graph with e edges, then G_n contains at least $\lceil c_{a,b} e^{ab} / n^{2ab-a-b} \rceil$ copies of $K_{a,b}$.*

Corollary 11.2. *Let $c > 0$. If $e(G_n) = e > (1+c)\mathbf{ex}(n, C_4)$, then G_n contains at least $\gamma e^4 / n^4$ copies of C_4 , for some $\gamma(c) > 0$. The random graph with e edges shows that this is sharp.*

²⁸If $r = 3$, then we delete the subscript in f_3 .

Proof of the Cube Theorem (Sketch). Apply Theorem 2.19 obtaining a Δ -almost-regular (bipartite) $\tilde{G}_n \subseteq G_n$. Apply the corollary to this \tilde{G}_n . It contains $\gamma \frac{e^4}{n^4}$ C_4 's. On the average, an edge of G_n is contained in $\gamma e^3/n^4$ copies of C_4 . Take a typical edge xy : the bipartite graph $G[U, V]$ spanned by the neighbors $U := N(x)$ and $V := N(y)$ – by $\mathbf{ex}(m, C_6) = O(m^{4/3})$, – will contain a C_6 . Now, xy and this C_6 will provide a Q_8^+ : a cube with a diagonal. ■

Basically the same argument proves Theorem 7.2.

11.1. Erdős–Simonovits–Sidorenko conjecture

In this part $\chi(L) = 2$. Erdős and Simonovits [225] formulated three conjectures and also that the main idea behind these conjectures is that the number of copies of subgraphs $L \in \mathcal{G}_n$ is minimized by the random graph if $E = e(G_n)$ is fixed and is not too small.

To formulate these conjectures, first we calculate the “expected number of copies” of $L \subseteq R_n$ if R_n is a random graph with edge probability $p = E/\binom{n}{2}$. Let $v = v(L)$, and $e = e(L)$. Clearly, if the edges are selected independently, with probability p , then R_n contains $\binom{n}{v}$ possible v -tuples, each containing the same number a_L of copies of L , and therefore

$$(11.1) \quad \mathbb{E}(\#(L \subseteq R_n)) = (a_L + o(1)) \frac{n^v}{v!} p^e = (a_L + o(1)) \frac{n^v}{v!} \left(\frac{2E}{n^2} \right)^e = a_L^* \frac{E^e}{n^{2e-v}}$$

Conjecture 11.3 (Erdős–Simonovits, [225]). *There are two constants, $\Omega = \Omega_L > 0$ and $c = c_L > 0$ such that if $E > \Omega \cdot \mathbf{ex}(n, L)$, then any graph G_n with E edges contains at least*

$$c_L \frac{E^e}{n^{2e-v}}$$

copies of L .

This was the weakest form. The strongest form of this conjecture was

Conjecture 11.4 (Erdős and Simonovits). *For every $\varepsilon > 0$, if $E > (1 + \varepsilon) \cdot \mathbf{ex}(n, L)$, then any graph G_n with E edges contains at least $(1 + \varepsilon) \mathbb{E}\mathbb{R}(n, L, E)$ copies of L , if $n > n_0(\varepsilon)$, where $\mathbb{E}\mathbb{R}(n, L, E)$ denotes the expected number of edges of a random Erdős–Rényi graph with n vertices and E edges.*

Obviously, one has to assume that $e(G_n) > \mathbf{ex}(n, L)$.

Remark 11.5 (Relation to Sidorenko's Conjecture). At first sight Sidorenko's conjecture [216] seems to be sharper than the above one. This is not the case. In fact, Sidorenko's Conjecture applies only to dense host graphs. There, as Sidorenko points out in his papers, the two versions are equivalent.

Also, it is obvious that there is not much difference if we consider above the hypergeometric model of random graphs, where the number of vertices and edges are given, or if we fix only n but the edges are taken independently, and therefore $e(G_n)$ follows a binomial distribution.

Sidorenko, working on applications of extremal graph theorem to probability distribution translated the above conjecture to integrals and arrived at a conjecture [216], where the error terms disappeared. The meaning of his version was that if one considers dense graphs and defines $L \subseteq R$ for the case when G is a function, generalizing the notion of graphs, then the Random Continuous graph will have the least number of copies of L , more precisely, that will minimize the corresponding integral.

We skip the formulation of this problem, just refer to some papers of Lovász, and Hatami [140], and to the book of Lovász [177].

Jagger, Šťovíček, and Thomason [148] investigated the following problem originating from a conjecture of Erdős, disproved by Thomason.

Problem 11.6. Given a sample graph L , denote by $\rho_L(G_n)$ the sum of copies of L in G_n and in its complementary graph. What is the minimum $\Gamma_n(L)$ of this, taken over all n -vertex graphs?

Erdős conjectured that the random graph yields the minimum, for K_4 . This was disproved by Thomason [237]. Investigating the case of general L , Jagger, Šťovíček, and Thomason proved some interesting results in connection with Sidorenko's conjecture.

Here we should emphasize that there is a slight difference between looking for copies of an L in G_n or for copies of homomorphic images: In the second case we allow vertices to map into G_n with some coincidences.

As to the Sidorenko Conjecture, the first unknown case (as Sidorenko mentions) is when we delete the edges of a Hamiltonian cycle from $K_{5,5}$.

Theorem 11.7 (Conlon, Fox, Sudakov [52] [53]). *The Sidorenko Conjecture holds if $L = L[A, B]$ is a bipartite graph with a vertex $x \in A$ completely joined to B .*

Remark 11.8. Unfortunately, we do not have sufficiently good lower bounds for the extremal problem of the cube. The Erdős–Simonovits Conjecture was proved for Q_8 in [92].

Hatami proved the Sidorenko conjecture for any cube (i.e. of any dimension), yet, that was not really enough to provide a reasonable upper bound for the 4-dimensional cube. This reflects some difference between extremal problems and the corresponding Supersaturated Graph Problems (at least, for dense host graphs).

12. ORDERED STRUCTURES

12.1. Directed graphs, ordered graphs

There is an extensive literature on Digraph extremal problems, see e.g., the survey of Brown and Simonovits [42], or [41]. We skip here the general theory.

Denote $\overrightarrow{\text{ext}}(n, \vec{L})$ the maximum number of edges in a directed graph not containing the oriented subgraph \vec{L} . For every \vec{L} containing a directed path of length 2 one has $\overrightarrow{\text{ext}}(\vec{L}) \geq \lfloor n^2/4 \rfloor$. Indeed, orient the edges of $K_{n/2, n/2}$ simultaneously into one direction. For bipartite \vec{L} it is more interesting to consider the minimum outdegree.

Consider the following directed graph $\vec{L}_{1,a,b}$ on $1+a+b$ vertices $w, x_1, \dots, x_a, y_1, \dots, y_b$. The oriented edges are w to x_i and x_i to y_j ($1 \leq i \leq a, 1 \leq j \leq b$).

Theorem 12.1 (Erdős, Harcos and Pach [82]). *Given integers a and b , there exists a $c = c_{a,b} > 0$ such that the following holds. Any oriented graph with minimum out-degree $\delta^+ \geq cn^{1-(1/a)}$ contains a copy of $\vec{L}_{1,a,b}$.*

This result opened up a new interesting field with many open problems.

Another ordered Turán function was defined by Timmons [239]. He showed that if a graph with vertex set $\{1, 2, \dots, n\}$ has at least

$$(1 + o(1))(2/3)n^{3/2}$$

edges, then it contains a C_4 with vertices $a_1b_1a_2b_2$ such that $a_1, a_2 < b_1, b_2$. He extended other ordinary Turán problems to these zig-zag type questions. Many problems remain unsolved.

12.2. Erdős–Moser conjecture on unit distances

Erdős and Leo Moser [83] conjectured that

Conjecture 12.2. *If n points of the plane are in convex position, then the number of unit distances among them is $O(n)$.*

Füredi proved a slightly weaker result:

Theorem 12.3 (Füredi [106]). *If n points are in convex position in the plane, then there are at most $O(n \log n)$ unit distances among them.*

To prove this, Füredi directly formulated the excluded Ordered Matrix Property and solved a matrix-containment problem. The crucial point of his proof was Theorem 12.6 below.

The best known lower bound in the Erdős–Moser problem, $2n - 7$, is due to Edelsbrunner and P. Hajnal [58].

12.3. Ordered submatrices

The ordered matrix problems partly came from geometric problems (see Bienstock and Györi, [23], Füredi [106]), but they are interesting on their own, too. A geometric application, called Erdős–Moser conjecture, is discussed above in Subsection 12.2.

We have already indicated that most extremal graph problems have matrix forms, too: To determine $\text{ex}^*(m, n, L)$ we considered all $m \times n$ 0-1 matrices not containing any permutation of the bipartite adjacency matrix of L .

In the ordered case here we exclude only those submatrices where the indexing of the rows and columns of \mathbf{M} is fixed. This way we exclude fewer subconfigurations.

Definition 12.4 (Matrix containment). Let \mathbf{M} and \mathbf{P} be two 0-1 matrices. We say that \mathbf{M} contains \mathbf{P} if we can delete some rows and columns of \mathbf{M} and then perhaps switch some 1's into 0 so that the resulting matrix be \mathbf{P} . Otherwise we say that \mathbf{M} avoids \mathbf{P} .

So, we can delete rows and columns of \mathbf{M} but can not permute them. Now we can define the Matrix Extremal Problems:

Problem 12.5 (Ordered Matrix Problem). Given an $a \times b$ 0-1 (sample) matrix \mathbf{P} , and a (huge) $m \times n$ 0-1 matrix \mathbf{M} , how many 1's can occur in \mathbf{M} under the condition that \mathbf{M} does not contain \mathbf{P} in the “ordered” way. Denote by $\text{ext}_{\text{mat}}(m, n, \mathbf{P})$ the maximum.

One of the first nontrivial results was

Theorem 12.6 (Füredi [106]). *Let*

$$\mathbf{P} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

If the $n \times n$ 0-1 matrix \mathbf{M} does not contain \mathbf{P} , then it has at most $O(n \log n)$ 1's. In fact, $\text{ext}_{\text{mat}}(n, n, \mathbf{P}) = \Theta(n \log n)$.

Tardos [236] proved that $\text{ext}_{\text{mat}}(n, \mathbf{P}) = n \log_2 n + O(n)$.

Completing earlier works of Füredi and Péter Hajnal [116] Tardos [236] classified the ordered matrix Turán numbers for all small submatrices. The extremely slow growing inverse Ackermann function is denoted by $\alpha(n)$.

Theorem 12.7 ([116], [236]). *If \mathbf{P} is a 0-1 matrix with at most four 1's, then*

$$\text{ext}_{\text{mat}}(n, n, \mathbf{P}) = \begin{cases} 0 & \text{or} \\ \Theta(n), & \text{or} \\ \Theta(n\alpha(n)), & \text{or} \\ \Theta(n \log n), & \text{or} \\ \Theta(n^{3/2}). \end{cases}$$

12.4. Ordered matrices and the Stanley–Wilf conjecture on sub-permutations

Trying to prove the Erdős–Moser Conjecture, Füredi and Péter Hajnal [116] arrived at the following conjecture, proved by Marcus and Tardos.

Theorem 12.8 (Füredi–Hajnal conj. [116]/Marcus–Tardos theorem [187]). *For all permutation matrices \mathbf{P} we have $\text{ext}_{\text{mat}}(n, n, \mathbf{P}) = O(n)$.*

This time there was a famous Stanley–Wilf conjecture around, on the number of permutations “avoiding” a fixed permutation. To formulate it, we need to define the Permutation containment:

Definition 12.9 (Permutation containment). We say that a permutation $\sigma : [1, n] \rightarrow [1, n]$ contains a permutation $\pi : [1, k] \rightarrow [1, k]$, if there exist $1 \leq x_1 < x_2 < \dots < x_k \leq n$ for which

$$\sigma(x_i) < \sigma(x_j) \quad \text{if and only if} \quad \pi(i) < \pi(j).$$

The famous Stanley–Wilf conjecture²⁹ states that

²⁹Marcus and Tardos [187] write that it is difficult to locate the corresponding reference.

Conjecture 12.10 (Stanley–Wilf). *For any permutation pattern q , if $S_n(q)$ is the number of permutations of length n avoiding the pattern q , then there is a constant c_q so that $S_n(q) \leq c_q^n$.*

Theorem 12.11 (Klazar [158]). *The Füredi–Hajnal conjecture implies the Stanley–Wilf conjecture.*

So Marcus and G. Tardos settled this conjecture as well.

Remark 12.12. The permutation containment is just a subcase of the more general question. In some other cases there are definite differences between ordinary Turán type extremal problems and the ordered matrix problems. For a special matrix, where the corresponding graph is a tree, hence it has linear Turán function, our threshold function turns out to be $\Theta(n \log n)$.

13. APPLICATIONS IN GEOMETRY

13.1. Applicability of the Kővári–T. Sós–Turán bound

We have mentioned that Theorem 2.22 is applicable in several cases. Here we mention only two.

(A) The Unit Distance Graph of the Plane contains no $K(2, 3)$. Erdős used this to estimate the number of unit distances by $O(n^{3/2})$.

(B) G. Megyesi and Endre Szabó³⁰ answered a question of F. E. P. Hirzebruch using this theorem.

Assume that we are given k smooth curves in the the Complex Projective Plane and assume that their union has only nodes and tacnodes³¹ as singularities. Let $t(k)$ denote the maximum number of tacnodes in such cases. Hirzebruch proved that $t(k) \leq \frac{4}{9}k^2 + \frac{4}{4}k$. Hirzebruch asked if $\limsup t(k)/k > 0$.

Theorem 13.1 (G. Megyesi, and E. Szabó [191]). *There exist three positive constants, A , B and C for which*

$$Ak^{1+(B/\log \log k)} \leq t(k) \leq Ck^{2-(1/7633)}.$$

³⁰We use the longer versions of the names whenever we see chances to mix up authors of similar names.

³¹Tacnode means roughly that the curve is touching itself.

13.2. Unit Distances

Erdős was interested in the following problem:

Problem 13.2 (Unit distances). Given an n -element set in the d -dimensional Euclidean space \mathbb{E}^d , how many of the distances can be the same, say equal to 1?

Conjecture 13.3 (Unit distances). For any $\varepsilon > 0$, there is an n_0 such that if $n > n_0$ and given an n -element set in the plane \mathbb{E}^2 , then the number of unit distances is at most $n^{1+\varepsilon}$.

The motivation of this conjecture is – as Erdős observed – that if we arrange the $n = k \times k$ points into a $k \times k$ grid, and rescale this grid so that the “most popular” distance be 1, then this distance will occur at most $n^{1+\varepsilon}$ times, (actually, approximately $n^{1+(c/(\log \log n))}$ times). So Erdős conjectured that the number of unit distances in the plane has an upper bound of roughly this form.

The first upper bound was a trivial application of Theorem 2.22:

Theorem 13.4 (Unit distances, Erdős 1946). Given n points in the plane, the number of unit distances among them is at most

$$\text{ex}(n, K_{2,3}) < \left(\frac{1}{\sqrt{2}} + o(1) \right) n^{3/2}.$$

In \mathbb{E}^3 the number of unit distances is at most

$$(13.1) \quad \text{ex}(n, K_{3,3}) < c_{3,3} n^{5/3}.$$

Proof. Since two circles intersect in at most 2 points, the Unit Distance Graph of \mathbb{E}^2 contains no $K_{2,3}$. This implies the first inequality. Since 3 unit balls intersect in at most 2 points, the Unit Distance Graph of \mathbb{E}^3 does not contain any $K_{3,3}$. This implies (13.1). ■

Remarks 13.5. (a) Everything is different for the higher dimensions: \mathbb{E}^4 contains two orthogonal circles of radii $\frac{1}{\sqrt{2}}$, and these form a $K(\infty, \infty)$ in the corresponding unit graphs of \mathbb{E}^d , for $d \geq 4$. (This is the so called Lenz Construction.) (See also Section 3.6.)

(b) How sharp is this application? As the reader can see, it is very far from the conjectured upper bound. However, just to improve it to $o(n^{3/2})$ is non-trivial (Józsa-Szemerédi [152]). Actually, for the plane an $O(n^{4/3})$

upper bound was proved by Beck and Spencer [20] and Spencer, Szemerédi and Trotter [233], which is sharp if we do not insist on Euclidean metric, only on “normed spaces”. For this see the results of Peter Brass [35] and Pavel Valtr [245].

13.3. Cells in line arrangements

Let $\mathcal{I}(m, n)$ denote the maximum number of edges in m distinct cells determined by an arrangement of n lines in the plane. Canham [45] showed that for an absolute constant $c > 0$

$$(13.2) \quad \mathcal{I}(m, n) < c(m\sqrt{n} + n).$$

Indeed, if we construct a bipartite graph where one side of the vertex set consists of the m cells (or any other family of m convex sets with disjoint interiors), the other side of the vertex set consists of the n (tangent) lines and two vertices are joined if the corresponding geometric objects are incident, then it is easy to see that this graph does not contain a $K_{5,2}$. ■

This was a first nontrivial step toward the determination of the exact order of the magnitude of $\mathcal{I}(m, n)$ by Clarkson, Edelsbrunner, Guibas, Sharir, and Welzl [49]; it is $\Theta(n^{2/3}m^{2/3} + n)$. More about this and other geometric applications see the monograph of Pach and Agarwal [199].

14. FURTHER CONNECTIONS AND PROBLEMS

14.1. Connections of hypergraphs and critical graphs

We discussed Degenerate Hypergraph Extremal Problems in Section 10. Here we continue that line.

Excluding the 3-uniform hypergraph cones. Many of the other results, problems of [39] were also degenerate ones. One of them was where \mathcal{T} is the family of triangulations of the 3-dimensional sphere. This problem gave the name to this paper [39]. The crucial point was excluding the double cones:

Definition 14.1 (r -cones). The vertices of the 3-uniform hypergraph $Q_{r,t}$ are x_1, x_2, \dots, x_r , and y_1, y_2, \dots, y_t for some t , and the hyperedges are $x_i y_j y_{j+1}$, for all the possible i, j , where $y_{t+1} = y_1$. Further, $Q_r := \{Q_{r,t} : t = 3, 4, 5, \dots\}$.

Theorem 14.2 (Brown, Erdős, Sós, $r = 2$, [39], Simonovits [219] $r \geq 2$).

$$\mathbf{ex}(n, \mathbb{Q}_r) := O(n^{3-(1/r)}).$$

For $r = 2, 3$ there are matching lower bounds here. Actually, for $r = 2$ Brown, Erdős and Sós gave a construction, where not only the double-cone was excluded, but all the triangulations of the sphere. In Simonovits' lower bound only the double cone was considered.

In [220] Simonovits returned to this question and – using the main idea of Brown's construction [36] – he proved

Theorem 14.3 (Simonovits [220]). *There are (finite geometric) 3-uniform hypergraphs without triple-cones (i.e. without hypergraphs from \mathbb{Q}_3) and still having at least $cn^{3-(1/3)}$ triples.*

We saw that for the family of triangulations of the sphere, and for the family of Double Cones the extremal number is $O(n^{3-(1/2)})$ [39], (see [219]).

Brown, Erdős and T. Sós arrived at their question (most probably) since they wanted to generalize certain results from ordinary graphs to hypergraphs. Simonovits came from a completely different direction: he used this to disprove a conjecture of Gallai on independent vertices in 4-colour-critical graphs.

G is colour-edge-critical, if deleting any edge of G , we get a $(\chi(G) - 1)$ -chromatic graph. The 3-colour-critical graphs are the odd cycles, so the problem of critical graphs becomes interesting for the 4-chromatic case. Here we shall restrict ourselves to this case and suggest the reader to read Bjarne Toft's results on this topic in general.

Erdős asked if a 4-colour-critical graph can have cn^2 edges and Bjarne Toft constructed such a 4-chromatic graph [240] of $\approx \frac{n^2}{16}$ edges. This and some related questions can also be found in Lovász' book: Combinatorial Exercises [176].

Gallai had many beautiful conjectures on 4-colour-critical graphs. One of them, however, was “completely demolished”. He conjectured that if G_n is 4-colour-critical, then $\alpha(G_n) \leq n/2$. $G_{4m+2} = C_{2m+1} \otimes C_{2m+1}$ is 6-critical, with $d_{\min}(G_{4m+2}) = 2m + 3$. Simonovits – “blowing up” the vertices in one of the two odd cycles, – proved that there are 6-critical graphs G_n with $\alpha(G_n) = n - o(n)$.

It turned out that slightly earlier Brown and Moon [38] already disproved Gallai's conjecture for the 4-chromatic case, with a “clever but simple” construction.

Theorem 14.4 (Brown and Moon [38]). *There exist 4-chromatic edge-critical graphs G_n with $\alpha(G_n) > n - c\sqrt{n}$, for some constant $c > 0$.*

Next, Bjarne Toft came up with his construction, mentioned above. Using this and a hypergraph extremal theorem, Simonovits proved

Theorem 14.5. *There exists a constant $c_2 > 0$ such that if G_n is 4-colour-critical, then $\alpha(G_n) \leq n - c_2n^{2/5}$.*

This was obtained as follows: Simonovits reduced the original problem to estimating the number of independent vertices of degree 3 in a 4-colour-critical graph. The neighborhoods of these vertices generated a 3-uniform hypergraph $\mathcal{H}_m^{(3)}$ on the remaining vertices. Simonovits – using the Sperner Lemma from Topology proved that if I is a set of independent vertices of degree 3, in $V(G_n)$, then for $m := n - |I|$, $|I| < \mathbf{ex}_3(m, \mathbb{Q}_2) = O(m^{5/2})$, see [219]. He observed that $\mathcal{H}_m^{(3)}$ cannot contain double cones. This proved that $|I| < n - cn^{2/5}$. ■

(b) Lovász observed that instead of excluding the graphs from \mathbb{Q}_2 one can exclude a larger family, $\tilde{\mathbb{Q}}$: those 3-uniform hypergraphs which obey the conclusion of Sperner’s lemma [175]: each pair (x, y) is contained in an even number of hyperedges. This enabled him to completely settle *this* Gallai problem on colour-critical graphs. He proved that $\mathbf{ex}(n, \tilde{\mathbb{Q}}) \leq \binom{n}{2}$. So he obtained $|I| < n - c\sqrt{n}$, in Gallai’s problem. Besides proving and using a more applicable extremal graph theorem he also generalized the Brown–Moon construction.

(c) It was an interesting feature of Lovász’ solution that to get an upper bound on $\mathbf{ex}(n, \tilde{\mathbb{Q}})$ he used linear algebra.

We finish this part by sketching the proof of Lovász on the upper bound.

Theorem 14.6 (Lovász [175]). *Let $\mathbb{E}^{(3)}$ denote the family of 3-uniform hypergraphs H in which each pair of vertices is contained in an even number of triplets (i.e. hyperedges). Then $\mathbf{ex}(n, \mathbb{E}^{(3)}) \leq \binom{n}{2}$.*

Proof (Sketch). Assume that $H_n^{(3)}$ contains no subgraphs from $\mathbb{E}^{(3)}$. Consider that vectorspace over $GF(2)$ of dimension $\binom{n}{2}$ where the coordinates are indexed by pairs from $1, \dots, n$. Represent each triple of $H_n^{(3)}$ by such a vector, where we have 1 in those coordinates which are pairs from our triple. The condition that $H_n^{(3)}$ contains no subgraphs from $\mathbb{E}^{(3)}$ translates into the fact, that these vectors are linearly independent. Hence their number is at most the dimension of the vector-space. ■

Now, repeating the original argument of Simonovits, Lovász obtained

Theorem 14.7. *There exists a constant $c_3 > 0$ such that if G_n is 4-colour-critical, then $\alpha(G_n) \leq n - c_3 n^{1/2}$,*

This with the Brown-Moon construction completely settles Gallai's original problem, providing a matching lower bound. Lovász proved a more general theorem, and extended the Brown-Moon construction as well. We close this part with a beautiful conjecture of Erdős:

Problem 14.8. Is it true that if (G_n) is a sequence of 4-colour-critical graphs, then $d_{\min}(G_n) = o(n)$?

(Simonovits [219] and Toft [240] succeeded in constructing 4-color-critical graphs with minimum degrees around $c\sqrt[3]{n}$.)

Further sources to read: Several related results can be found in Lovász [176].

14.2. A multiplicative Sidon problem and C_{2k} -free graphs

As it was explained in Subsection 1.5, the Erdős problem about $\mathbf{ex}(n, C_4)$ in [60] was obtained from a multiplicative Sidon type question. He investigated subsets of integers of $A \subset \{1, 2, \dots, n\}$ with the property that for any four members of A the pairwise products are distinct, $a_i a_j \neq a_k a_\ell$.

A. Sárközy, P. Erdős, and V. T. Sós [88] started investigating the more general problem.

Problem 14.9. Fix an integer k . How many integers can we take from $[1, n]$ if the product of no k of them is a square.

Interestingly, this Problem also lead to Turán type questions, namely to $\mathbf{ex}(m, n, C_{2k})$ with $m \gg n$. Their conjecture (Conjecture 4.32 above) was proved by Győri [134], see Theorem 4.33. We shall not go into the number theoretic details; just refer the reader again to [134].

14.3. Cycle-free subgraphs of the d -dimensional hypercube

The d -dimensional hypercube, Q^d , is the graph whose vertex set is $\{0, 1\}^d$ and whose edge set is the set of pairs that differ in exactly one coordinate, $e(Q^d) = d2^{d-1}$. Let $\gamma(C_\ell) = \lim_{d \rightarrow \infty} \mathbf{ex}(Q^d, C_\ell)/e(Q^d)$. Note that $\gamma(C_\ell)$ exists, because $\mathbf{ex}(Q^d, C_\ell)/e(Q^d)$ is a non-increasing and bounded function of d . Considering the edges between the levels $2i$ to $2i + 1$ one can see that $\mathbf{ex}(Q^d, C_4) \geq (1/2)e(Q^d)$. The following conjecture is still open.

Conjecture 14.10 (Erdős [74]). $\text{ex}(Q^d, C_4) = (\frac{1}{2} + o(1)) e(Q^d)$.

The best upper bound $\gamma(C_4) \leq 0.6226$ was obtained by Thomason and Wagner [238], slightly improving the result of Chung [46].

Erdős [74] also asked whether $\text{ex}(Q^d, C_{2k})$ is $o(d)2^d$ for $k > 2$. This was answered negatively for C_6 by Chung [46], showing that $\gamma(C_6) \geq 1/4$. The best known results for C_6 are $1/3 \leq \gamma(C_6) < 0.3941$ due to Conder [50] and Lu [180], respectively.

On the other hand, for every $t \geq 2$ the inequalities

$$(14.1) \quad \text{ex}(Q^d, C_{4t}) \leq O(d^{\frac{1}{2} - \frac{1}{2t}} 2^d) \quad \text{and} \quad \text{ex}(Q^d, C_{4t+6}) = O(d^{\frac{15}{16} - \frac{1}{16t}} 2^d)$$

were proved by Chung [46] and Füredi and Özkahya [118], respectively. Hence $\gamma(C_{2k}) = 0$, except $\gamma(C_4) \geq 1/2$, $\gamma(C_6) \geq 1/3$ and the problem of deciding whether $\gamma(C_{10}) = 0$ is still open.

Conlon [51] generalized (14.1) by showing $\text{ex}(Q^d, H) = o(e(Q^d))$ for all H that admit a k -partite representation, also satisfied by each $H = C_{2k}$ except for $k \in \{2, 3, 5\}$.

14.4. Two problems of Erdős

Of course, we should close with two open problem of Erdős. The first one is the general version of that problem which was solved in [124] and [141], see Section 1.1.

Conjecture 14.11 (Erdős [75]). *Suppose that G is a graph on $(2k + 1)n$ vertices and of odd girth $2k + 1$. Then G contains at most n^{2k+1} induced cycles of length $2k + 1$.*

The next conjecture is also very famous and is motivated by the blown up pentagon (if we restrict it to $k = 2$.)

Conjecture 14.12 (Erdős [75]). *Suppose that G is a graph on $(2k + 1)n$ vertices and of odd girth at least $2k + 1$. Then G can be made bipartite by omitting at most n^2 edges.*

For the best known results here, for $k = 1$, see Erdős, Faudree, Pach, and Spencer [77] and Erdős, Györi, and Simonovits [81].

Acknowledgements. The authors are greatly indebted for fruitful discussions and helps to a great number of colleagues, among others to R. Faudree, E. Györi, and Z. Nagy.

REFERENCES

- [1] M. Ajtai, J. Komlós, M. Simonovits, and E. Szemerédi: On the approximative solution of the Erdős–Sós conjecture on trees, (manuscript).
- [2] M. Ajtai, J. Komlós, M. Simonovits, and E. Szemerédi: Some elementary lemmas on the Erdős–T. Sós conjecture for trees, (manuscript).
- [3] M. Ajtai, J. Komlós, M. Simonovits, and E. Szemerédi: The solution of the Erdős–Sós conjecture for large trees, (manuscript, in preparation).
- [4] M. Ajtai, J. Komlós, and E. Szemerédi: On a conjecture of Loeb, in *Graph theory, Combinatorics, and Algorithms*, Vol. 1, 2 (Kalamazoo, MI, 1992), Wiley-Intersci. Publ., pp. 1135–1146. Wiley, New York, 1995.
- [5] P. Allen, P. Keevash, B. Sudakov, and J. Verstraëte: Turán numbers of bipartite graphs plus an odd cycle, submitted.
- [6] N. Alon: Eigenvalues and expanders, *Combinatorica* 6 (1983), 83–96.
- [7] N. Alon: Tools from higher algebra, in: "Handbook of Combinatorics", R. L. Graham, M. Grötschel and L. Lovász, eds, North Holland (1995), Chapter 32, pp. 1749–1783.
- [8] N. Alon, S. Hoory, and N. Linial: The Moore bound for irregular graphs, *Graphs Combin.* 18 (2002), no. 1, 53–57.
- [9] N. Alon, M. Krivelevich, and B. Sudakov: Turán numbers of bipartite graphs and related Ramsey-type questions, *Combin. Probab. Comput.* 12 (2003), no. 5–6, 477–494.
- [10] N. Alon and V. D. Milman: λ_1 -isoperimetric inequalities for graphs and superconcentrators, *J. Combin. Theory Ser. B* 38 (1985), 73–88.
- [11] N. Alon, L. Rónyai, and T. Szabó: Norm-graphs: variations and applications, *J. Combin. Theory Ser. B* 76 (1999), 280–290.
- [12] L. Babai and B. Guiduli: Spectral extrema for graphs: the Zarankiewicz problem, *Electronic J. Combin.* 15 (2009), R123.
- [13] R. Baer: Polarities in finite projective planes, *Bull. Amer. Math. Soc.* 52 (1946), 77–93.
- [14] C. Balbuena, P. García-Vázquez, X. Marcote, and J. C. Valenzuela: New results on the Zarankiewicz problem, *Discrete Math.* 307 (2007), no. 17–18, 2322–2327.
- [15] C. Balbuena, P. García-Vázquez, X. Marcote, and J. C. Valenzuela: Counterexample to a conjecture of Györi on C_{2l} -free bipartite graphs, *Discrete Math.* 307 (2007), no. 6, 748–749.
- [16] C. Balbuena, P. García-Vázquez, X. Marcote, and J. C. Valenzuela: Extremal $K(s, t)$ -free bipartite graphs, *Discrete Math. Theor. Comput. Sci.* 10 (2008), no. 3, 35–48.
- [17] P. N. Balister, B. Bollobás, O. M. Riordan, and R. H. Schelp: Graphs with large maximum degree containing no odd cycles of a given length, *J. Combin. Theory B* 87 (2003), 366–373.
- [18] P. N. Balister, E. Györi, J. Lehel, and R. H. Schelp: Connected graphs without long paths, *Discrete Math.* 308 (2008), no. 19, 4487–4494.

- [19] S. Ball and V. Pepe: Asymptotic improvements to the lower bound of certain bipartite Turán numbers, *Combin. Probab. Comput.* 21 (2012), no. 3, 323–329.
- [20] J. Beck and J. Spencer: Unit distances, *J. Combin. Theory Ser. A* 37 (1984), 231–238.
- [21] F. Behrend: On sets of integers which contain no three terms in arithmetic progression, *Proc. Nat. Acad. Sci. US.* 32 (1956), 331–332.
- [22] C. T. Benson: Minimal regular graphs of girths eight and twelve, *Canad. J. Math.* 18 (1966), 1091–1094.
- [23] D. Bienstock and E. Györi: An extremal problem on sparse 0-1 matrices, *SIAM J. Discrete Math.* 4 (1991), no. 1, 17–27.
- [24] P. Blagojević, B. Bukh, and R. Karasev: Turán numbers for $K_{s,t}$ -free graphs: topological obstructions and algebraic constructions, arXiv:1108.5254v3, 3 Jun 2012.
- [25] B. Bollobás: Cycles modulo k , *Bull. London Math. Soc.* 9 (1977), no. 1, 97–98.
- [26] B. Bollobás: *Extremal Graph Theory*, Academic Press, London, 1978.
- [27] B. Bollobás: *Random Graphs*, Academic Press, London, 1985.
- [28] B. Bollobás: Extremal graph theory, in: R. L. Graham, M. Grötschel, and L. Lovász (Eds.), *Handbook of Combinatorics*, Elsevier Science, Amsterdam, 1995, pp. 1231–1292.
- [29] B. Bollobás and A. Thomason: Proof of a conjecture of Mader, Erdős and Hajnal on topological subgraphs, *European J. Combin* 19 (1998), 883–887.
- [30] J. A. Bondy: Basic graph theory: paths and circuits, *Handbook of Combinatorics*, Vol. I., pp. 3–110, Elsevier, Amsterdam, 1995.
- [31] J. A. Bondy: Extremal problems of Paul Erdős on circuits in graphs, *Paul Erdős and his mathematics, II* (Budapest, 1999), 135–156, *Bolyai Soc. Math. Stud.*, 11, János Bolyai Math. Soc., Budapest, 2002.
- [32] J. A. Bondy and M. Simonovits: Cycles of even length in graphs, *J. Combin. Theory Ser. B* 16 (1974), 97–105.
- [33] J. A. Bondy and A. Vince: Cycles in a graph whose lengths differ by one or two, *J. Graph Theory* 27 (1998), 11–15.
- [34] S. Brandt and E. Dobson: The Erdős–Sós conjecture for graphs of girth 5, *Selected papers in honour of Paul Erdős on the occasion of his 80th birthday* (Keszthely, 1993), *Discrete Math.* 150 (1996), no. 1–3. 411–414.
- [35] P. Brass: Erdős distance problems in normed spaces, *Comput. Geom.* 6 (1996), no. 4, 195–214.
- [36] W. G. Brown: On graphs that do not contain a Thomsen graph, *Canad. Math. Bull.* 9 (1966), 281–285.
- [37] W. G. Brown: On the non-existence of a type of regular graphs of girth 5, *Canad. J. Math.* 19 (1967), 644–648.
- [38] W. G. Brown and J. W. Moon: Sur les ensembles de sommets indépendants dans les graphes chromatiques minimaux, (French), *Canad. J. Math.* 21 (1969), 274–278.
- [39] W. G. Brown, P. Erdős and V. T. Sós: On the existence of triangulated spheres in 3-graphs, and related problems, *Period Math. Hungar.* 3 (1973), 221–228.

- [40] W. G. Brown, P. Erdős and V. T. Sós: Some extremal problems on r -graphs, *New Directions in the Theory of Graphs* (ed. F. Harary), Academic Press, New York, 1973, pp. 53–63.
- [41] W. G. Brown and M. Simonovits: Digraph extremal problems, hypergraph extremal problems, and the densities of graph structures, *Discrete Math.* 48 (1984), no. 2–3, 147–162.
- [42] W. G. Brown, and M. Simonovits: Extremal multigraph and digraph problems, *Paul Erdős and his mathematics, II* (Budapest, 1999), pp. 157–203, *Bolyai Soc. Math. Stud.*, 11, János Bolyai Math. Soc., Budapest, 2002.
- [43] L. Caccetta and K. Vijayan: Long cycles in subgraphs with prescribed minimum degree, *Discrete Math.* 97 (1991), no. 1–3, 69–81.
- [44] D. de Caen and L. A. Székely: The maximum size of 4- and 6-cycle free bipartite graphs on m, n vertices, *Sets, Graphs and Numbers* (Budapest, 1991), *Colloquium Mathematical Society János Bolyai*, vol. 60, North-Holland, Amsterdam, 1992, pp. 135–142.
- [45] R. Canham: A theorem on arrangements of lines in the plane, *Israel J. Math.* 7 (1969), 393–397.
- [46] F. Chung: Subgraphs of a hypercube containing no small even cycles, *J. Graph Theory* 16 (1992), 273–286.
- [47] F. R. K. Chung and R. L. Graham: *Erdős on Graphs: His Legacy of Unsolved Problems*, A. K. Peters Ltd., Wellesley, MA, 1998.
- [48] C. R. J. Clapham, A. Flockart, and J. Sheehan: Graphs without four-cycles, *J. Graph Theory* 13 (1989), 29–47.
- [49] K. Clarkson, H. Edelsbrunner, L. J. Guibas, M. Sharir, and E. Welzl: Combinatorial complexity bounds for arrangements of curves and spheres, *Discrete Comput. Geom.* 5 (1990), no. 2, 99–160.
- [50] M. Conder: Hexagon-free subgraphs of hypercubes, *J. Graph Theory* 17 (1993), 477–479.
- [51] D. Conlon: An extremal theorem in the hypercube, *Electron. J. Combin.* 17 (2010), Research Paper 111.
- [52] D. Conlon, J. Fox, and B. Sudakov: An approximate version of Sidorenko’s conjecture, *Geom. Funct. Anal.* 20 (2010), no. 6, 1354–1366.
- [53] D. Conlon, J. Fox, and B. Sudakov: Sidorenko’s conjecture for a class of graphs: an exposition, <http://arxiv.org/abs/1209.0184>
- [54] O. Cooley: Proof of the Loebel–Komlós–Sós conjecture for large, dense graphs, *Discrete Math.* 309 (2009), no. 21, 6190–6228.
- [55] K. Čulík: Teilweise Lösung eines verallgemeinerten Problems von K. Zarankiewicz, *Ann. Polon. Math.* 3 (1956), 165–168.
- [56] D. M. Cvetković, M. Doob, and H. Sachs: *Spectra of Graphs*, Academic Press Inc., New York, 1980.
- [57] G. Damásdi, T. Héger, and T. Szőnyi: The Zarankiewicz problem, cages, and geometries, manuscript 2013.
- [58] H. Edelsbrunner and P. Hajnal: A lower bound on the number of unit distances between the vertices of a convex polygon, *J. Combin. Theory Ser. A* 56 (1991), no. 2, 312–316.

- [59] Paul Erdős: Erdős homepage (his scanned in papers up to 1989):
www.renyi.hu/~p_erdos.
- [60] P. Erdős: On sequences of integers no one of which divides the product of two others, and some related problems, *Mitt. Forschungsinst. Math. u. Mech. Tomsk* 2 (1938), 74–82.
- [61] P. Erdős: Graph theory and probability I, *Canad. J. Math.* 11 (1959), 34–38.
- [62] P. Erdős: Graph theory and probability II, *Canad. J. Math.* 13 (1961), 346–352.
- [63] P. Erdős: Extremal problems in graph theory, *Proc. Sympos. Smolenice, 1963*, pp. 29–36, *Publ. House Czechoslovak Acad. Sci., Prague, 1964*.
- [64] P. Erdős: Some applications of probability to graph theory and combinatorial problems, *Theory of Graphs and its Applications (Proc. Sympos. Smolenice, 1963)*, pp. 133–136, *Publ. House Czech. Acad. Sci., Prague, 1964*.
- [65] P. Erdős: On some extremal problems in graph theory, *Israel J. Math.* 3 (1965), 113–116.
- [66] P. Erdős: Some recent results on extremal problems in graph theory, *Theory of Graphs (ed P. Rosenstiehl), (Internat. Sympos., Rome, 1966), Gordon and Breach, New York, and Dunod, Paris, 1967*, pp. 117–123.
- [67] P. Erdős: On some new inequalities concerning extremal properties of graphs, *Theory of Graphs (P. Erdős and G. Katona, Eds.), Academic Press, Nev. York, 1968*, pp. 77–81.
- [68] P. Erdős: *The Art of Counting (ed. J. Spencer), The MIT Press, Cambridge, Mass., 1973*.
- [69] P. Erdős: Problems and results on finite and infinite combinatorial analysis, *in Infinite and Finite Sets (Proc. Conf., Keszthely, Hungary, 1973)*, pp. 403–424, *Proc. Colloq. Math. Soc. J. Bolyai* 10, Bolyai–North-Holland, 1975.
- [70] P. Erdős: Some recent progress on extremal problems in graph theory, *Congr. Numerantium* 14 (1975), 3–14.
- [71] P. Erdős: Problems and results in combinatorial analysis, *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II, Atti dei Convegni Lincei, No. 17*, pp. 3–17, *Accad. Naz. Lincei, Rome, 1976*.
- [72] P. Erdős: Problems and results in graph theory and combinatorial analysis, *Graph theory and related topics (Proc. Conf., Univ. Waterloo, Waterloo, Ont., 1977)*, pp. 153–163, *Academic Press, New York-London*.
- [73] P. Erdős: On the combinatorial problems which I would most like to see solved, *Combinatorica* 1 (1981), no. 1, 25–42.
- [74] P. Erdős: On some problems in graph theory, combinatorial analysis and combinatorial number theory, *Graph Theory and Combinatorics (Cambridge, 1983)*, pp. 1–17, *Academic Press, London, 1984*.
- [75] P. Erdős: Two problems in extremal graph theory. *Graphs Combin.* 2 (1986), no. 1, 189–190.
- [76] P. Erdős: On some of my favourite theorems, *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, 97–132, *Bolyai Soc. Math. Stud.*, 2, *János Bolyai Math. Soc., Budapest, 1996*.
- [77] P. Erdős, R. J. Faudree, J. Pach, and J. Spencer: How to make a graph bipartite, *J. Combin. Theory Ser. B* 45 (1988), no. 1, 86–98.

- [78] P. Erdős, R. J. Faudree, R. H. Schelp, and M. Simonovits: An extremal result for paths, *Graph theory and its applications: East and West* (Jinan, 1986), 155–162, *Ann. New York Acad. Sci.*, 576, New York Acad. Sci., New York, 1989.
- [79] P. Erdős, Z. Füredi, M. Loebl, and V. T. Sós: Discrepancy of trees, *Studia Sci. Math. Hungar.* 30 (1995), no. 1–2, 47–57.
- [80] P. Erdős and T. Gallai: On maximal paths and circuits of graphs, *Acta Math. Acad. Sci. Hungar.* 10 (1959), 337–356.
- [81] P. Erdős, E. Györi, and M. Simonovits: How many edges should be deleted to make a triangle-free graph bipartite? *Sets, graphs and numbers* (Budapest, 1991), pp. 239–263, *Colloq. Math. Soc. János Bolyai*, 60, North-Holland, Amsterdam, 1992.
- [82] P. Erdős, G. Harcos, and J. Pach: Popular distances in 3-space, *Discrete Math.* 200 (1999), no. 1–3, 95–99.
- [83] P. Erdős and L. Moser: Problem 11, *Canad. Math. Bull.* 2 (1959), 43.
- [84] P. Erdős and A. Rényi: On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 5 (1960), 17–61.
- [85] P. Erdős and A. Rényi: On a problem in the theory of graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 7 (1962), 623–641.
- [86] P. Erdős, A. Rényi, and Vera T. Sós: On a problem of graph theory, *Stud Sci. Math. Hung.* 1 (1966), 215–235.
- [87] P. Erdős and H. Sachs: Reguläre Graphen gegebener Tailenweite mit minimaler Knotenzahl (in German), *Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe* 12 (1963), 251–257.
- [88] P. Erdős, A. Sárközy, and V. T. Sós: On product representation of powers, I, *European J. Combin.* 16 (1995), 567–588.
- [89] P. Erdős and M. Simonovits: A limit theorem in graph theory, *Studia Sci. Math. Hungar.* 1 (1966), 51–57.
- [90] P. Erdős and M. Simonovits: Some extremal problems in graph theory, *Combinatorial Theory and Its Applications, I*. (Proc. Colloq. Balatonfüred, 1969), North Holland, Amsterdam, 1970, pp. 377–390.
- [91] P. Erdős and M. Simonovits: An extremal graph problem, *Acta Math. Acad. Sci. Hungar.* 22 (1971/72), 275–282.
- [92] P. Erdős and M. Simonovits: Cube-supersaturated graphs and related problems, *Progress in Graph Theory* (Waterloo, Ont., 1982), pp. 203–218, Academic Press, Toronto, Ont., 1984.
- [93] P. Erdős and M. Simonovits: Compactness results in extremal graph theory, *Combinatorica* 2 (1982), no. 3, 275–288.
- [94] P. Erdős and M. Simonovits: Supersaturated graphs and hypergraphs, *Combinatorica* 3 (1983), 181–192.
- [95] P. Erdős and A. M. Stone: On the structure of linear graphs, *Bull. Amer. Math. Soc* 52 (1946), 1087–1091.
- [96] G. Fan: Distribution of cycle lengths in graphs, *J. Combin. Theory Ser. B* 84 (2002), 187–202.
- [97] G. Fan, Xuezheng Lv, and Pei Wang: Cycles in 2-connected graphs, *J. Combin. Theory Ser. B* 92 (2004), no. 2, 379–394.

- [98] R. J. Faudree and R. H. Schelp: Path Ramsey numbers in multicolorings, *J. Combin. Theory Ser. B* 19 (1975), no. 2, 150–160.
- [99] R. J. Faudree and M. Simonovits: On a class of degenerate extremal graph problems, *Combinatorica* 3 (1983), 83–93.
- [100] R. J. Faudree and M. Simonovits: On a class of degenerate extremal problems II, preprint.
- [101] J. Fox and B. Sudakov: Dependent random choice, *Random Structures Algorithms* 38 (2011), no. 1–2, 68–99.
- [102] F. A. Firke, P. M. Kosek, E. D. Nash, and J. Williford: Extremal graphs without 4-cycles, <http://arxiv.org/pdf/1201.4912v1.pdf>
- [103] Z. Füredi: Graphs without quadrilaterals, *J. Combin. Theory Ser. B* 34 (1983), 187–190.
- [104] Z. Füredi: Quadrilateral-free graphs with maximum number of edges, preprint 1988, http://www.math.uiuc.edu/~z-furedi/PUBS/furedi_C4from1988.pdf
- [105] Z. Füredi: Graphs of diameter 3 with the minimum number of edges, *Graphs Combin.* 6 (1990), no. 4, 333–337.
- [106] Z. Füredi: The maximum number of unit distances in a convex n -gon, *J. Combin. Theory Ser. A* 55 (1990), no. 2, 316–320.
- [107] Z. Füredi: On a Turán type problem of Erdős, *Combinatorica* 11 (1991), 75–79.
- [108] Z. Füredi: Turán type problems, in *Surveys in Combinatorics*, London Math. Soc. Lecture Note Ser. 166, Cambridge University Press, Cambridge, UK, 1991, pp. 253–300.
- [109] Z. Füredi: The maximum number of edges in a minimal graph of diameter 2, *J. Graph Theory* 16 (1992), no. 1, 81–98.
- [110] Z. Füredi: Extremal hypergraphs and combinatorial geometry, *Proceedings of the International Congress of Mathematicians*, Vol. 1, 2 (Zürich, 1994), pp. 1343–1352, Birkhäuser, Basel, 1995.
- [111] Z. Füredi: On the number of edges of quadrilateral-free graphs, *J. Combin. Theory Ser. B* 68 (1996), 1–6.
- [112] Z. Füredi: An upper bound on Zarankiewicz problem, *Combin. Probab. Comput.* 5 (1996), no. 1, 29–33.
- [113] Z. Füredi: New asymptotics for bipartite Turán numbers, *J. Combin. Theory Ser. A* 75 (1996), no. 1, 141–144.
- [114] Z. Füredi: On the number of fivecycles, manuscript, unpublished, superseded by [124].
- [115] Z. Füredi: On a theorem of Erdős and Simonovits on graphs not containing the cube, to appear.
- [116] Z. Füredi and Peter Hajnal: Davenport–Schinzel theory of matrices, *Discrete Math.* 103 (1992), 231–251.
- [117] Z. Füredi, A. Naor, and J. Verstraëte: On the Turán number for the hexagon, *Adv. Math.* 203 (2006), no. 2, 476–496.
- [118] Z. Füredi and L. Özkahya: On even-cycle-free subgraphs of the hypercube, *J. Combin. Theory Ser. A* 118 (2011), 1816–1819.

- [119] Z. Füredi, O. Pikhurko, and M. Simonovits: The Turán density of the hypergraph $\{abc, ade, bde, cde\}$, *Electronic J. Combin.* 10 (2003), R18.
- [120] Z. Füredi and M. Simonovits: Triple systems not containing a Fano configuration, *Combin. Probab. Comput.* 14 (2005), no. 4, 467–484.
- [121] Z. Füredi and D. West: Ramsey theory and bandwidth of graphs, *Graphs and Combin.* 17 (2001), 463–471.
- [122] D. K. Garnick, Y. H. H. Kwong, and F. Lazebnik: Extremal graphs without three-cycles or four-cycles, *J. Graph Theory* 17 (1993), no. 5, 633–645.
- [123] D. K. Garnick, and N. A. Nieuwejaar, Non-isomorphic extremal graphs without three-cycles and four-cycles, *J. Combin. Math. Combin. Comput.* 12 (1992), 33–56.
- [124] A. Grzesik: On the maximum number of five-cycles in a triangle-free graph, *J. Combin. Theory Ser. B* 102 (2012), no. 5, 1061–1066.
- [125] J. R. Griggs and Chih-Chang Ho: On the half-half case of the Zarankiewicz problem, *Discrete Math.* 249 (2002), no. 1–3, 95–104.
- [126] J. Griggs, J. Ouyang: $(0, 1)$ -matrices with no half-half submatrix of ones, *European J. Combin.* 18 (1997), 751–761.
- [127] J. R. Griggs, M. Simonovits, and George Rubin Thomas: Extremal graphs with bounded densities of small subgraphs, *J. Graph Theory* 29 (1998), no. 3, 185–207.
- [128] R. K. Guy: A problem of Zarankiewicz, in: *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pp. 119–150. Academic Press, New York 1968.
- [129] R. K. Guy and S. Znám: A problem of Zarankiewicz, *Recent Progress in Combinatorics (Proc. Third Waterloo Conf. on Combinatorics, 1968)*, pp. 237–243. Academic Press, New York 1969.
- [130] A. Gyárfás: Graphs with k odd cycle lengths, *Discrete Math.* 103 (1992), 41–48.
- [131] A. Gyárfás, J. Komlós, and E. Szemerédi: On the distribution of cycle lengths in graphs, *J. Graph Theory* 8 (1984), 441–462.
- [132] A. Gyárfás, C. C. Rousseau, and R. H. Schelp: An extremal problem for paths in bipartite graphs, *J. Graph Theory* 8 (1984), 83–95.
- [133] E. Győri: On the number of C_5 's in a triangle-free graph, *Combinatorica* 9 (1989), 101–102.
- [134] E. Győri: C_6 -free bipartite graphs and product representation of squares, *Graphs Combin.* (Marseille, 1995), *Discrete Math.* 165/166 (1997), 371–375.
- [135] E. Győri: Triangle-free hypergraphs, *Combin. Prob. Comput.* 15 (2006), 185–191.
- [136] E. Győri, B. Rothschild, and A. Ruciński: Every graph is contained in a sparsest possible balanced graph, *Math. Proc. Cambridge Philos. Soc.* 98 (1985), no. 3, 397–401.
- [137] R. Häggkvist and A. D. Scott: Arithmetic progressions of cycles, *Tech. Rep. Mat. Inst. Umeå Univ.* 16, (1998).
- [138] R. Häggkvist and A. Scott: Cycles of nearly equal length in cubic graphs, Preprint.
- [139] S. Hartman, J. Mycielski, C. Ryll-Nardzewski: Systèmes spéciaux de points à coordonnées entières, *Colloq. Math.* 3 (1954), 84–85, (Bericht Über di Tagung der Poln Math Gesellschaft, Wroclaw, am 20. September 1951.)

- [140] H. Hatami: Graph norms and Sidorenko's conjecture, *Israel J. Math.* 175 (2010), 125–150.
- [141] H. Hatami, J. Hladký, D. Král, S. Norine, and A. Razborov: On the number of pentagons in triangle-free graphs, *J. Combin. Theory Ser. A* 120 (2013), no. 3, 722–732.
- [142] H. Hatami and S. Norine: Undecidability of linear inequalities in graph homomorphism densities, *J. Amer. Math. Soc.* 24 (2011), no. 2, 547–565.
- [143] J. Hladký, J. Komlós, M. Simonovits, M. Stein, and E. Szemerédi: An approximate version of the Loeb–Komlós–Sós Conjecture for sparse graphs, submitted, on arXiv:1211.3050.v1, 2012, Nov 13.
- [144] J. Hladký and D. Piguet: Loeb–Komlós–Sós Conjecture: dense case, Manuscript (arXiv:0805:4834).
- [145] M. N. Huxley and H. Iwaniec: Bombieri's theorem in short intervals, *Mathematika* 22 (1975), 188–194.
- [146] C. Hyltén-Cavallius: On a combinatorial problem, *Colloq. Math.* 6 (1958), 59–65.
- [147] W. Imrich: Explicit construction of graphs without small cycles, *Combinatorica* 4 (1984), 53–59.
- [148] C. Jagger, P. Šťovíček, and A. Thomason: Multiplicities of subgraphs, *Combinatorica* 16 (1996), no. 1, 123–141.
- [149] S. Janson, T. Łuczak, and A. Ruciński: *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000. xii+333 pp.
- [150] T. Jiang: Compact topological minors in graphs, *J. Graph Theory* 67 (2011), 139–152.
- [151] T. Jiang and R. Seiver: Turán numbers of subdivided graphs, *SIAM J. Discrete Math.* 26 (2012), no. 3, 1238–1255.
- [152] S. Józsa and E. Szemerédi: The number of unit distance on the plane, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II, pp. 939–950. *Colloq. Math. Soc. János Bolyai*, Vol. 10, North-Holland, Amsterdam, 1975.
- [153] G. O. H. Katona: Turán's graph theorem and probability theory, *Turán Memorial: Number theory, Analysis and Combinatorics*, de Gruyter, Berlin, to appear.
- [154] Gy. Katona, T. Nemetz, and M. Simonovits: On a problem of Turán in the theory of graphs, *Mat. Lapok* 15 (1964), 228–238.
- [155] P. Keevash: *Hypergraph Turán problems*, *Surveys in Combinatorics*, Cambridge University Press, 2011, 83–140.
- [156] P. Keevash and B. Sudakov: The Turán number of the Fano plane, *Combinatorica* 25 (2005), 561–574.
- [157] P. Keevash, B. Sudakov, and J. Verstraëte: On a conjecture of Erdős and Simonovits: even cycles, *Combinatorica*, to appear.
- [158] M. Klazar: The Füredi–Hajnal conjecture implies the Stanley–Wilf conjecture, in: D. Krob, A. A. Mikhalev, A. V. Mikhalev (Eds.), *Formal Power Series and Algebraic Combinatorics*, Springer, Berlin, 2000, pp. 250–255.

- [159] J. Kollár, L. Rónyai, and T. Szabó: Norm graphs and bipartite Turán numbers, *Combinatorica* 16 (1996), 399–406.
- [160] J. Komlós and E. Szemerédi: Topological cliques in graphs, *Combin. Probab. Comput.* 3 (1994), no. 2, 247–256.
- [161] J. Komlós and E. Szemerédi: Topological cliques in graphs II, *Combin. Probab. Comput.* 5 (1996), 79–90.
- [162] G. N. Kopylov: Maximal paths and cycles in a graph, *Dokl. Akad. Nauk SSSR* 234 (1977), no. 1, 19–21. (English translation: *Soviet Math. Dokl.* 18 (1977), no. 3, 593–596.)
- [163] A. Kostochka and L. Pyber: Small topological complete subgraphs of “dense” graphs, *Combinatorica* 8 (1988), 83–86.
- [164] T. Kővári, V. T. Sós, and P. Turán: On a problem of K. Zarankiewicz, *Colloq. Math.* 3 (1954), 50–57.
- [165] D. Kühn and D. Osthus: Four-cycles in graphs without a given even cycle, *J. Graph Theory* 48 (2005), 147–156.
- [166] T. Lam and J. Verstraëte: A note on graphs without short even cycles, *Electron. J. Combin.* 12 (2005), Note 5, 6 pp.
- [167] F. Lazebnik and D. Mubayi: New lower bounds for Ramsey numbers of graphs and hypergraphs, *Adv. in Appl. Math.* 28 (2002), no. 3–4, 544–559.
- [168] F. Lazebnik and V. A. Ustimenko, New examples of graphs without small cycles and of large size, *European J. Combin.* 14 (1993), no. 5, 445–460.
- [169] F. Lazebnik, V. A. Ustimenko, and A. J. Woldar: Properties of certain families of $2k$ -cycle-free graphs, *J. Combin. Theory Ser. B* 60 (1994), no. 2, 293–298.
- [170] F. Lazebnik, V. A. Ustimenko, and A. J. Woldar: A new series of dense graphs of high girth, *Bull. Amer. Math. Soc.* 32 (1995), no. 1, 73–79.
- [171] F. Lazebnik, V. A. Ustimenko, and A. J. Woldar, Polarities and $2k$ -cycle-free graphs, *Discrete Math.* 197/198 (1999), 503–513.
- [172] F. Lazebnik and A. J. Woldar: General properties of some families of graphs defined by systems of equations, *J. Graph Theory* 38 (2001), no. 2, 65–86.
- [173] J. Lenz and D. Mubayi: Multicolor Ramsey numbers for complete bipartite versus complete graphs, *arXiv* 1201.2123, 26 pp.
- [174] B. Lidický, Hong Liu, and C. Palmer: On the Turán number of forests, *arXiv* 1204.3102.
- [175] L. Lovász: Independent sets in critical chromatic graphs, *Studia Sci. Math. Hungar.* 8 (1973), 165–168.
- [176] L. Lovász: *Combinatorial Problems and Exercises*, 2nd Ed., North-Holland, Amsterdam, 1993.
- [177] L. Lovász: *Large Networks and Graph Limits*, Colloquium Publications 2012, 475 pp.
- [178] L. Lovász and M. Simonovits: On the number of complete subgraphs of a graph II, *Studies in Pure Mathematics*, pp. 459–495, (dedicated to the memory of P. Turán), Akadémiai Kiadó and Birkhäuser Verlag 1982.
- [179] A. A. Razborov: Flag algebras, *J. Symbolic Logic* 72 (2007), no. 4, 1239–1282.

- [180] Linyuan Lu: Hexagon-free subgraphs in hypercube Q_n , private communication.
- [181] A. Lubotzky, R. Phillips, and P. Sarnak: Ramanujan graphs, *Combinatorica* 8 (1988), no. 3, 261–277.
- [182] A. McLennan: The Erdős–Sós conjecture for trees of diameter four, *J. Graph Theory* 49 (2005), no. 4, 291–301.
- [183] W. Mader: Homomorphieeigenschaften und mittlere Kantendichte von Graphen, *Math. Ann.* 174 (1967), 265–268.
- [184] W. Mader: Topological subgraphs in graphs of large girth, *Combinatorica* 18 (1998), no. 3, 405–412.
- [185] W. Mader: Topological minors in graphs of minimum degree n , *Contemporary trends in discrete mathematics (Štirín Castle, 1997)*, 199–211, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., 49, Amer. Math. Soc., Providence, RI, 1999.
- [186] W. Mader: Graphs with $3n - 6$ edges not containing a subdivision of K_5 , *Combinatorica* 25 (2005), no. 4, 425–438.
- [187] A. Marcus and G. Tardos: Excluded permutation matrices and the Stanley–Wilf conjecture, *J. Combin. Theory Ser. A* 107 (2004), no. 1, 153–160.
- [188] G. A. Margulis: Explicit construction of graphs without short cycles and low density codes, *Combinatorica* 2 (1982), 71–78.
- [189] G. A. Margulis: Arithmetic groups and graphs without short cycles, in: *6th Int. Symp. on Information Theory, Tashkent, Abstracts 1, 1984*, pp. 123–125 (in Russian).
- [190] G. A. Margulis: Explicit group-theoretical construction of combinatorial schemes and their application to the design of expanders and concentrators, *J. Problems of Inform. Trans.* 24 (1988), 39–46; translation from *Problemy Peredachi Informatsii* 24 (January–March 1988), 51–60.
- [191] G. Megyesi and E. Szabó: On the tacnodes of configurations of conics in the projective plane, *Math. Ann.* 305 (1996), no. 4, 693–703.
- [192] M. Molloy and B. Reed: *Graph Colouring and the Probabilistic Method*, Algorithms and Combinatorics, 23. Springer-Verlag, Berlin, 2002, xiv+326 pp.
- [193] B. Montágh: Unavoidable substructures, PHD Thesis, University of Memphis, May 2005.
- [194] M. Mörs: A new result on the problem of Zarankiewicz, *J. Combin. Theory Ser. A* 31 (1981), no. 2, 126–130.
- [195] D. Mubayi and Gy. Turán: Finding bipartite subgraphs efficiently, *Inform. Process. Lett.* 110 (2010), no. 5, 174–177.
- [196] Z. L. Nagy: A multipartite version of the Turán problem – density conditions and eigenvalues, *Electron. J. Combin.* 18 (2011), no. 1, Paper 46, 15 pp.
- [197] V. Nikiforov: Bounds on graph eigenvalues II, *Linear Algebra Appl.* 427 (2007), 183–189.
- [198] V. Nikiforov: A contribution to the Zarankiewicz problem, *Linear Algebra Appl.* 432 (2010), no. 6, 1405–1411.
- [199] J. Pach and P. K. Agarwal: *Combinatorial Geometry*, Wiley-Interscience, New York, 1995. xiv+354 pp.

- [200] P. P. Pálffy and M. Szalay: in Turán Memorial: Number theory, Analysis and Combinatorics, de Gruyter, Berlin, to appear.
- [201] D. Piguet and M. J. Stein: Loebel–Kömös–Sós conjecture for trees of diameter 5, *Electron. J. Combin.*, 15 (2008), Research Paper 106, 11 pp.
- [202] D. Piguet and M. J. Stein: An approximate version of the Loebel–Kömös–Sós conjecture, *J. Combin. Theory Ser. B* 102 (2012), no. 1, 102–125.
- [203] O. Pikhurko: A note on the Turán Function of even cycles, *Proc. Amer. Math Soc.* 140 (2012), 3687–3992.
- [204] R. Pinchasi and M. Sharir: On graphs that do not contain the cube and related problems, *Combinatorica* 25 (2005), no. 5, 615–623.
- [205] C. Reiher: The clique density theorem, arxiv1212.2454.
- [206] I. Reiman: Über ein Problem von K. Zarankiewicz, *Acta Math. Acad. Sci. Hungar.* 9 (1958), no. 3–4, 269–273.
- [207] I. Reiman: An extremal problem in graph theory, (Hungarian). *Mat. Lapok* 12 (1961), 44–53.
- [208] A. Rényi: Selected Papers of Alfréd Rényi, Akadémiai Kiadó, 1976 (ed. Paul Turán).
- [209] V. Rödl and M. Schacht: Extremal results for random graphs, in this volume.
- [210] I. Z. Ruzsa and E. Szemerédi: Triple systems with no six points carrying three triangles, *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976)*, Vol. II, pp. 939–945, *Colloq. Math. Soc. János Bolyai*, 18, North-Holland, Amsterdam-New York, 1978.
- [211] J.-F. Saclé and M. Woźniak: A note on the Erdős–Sós conjecture for graphs without C_4 , *J. Combin. Theory Ser. B* 70 (1997), no. 2, 367–372.
- [212] G. N. Sárközy: Cycles in bipartite graphs and an application in number theory, *J. Graph Theory*, 19 (1995), 323–331.
- [213] A. Scott: Szemerédi’s regularity lemma for matrices and sparse graphs, *Combin. Probab. Comput.* 20 (2011), no. 3, 455–466.
- [214] Jian Shen: On two Turán numbers, *J. Graph Theory* 51 (2006), 244–250.
- [215] A. F. Sidorenko: Asymptotic solution for a new class of forbidden r -graphs, *Combinatorica* 9 (1989), no. 2, 207–215.
- [216] A. Sidorenko: A correlation inequality for bipartite graphs, *Graphs Combin.* 9 (1993), no. 2, 201–204.
- [217] A. F. Sidorenko: What do we know and what we do not know about Turán Numbers, *Graphs Combin.* 11 (1995), no. 2, 179–199.
- [218] M. Simonovits: A method for solving extremal problems in graph theory, *Theory of Graphs, Proc. Colloq. Tihany, (1966)*, (P. Erdős and G. Katona, Eds.), pp. 279–319, Acad. Press, New York, 1968.
- [219] M. Simonovits: On colour-critical graphs, *Studia Sci. Math. Hungar.* 7 (1972), 67–81.
- [220] M. Simonovits: Note on a hypergraph extremal problem, *Hypergraph Seminar, Columbus Ohio USA, 1972*, (C. Berge and D. K. Ray-Chaudhuri, Eds.), *Lecture Notes in Mathematics* 411, pp. 147–151, Springer Verlag, 1974.

- [221] M. Simonovits: Extremal graph problems with symmetrical extremal graphs, additional chromatic conditions, *Discrete Math.* 7 (1974), 349–376.
- [222] M. Simonovits: On Paul Turán’s influence on graph theory, *J. Graph Theory* 1 (1977), no. 2, 102–116.
- [223] M. Simonovits: Extremal graph problems and graph products, *Studies in Pure Mathematics*, pp. 669–680, (dedicated to the memory of P. Turán), Akadémiai Kiadó and Birkhäuser Verlag 1982.
- [224] M. Simonovits: Extremal graph theory, in: L. W. Beineke, R. J. Wilson (Eds.), *Selected Topics in Graph Theory II.*, pp. 161–200, Academic Press, London, 1983.
- [225] M. Simonovits: Extremal graph problems, degenerate extremal problems and supersaturated graphs, *Progress in graph Theory*, (Bondy and Murty, Eds.), pp. 419–438, Academic Press, 1984.
- [226] M. Simonovits: How to solve a Turán type extremal graph problem? (linear decomposition), *Contemporary trends in discrete mathematics (Stirin Castle, 1997)*, pp. 283–305, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., 49, Amer. Math. Soc., Providence, RI, 1999.
- [227] M. Simonovits: Paul Erdős’ influence on extremal graph theory, *The mathematics of Paul Erdős, II.*, pp. 148–192, *Algorithms Combin.*, 14, Springer, Berlin, 1997.
- [228] M. Simonovits: Paul Erdős’ influence on extremal graph theory, The new version of the old paper [227].
- [229] M. Simonovits: Paul Turán’s influence in Combinatorics, in *Turán Memorial: Number Theory, Analysis, and Combinations*, De Gruyter, to appear.
- [230] M. Simonovits and V. T. Sós: Ramsey–Turán theory, *Combinatorics, graph theory, algorithms and applications*, *Discrete Math.* 229 (2001), no. 1–3, 293–340.
- [231] R. R. Singleton: On minimal graphs of maximum even girth, *J. Combinatorial Theory* 1 (1966), 306–332.
- [232] V. T. Sós: Remarks on the connection of graph theory, finite geometry and block designs, *Colloquio Internazionale sulle Teorie Combinatorie (Roma, 1973)*, Tomo II, pp. 223–233, *Atti dei Convegni Lincei*, No. 17, Accad. Naz. Lincei, Rome, 1976.
- [233] J. Spencer, E. Szemerédi, and W. T. Trotter: Unit distances in the Euclidean plane, *Graph theory and combinatorics (Cambridge, 1983)*, pp. 293–303, Academic Press, London, 1984.
- [234] B. Sudakov and J. Verstraëte: Cycle lengths in sparse graphs, *Combinatorica* 28 (2008), no. 3, 357–372.
- [235] E. Szemerédi: Regular partitions of graphs, *Problemes Combinatoires et Theorie des Graphes* (ed. I.-C. Bermond et al.), pp. 399–401, CNRS, Paris, 1978.
- [236] G. Tardos: On 0-1 matrices and small excluded submatrices, *J. Combin. Th. Ser. A* 111 (2005), 266–288.
- [237] A. G. Thomason: A disproof of a conjecture of Erdős in Ramsey Theory, *J. London Math. Soc.* 39 (1989), 246–255.
- [238] A. Thomason and P. Wagner: Bounding the size of square-free subgraphs of the hypercube, *Discrete Math.* 309 (2009), 1730–1735.
- [239] C. M. Timmons: Ordered Turán Problems, Lecture no. 1086-05-1067 on the Joint Mathematics Meetings, San Diego, CA, January 9, 2013.

- [240] B. Toft: Two theorems on critical 4-chromatic graphs, *Studia Sci. Math. Hungar.* 7 (1972), 83–89.
- [241] P. Turán: On a theorem of Hardy-Ramanujan, *Journal of London Math Soc.* 9 (1934), 274–276.
- [242] P. Turán: On an extremal problem in graph theory, (Hungarian), *Mat. Fiz. Lapok* 48 (1941), 436–452.
- [243] P. Turán: On the theory of graphs, *Colloq. Math.* 3 (1954), 19–30.
- [244] P. Turán: A note of welcome, *J. Graph Theory* 1 (1977), 7–9.
- [245] P. Valtr: Strictly convex norms allowing many unit distances and related touching questions, manuscript.
- [246] J. Verstraëte: On arithmetic progressions of cycle lengths in graphs, *Combin. Probab. Comput.* 9 (2000), no. 4, 369–373.
- [247] R. Wenger: Extremal graphs with no C_4 's, C_6 's, or C_{10} 's, *J. Combin. Theory Ser. B* 52 (1991), no. 1, 113–116.
- [248] R. M. Wilson: An existence theory for pairwise balanced designs, III. Proof of the existence conjectures, *J. Combin. Theory Ser. A* 18 (1975), 71–79.
- [249] D. R. Woodall: Maximal circuits of graphs I, *Acta Math. Acad. Sci. Hungar.* 28 (1976), no. 1–2, 77–80.
- [250] D. R. Woodall: Maximal circuits of graphs II, *Studia Sci. Math. Hungar.* 10 (1975), no. 1–2, 103–109.
- [251] M. Woźniak: On the Erdős-Sós conjecture, *J. Graph Theory*, 21 (1996), no. 2, 229–234.
- [252] Y. Yuansheng and P. Rowlinson: On extremal graphs without four-cycles, *Utilitas Math.* 41 (1992), 204–210.
- [253] Y. Yuansheng and P. Rowlinson: On graphs without 6-cycles and related Ramsey numbers, *Utilitas Math.* 44 (1993), 192–196.
- [254] K. Zarankiewicz: Problem 101, *Colloquium Mathematicum* 2 (1951), p. 301.
- [255] Yi Zhao: Proof of the $(n/2 - n/2 - n/2)$ conjecture for large n , *Electron. J. Combin.* 18 (2011), Paper 27.
- [256] Š. Znám: On a combinatorical problem of K. Zarankiewicz, *Colloq. Math.* 11 (1963), 81–84.
- [257] Š. Znám: Two improvements of a result concerning a problem of K. Zarankiewicz, *Colloq. Math.* 13 (1964/1965), 255–258.

Zoltán Füredi

*Alfréd Rényi Institute of
Mathematics,
Hungarian Academy of Sciences,
Budapest, Reáltanoda u. 13–15,
H-1053, Hungary*

e-mail: z-furedi@illinois.edu

Miklós Simonovits

*Alfréd Rényi Institute of
Mathematics,
Hungarian Academy of Sciences,
Budapest, Reáltanoda u. 13–15,
H-1053, Hungary*

e-mail:

simonovits.miklos@renyi.mta.hu

ERDŐS AND ARITHMETIC PROGRESSIONS

W. TIMOTHY GOWERS

Two of Erdős's most famous conjectures concern arithmetic progressions. In this paper we discuss some of the progress that has been made on them.

1. INTRODUCTION

Possibly the best known of all of Erdős's many conjectures is the following striking statement.

Conjecture 1.1. *Let A be a set of positive integers such that $\sum_{n \in A} n^{-1} = \infty$. Then A contains arbitrarily long arithmetic progressions.*

This conjecture is still wide open. Indeed, it is not even known whether A must contain an arithmetic progression of length 3.

There is another conjecture of Erdős about arithmetic progressions. It is not as famous as the first, but it is still well known and extremely interesting. It is sometimes referred to as *Erdős's discrepancy problem*.

Conjecture 1.2. *Let $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots$ be a sequence taking values in the set $\{-1, 1\}$. Then for every constant C there exist positive integers n and d such that $|\sum_{m=1}^n \varepsilon_{md}| \geq C$.*

The purpose of this paper is to say a little bit about the two conjectures and to discuss some known results and related problems.

2. ARITHMETIC PROGRESSIONS IN SPARSE SETS

What does it tell us about a set A if $\sum_{n \in A} n^{-1}$ diverges? Clearly it tells us that in some sense A is not too small, since the larger it is, the more likely the sum of its reciprocals is to diverge. A rough interpretation of the condition turns out to be that the density $\delta(n) = n^{-1}|A \cap \{1, 2, \dots, n\}|$ decreases not too much faster than $(\log n)^{-1}$. One way of seeing this is as follows. Writing 1_A for the characteristic function of A , we have the trivial identity

$$1_A(n) = n\delta(n) - (n-1)\delta(n-1),$$

from which (if we adopt the convention that $\delta(0) = 0$) it follows that

$$\begin{aligned} \sum_{n \in A} n^{-1} &= \sum_n n^{-1} 1_A(n) = \sum_n (\delta(n) - \delta(n-1) + \delta(n-1)/n) \\ &= \sum_n \delta(n-1)/n. \end{aligned}$$

Thus, if the density decreases like $(\log n)^{-1}$ then we get a sum like

$$\sum_n 1/n \log n,$$

which diverges, while if it decreases like, say, $(\log n)^{-1}(\log \log n)^{-2}$, then we get a convergent sum.

Of course, the density does not have to decrease smoothly in this way, but this nevertheless gives a good general picture of what the conjecture is saying. In particular, the simple calculation just given tells us that if $\sum_{n \in A} n^{-1} = \infty$, then there must be infinitely many n for which $\delta(n) \geq (\log n)^{-1}(\log \log n)^{-2}$, so to prove Erdős's conjecture it is sufficient to prove the following statement.

Conjecture 2.1. *For every k there exists n such that if A is any subset of $\{1, \dots, n\}$ of cardinality at least $n/\log n(\log \log n)^2$, then A contains an arithmetic progression of length k .*

It is also not hard to show that to *disprove* Erdős's conjecture, it would be sufficient to show that for every k and every sufficiently large n there exists a subset $A \subset \{1, \dots, n\}$ of cardinality at least $n/\log n$ that does not contain an arithmetic progression of length k . To do this, for each sufficiently large r let A_r be a subset of $\{2^r + 1, \dots, 2^{r+1}\}$ of size at least

$cr^{-1}2^r$ that contains no arithmetic progression of length k and let A be the infinite set $A_s \cup A_{s+2} \cup A_{s+4} \cup \dots$ for a sufficiently large s . Then for every sufficiently large n we have $\delta(n) \geq c'(\log n)^{-1}$ and A contains no arithmetic progression of length k .

Thus, Erdős's conjecture is basically addressing the following problem, and suggesting an approximate answer.

Problem 2.2. Let k and n be positive integers. How large does a subset $A \subset \{1, 2, \dots, n\}$ have to be to guarantee that it contains an arithmetic progression of length k ?

The suggested answer is that a cardinality of somewhere around $n/\log n$ should be enough.

A natural starting point would be to prove *any* bound of the form $o(n)$. This gives us another famous conjecture of Erdős, made with Paul Turán in 1936 [10].

Conjecture 2.3. For every positive integer k and every $\delta > 0$ there exists n such that every subset $A \subset \{1, 2, \dots, n\}$ of cardinality at least δn contains an arithmetic progression of length k .

Even this much weaker conjecture turned out to be very hard, and very interesting indeed: it can be seen as having given rise to several different branches of mathematics.

The first progress on the Erdős-Turán conjecture was due to Roth, who proved in 1953 that it is true when $k = 3$ [31]. Roth's proof, which used Fourier analysis, showed that δ could be taken to be $C/\log \log n$ for an absolute constant C . The problem for longer progressions turned out to be much harder, and it was not until 1969 that there was further progress, when Szemerédi proved the result for $k = 4$ [36], this time with a bound for δ that was too weak to be worth stating explicitly. And a few years later (the paper was published in 1975), Szemerédi managed to prove the general case [37].

2.1. Other proofs of Szemerédi's theorem

This result was hailed at the time and is still regarded as one of the great mathematical results of the second half of the twentieth century, but it was by no means the end of the story: over the last four decades its significance has steadily grown. In this respect, the Erdős-Turán conjecture is like many conjectures of Erdős. Initially it seems like an amusing puzzle, but the more you think about it, the more you come to understand that the "amusing

puzzle” is a brilliant distillation of a much more fundamental mathematical difficulty. There are few direct applications of Szemerédi’s theorem (though they do exist), but an enormous number of applications of the methods that Szemerédi developed to prove the theorem, and in particular of his famous regularity lemma.

Since then, there have been several other proofs of the theorem, which have also introduced ideas with applications that go well beyond Szemerédi’s theorem itself. In 1977, Furstenberg pioneered an ergodic-theoretic approach [11], giving a new proof of the theorem and developing a method that went on to yield the first proofs of many generalizations, of which we mention three notable ones.

The first is a natural multidimensional version of Szemerédi’s theorem, due to Furstenberg and Katznelson [12].

Theorem 2.4. *For every $\delta > 0$, every positive integer d and every subset $K \subset \mathbb{Z}^d$ there exists n such that every subset $A \subset \{1, \dots, n\}^d$ of size at least δn^d contains a homothetic copy of K : that is, a set of the form $aK + b$ for some positive integer a and some $b \in \mathbb{Z}^d$.*

Next, we have the “density Hales–Jewett theorem”, also due to Furstenberg and Katznelson [13]. For this we need a definition. If x is a point in $\{1, \dots, k\}^n$ and E is a subset of $\{1, 2, \dots, n\}$, then for each $1 \leq j \leq k$ let $x \oplus jE$ be the point $y \in \{1, \dots, k\}^n$ such that $y_i = j$ for every $i \in E$ and $y_i = x_i$ otherwise. A *combinatorial line* in $\{1, \dots, k\}^n$ is a set of points of the form $\{x \oplus jE : j = 1, \dots, k\}$.

Theorem 2.5. *For every $\delta > 0$ and every k there exists n such that every subset $A \subset \{1, \dots, k\}^n$ of cardinality at least δk^n contains a combinatorial line.*

Finally, the Bergelson-Leibman theorem [2] is the following remarkable “polynomial version” of Szemerédi’s theorem.

Theorem 2.6. *For every $\delta > 0$ and every sequence P_1, \dots, P_k of polynomials with integer coefficients and no constant term there exists n such that every subset $A \subset \{1, 2, \dots, n\}$ of cardinality at least δn contains a subset of the form $\{a + P_1(d), a + P_2(d), \dots, a + P_k(d)\}$ with $d \neq 0$.*

If we take $P_i(d)$ to be $(i - 1)d$, then we recover Szemerédi’s theorem, but this result is considerably more general. For example, amongst many other things it implies that in Szemerédi’s theorem we can ask for the common difference of the arithmetic progression we obtain to be a perfect cube.

Another approach to Szemerédi’s theorem was discovered approximately twenty years later by the author [16, 17]. One of the reasons that Roth’s

proof for progressions of length 3 was not quickly followed by a proof of the general case was that while the number of arithmetic progressions of length 3 in a set can be expressed very nicely in terms of Fourier coefficients, there is no useful Fourier expression for the number of arithmetic progressions of length 4 (or more). The proofs in [16, 17] replaced the trigonometric functions that Roth used by polynomial phase functions (that is, functions of the form $\exp(2\pi ip(x))$ for some polynomial p) restricted to arithmetic progressions. This strongly suggested that there should be a kind of “higher-order Fourier analysis”, and, in a major recent achievement, such a theory was worked out by Green, Tao and Ziegler [22] (see also [20, 4]). Their *inverse theorem for the uniformity norms* had a very important application that we shall describe briefly later.

A fourth approach to the theorem had its roots in a fascinating argument of Ruzsa and Szemerédi [32], who used Szemerédi’s regularity lemma to prove the following result, which is now known as the *triangle removal lemma*.

Theorem 2.7. *For every $\varepsilon > 0$ there exists $\delta > 0$ such that if G is any graph with n vertices and at most δn^3 triangles, then there is a triangle-free graph that differs from G by at most εn^2 edges.*

By applying the triangle removal lemma to a suitably chosen graph, one can deduce Roth’s theorem (with a much worse bound).

It is natural to wonder whether this idea can be generalized to give a proof of the general case of Szemerédi’s theorem. This thought led Rödl to formulate an approach to the theorem in which the regularity lemma was generalized from graphs to hypergraphs. The generalization is not straightforward to state, and proving both it and an associated “counting lemma” turned out to be hard. Frankl and Rödl proved a hypergraph regularity lemma in 1992 [14] and in 2002 managed to use it to prove Szemerédi’s theorem for progressions of length 4 [15]. The general case was proved by this method in independent work of Nagle, Rödl and Schacht [27] and the author [18]. (In the latter proof the formulation of the hypergraph regularity lemma was different, which made it harder to prove but made the counting lemma easier to prove.) Hypergraph regularity has gone on to have several other applications.

An important development in our understanding of the regularity lemma came with work of Lovász and others on *graph limits*. Loosely speaking, with the help of the regularity lemma one can show that very large graphs look like measurable functions from $[0, 1]^2$ to $[0, 1]$. In a way this is not too surprising, because the regularity lemma allows one to approximate any graph with just a bounded amount of information about densities between

subsets. What is more surprising, however, is that the graph-limits point of view leads to a simpler proof of the regularity lemma itself [24]: for the limiting arguments one can use a weaker regularity lemma, and once one has passed to a measurable function on $[0, 1]^2$, one has a limit of step functions, which implies that if one partitions into a very fine grid, then the function will be approximately constant on most squares.

Once one is given the statement of Szemerédi's regularity lemma and the basic idea of the standard proof, working out the details is not especially hard to begin with. However, the limits approach generalizes to hypergraphs [9], where proving corresponding results is much harder, and gives rise to similar simplifications. The resulting hypergraph-limits approach to Szemerédi's theorem has a strong claim to be the simplest known proof of the theorem. More generally, graph and hypergraph limits have become a very active area of research with several other applications.

We briefly mention one other candidate for the simplest known proof of Szemerédi's theorem, which is a combinatorial proof of the density Hales–Jewett theorem, discovered by a “massive online collaboration” [28]. It is easy to see that the density Hales–Jewett theorem implies Szemerédi's theorem: one just needs to interpret the points in $\{1, \dots, k\}^n$ as base- k representations of integers, and then every combinatorial line is an arithmetic progression of length k (but not vice versa). Recently, this proof has been simplified yet further [8].

2.2. Quantitative considerations

As we saw earlier, Conjecture 1.1 is roughly saying that a density of $(\log n)^{-1}$ is enough to guarantee an arithmetic progression. But what is special about this bound? Indeed, *is* it special?

There are two sensible answers to this question: yes and no. The reason the bound is special, and the reason that Erdős asked the question, is that the primes have density around $(\log n)^{-1}$ in the first n integers. One of Erdős's formative mathematical experiences was proving for himself that the sum of the reciprocals of the primes diverges, and it is clear that his main motivation for the sum-of-reciprocals conjecture was that it would imply that the prime numbers contain arbitrarily long arithmetic progressions. This would be an example of a result of a kind that Erdős particularly liked: a result that appears to be number-theoretic but turns out to be true for purely combinatorial reasons.

It would have been fascinating to know how Erdős would have reacted to the proof by Green and Tao [19] that the primes do indeed contain

arbitrarily long arithmetic progressions. In fact, Green and Tao proved the following stronger result.

Theorem 2.8. *For every $\delta > 0$ and every k there exists n such that if A is any set of at least $\delta n / \log n$ primes between 1 and n , then A contains an arithmetic progression of length k .*

That is, not only do the primes contain arbitrarily long arithmetic progressions, but so does any subset of the primes of positive relative density. (Of course, this too is implied by the sum-of-reciprocals conjecture.)

The proof of this celebrated result did not go according to Erdős's plan, in that it made significant use of distribution properties of the primes. However, despite this, it would almost certainly have appealed to Erdős's love of combinatorial arguments, since the main new ingredient in the proof was in a sense "purely combinatorial": they proved a "relative version" of Szemerédi's theorem, showing that a set A that is a relatively dense subset of a set B must contain an arithmetic progression of length k , provided that B is sufficiently large and sufficiently "pseudorandom" in a technical sense that they defined. (The result they stated and used was actually more general than this: B was replaced by a "pseudorandom measure".) In order to prove this result, they used Szemerédi's theorem as well as techniques from several of the proofs of the theorem. Thus, the work on the Erdős-Turán conjecture did in the end result in a solution to the problem that so fascinated Erdős.

Green and Tao followed this theorem with a project to obtain asymptotic bounds for the number of arithmetic progressions of length k (and many other configurations) in the primes up to n . Over several years, they published a sequence of major papers, culminating in a proof, with Tamar Ziegler, of the inverse theorem for the uniformity norms [22], mentioned earlier, at which point the project was completed.

2.2.1. How natural is Erdős's conjecture? The fact that Erdős's conjecture implies an extremely striking result about the primes is not really evidence that the correct bound in Szemerédi's theorem is anywhere near $\delta = (\log n)^{-1}$. Obtaining such a bound would be wonderful, but there is no strong reason to suppose that it would be the last word on the subject.

In particular, the best known *lower* bound for Szemerédi's theorem is far smaller than $(\log n)^{-1}$. It comes from a construction of Behrend in 1946 [1]. Behrend started from the observation that the surface of a sphere contains no three points in a line, and in particular no three points such that one is the midpoint of the other two. The argument proceeds as follows. For suitable integers m and d , to be optimized at the end of the argument, one shows by

the pigeonhole principle that there exists r such that the sphere of radius r contains many points in the grid $\{1, \dots, m\}^d$. Next, one embeds that grid “isomorphically” into the set $\{1, 2, \dots, (2m)^d\}$ by thinking of the points in $\{1, \dots, m\}^d$ as base- $2m$ representations of integers. The main property of this “isomorphism” is that it does not create any arithmetic progressions of length 3 that were not present before. Finally, one maximizes the number of points in the spherical surface subject to the constraint that $(2m)^d = n$. The resulting bound is $\delta = \exp(-c\sqrt{\log n})$.

This bound helps to explain why it is so hard to determine optimal bounds for Szemerédi’s theorem, even when the progressions have length 3. On a first acquaintance with the problem, it is natural to conjecture that the extremal example would be given by a simple probabilistic construction. If that were the case, then there would be hope of proving that that construction was best possible by showing that “quasirandom sets are best”. An approach like this works, for example, if one wishes to minimize, for a given cardinality of a subset $A \subset \mathbb{Z}/n\mathbb{Z}$, the number of quadruples $(a_1, a_2, a_3, a_4) \in A^4$ such that $a_1 + a_2 = a_3 + a_4$, at least when that cardinality is significantly greater than \sqrt{n} . However, random sets do not work for progressions of length 3: the standard method of choosing points randomly with probability p , where p is chosen such that the expected number of progressions of length 3 is at most half the expected number of points, and then deleting a point from each progression, gives a lower bound of $\delta = cn^{-2/3}$, far smaller than the Behrend bound.

The Behrend bound can be slightly improved when the progressions are longer, but for now let us focus on progressions of length 3. What is the correct bound for the first non-trivial case of Szemerédi’s theorem? This is a fascinating question that is still wide open, despite the attention of many mathematicians. However, there has been some very interesting progress.

As mentioned earlier, the original argument of Roth gave an upper bound of $C(\log \log n)^{-1}$. This bound was improved to one of the form $(\log n)^{-c}$ by Heath-Brown [23] and Szemerédi [38]. An important new technique, the use of regular Bohr sets, was introduced by Bourgain in 1999 [6], to improve the constant c . More precisely, he obtained a bound of $C(\log \log n / \log n)^{1/2}$. A difficulty with the problem is that cyclic groups are not rich in subgroups, so dropping down to a subgroup is not an option. Regular Bohr sets are a kind of substitute for subgroups, allowing Bourgain to get round this difficulty. They have subsequently been used in many other proofs.

For a while, Bourgain’s result was seen as the limit of what could be achieved without a radical change of approach. It therefore came as a surprise in 2008 when Bourgain introduced an idea that allowed him to

carry out the general scheme of his proof more efficiently and obtain a power of $2/3$ instead of $1/2$. Sanders [33] pushed this approach further and obtained a power of $3/4$.

Sanders followed up this improvement with a major advance on the problem [34]. He found an argument that was substantially different from Bourgain's and used it to obtain a bound of $C(\log \log n)^5 / \log n$. Thus, he was tantalizingly close to the logarithmic barrier. In fact, even a bound of $c \log \log n / \log n$ would be enough to prove purely combinatorially that the primes contain infinitely many arithmetic progressions of length 3, since if m is a number with many small prime factors, then most arithmetic progressions with common difference m contain almost no primes, which means that some have a high density of primes. Working out the details, one can find arithmetic progressions of length n in which the primes have density $c \log \log n / \log n$.

2.2.2. What is the right bound for Roth's theorem? That is where things stand today. Is the Behrend bound correct, or is Sanders's upper bound close to optimal? Nobody knows, but there there are two recent results that give weakish evidence that the Behrend bound is more like the truth of the matter.

The first of these concerns a closely related problem about subsets of \mathbb{F}_3^n (where \mathbb{F}_3 is the field with three elements). How large must a subset of \mathbb{F}_3^n be to guarantee that it contains an affine line, or equivalently three points x, y, z such that $x + y + z = 0$? (Such a triple can also be thought of as an arithmetic progression, since if $x + y + z = 0$, then $2y = x + z$.)

It was observed by Meshulam that Roth's original argument works very cleanly in this context (the main reason being that, in contrast with the cyclic group $\mathbb{Z}/n\mathbb{Z}$, the group \mathbb{F}_3^n is very rich in subgroups), and yields the following theorem [26].

Theorem 2.9. *There exists a constant C such that every subset $A \subset \mathbb{F}_3^n$ of density at least C/n contains an affine line.*

Thus, in this context, we have a logarithmic bound (since n is logarithmic in the size, 3^n , of the set \mathbb{F}_3^n).

The gap between this and the best known lower bound is even more embarrassingly large than it is for Roth's theorem, since the lower bound is of the form α^n for some constant $\alpha < 3$. (To obtain such a lower bound, one finds a low-dimensional example and takes powers of that example.)

It was felt by many people that this was a better problem to attack than attempting to improve the bounds in Roth's theorem, since working in the group \mathbb{F}_3^n presented technical simplifications without avoiding the deeper

mathematical difficulties. And yet, despite the simplicity of the arguments for both the upper and lower bounds, for many years nobody could come up with any improvement. There was therefore considerable excitement in 2011 when Bateman and Katz [3] broke the logarithmic barrier for this problem, improving the upper bound to $C/n^{1+\varepsilon}$ for a small but fixed positive ε . Initially there was a hope that it might be possible to combine their ideas with those of Sanders to break the logarithmic barrier in Roth's theorem as well, thereby proving the first non-trivial case of Erdős's sum-of-reciprocals conjecture, but unfortunately good reasons emerged to suppose that this cannot be done without significant new ideas. However, the fact remains that the logarithmic barrier is not the right bound for the \mathbb{F}_3^n version of the problem, which makes it hard to think of a good reason for its being the right bound for Roth's theorem itself.

The second recent result, also from 2011, makes it look as though a Behrend-type bound might be correct. Roth's theorem can be thought of as a search for solutions to the equation $x + z = 2y$. Schoen and Shkredov, building on the methods that Sanders introduced to prove his near-logarithmic bound for Roth's theorem, showed that if we generalize this equation, then we can obtain a much better bound [35].

Theorem 2.10. *Let A be a subset of $\{1, 2, \dots, n\}$ of density*

$$\exp(-c(\log n)^{1/6-\varepsilon}).$$

Then A contains distinct elements x_1, x_2, x_3, x_4, x_5 and y such that $x_1 + x_2 + x_3 + x_4 + x_5 = 5y$.

Note that the Behrend lower bound is easily adapted to this equation (since if x_1, \dots, x_5, y are distinct and satisfy that equation then they cannot all lie on the surface of a sphere), so this result is within spitting distance of best possible.

Of course, one could state an Erdős-like corollary to this theorem: if A is a set of integers such that $\sum_{n \in A} n^{-1}$ diverges, then A contains a non-degenerate solution to the equation $x_1 + x_2 + x_3 + x_4 + x_5 = 5y$. However, the original result is more natural.

The result of Schoen and Shkredov is by no means conclusive evidence that the correct bound for Roth's theorem is of the form $\exp(-(\log n)^c)$, since convolutions of three or more functions are significantly smoother than convolutions of two functions, a phenomenon that also explains why the twin-prime conjecture and Goldbach's conjecture are much harder than Vinogradov's three-primes theorem. However, one can at least say, in the light of this result and the result of Bateman and Katz, that there is

a significant chance that the logarithmic barrier for Roth's theorem will eventually be surpassed and the first non-trivial case of Erdős's conjecture proved.

2.2.3. Arithmetic progressions of length 4 or more. What happens for longer progressions? As mentioned earlier, the bounds coming from Szemerédi's proof are very weak. Furstenberg's proof was infinitary and gave no bound at all (though a discrete version of his argument was later found by Tao [39], which in principle gave a weak quantitative bound). The first argument to give a "reasonable" bound was the one in [16, 17], where the following theorem was proved.

Theorem 2.11. *Let A be a subset of $\{1, 2, \dots, n\}$ of density at least $C(\log \log n)^{-1/2^{k+9}}$. Then A contains an arithmetic progression of length k .*

Green and Tao subsequently improved the bound for $k = 4$ to $\exp(-c\sqrt{\log \log n})$ [20]. And that is the current state of the art, though for a finite-field analogue of the problem (again with $k = 4$) they have a bound of the form $\exp(-(\log n)^c)$ [21].

Will Erdős's sum-of-reciprocals conjecture be proved any time soon? There seems at least a fair chance that the case $k = 3$ will be established within, say, the next ten years. There are significant extra difficulties involved when the progressions are longer, but a significant amount of technology for dealing with longer progressions has now been developed. Whether a bound for $k = 3$ will lead to a bound for longer progressions probably depends a lot on what the proof for $k = 3$ looks like, and by how much it beats the logarithmic bound. It may also depend on whether the inverse theorem for uniformity norms can be proved with good quantitative bounds.

3. ERDŐS'S DISCREPANCY PROBLEM

Let us now turn to Conjecture 1.2. Discrepancy problems are problems that ask how "balanced" a colouring of a set can be with respect to some class of subsets. If we have a red/blue colouring κ of a set X and $A \subset X$, then define the discrepancy $\text{disc}(\kappa, A)$ of κ on A to be the difference between the number of red elements of A and the number of blue elements of A . The discrepancy $\text{disc}(\kappa, \mathcal{A})$ of κ with respect to \mathcal{A} is then $\max_{A \in \mathcal{A}} \text{disc}(\kappa, A)$. The discrepancy problem for \mathcal{A} is the problem of determining the minimum of $\text{disc}(\kappa, \mathcal{A})$ over all 2-colourings κ . We can of course think of κ as a function from X to $\{-1, 1\}$ and then $\text{disc}(\kappa, A)$ is $|\sum_{x \in A} \kappa(x)|$. The Erdős discrepancy

problem is the discrepancy problem for the set \mathcal{A} of *homogeneous arithmetic progressions*: that is arithmetic progressions of the form $(d, 2d, 3d, \dots, md)$.

3.1. Known bounds

As with Szemerédi's theorem, it is tempting to conjecture, again wrongly, that random examples are best for this problem. If we choose a random sequence (ε_i) of 1s and -1s, then the expected size of $\sum_{m=1}^n \varepsilon_{md}$ is around \sqrt{n} , and occasionally the size will be slightly bigger by a logarithmic factor.

A simple example that gives rise to much slower growth of these sums is the following, observed by Borwein, Choi and Coons [5]. Every positive integer m can be written in a unique way as $(3a \pm 1)3^b$ for integers a and b . We let $\varepsilon_m = 1$ if m is of the form $(3a + 1)3^b$ and -1 if m is of the form $(3a - 1)3^b$. Note that this function is *completely multiplicative*: $\varepsilon_m \varepsilon_n = \varepsilon_{mn}$ for any two positive integers m and n . Therefore, $|\sum_{m=1}^n \varepsilon_{md}| = |\varepsilon_d \sum_{m=1}^n \varepsilon_m| = |\sum_{m=1}^n \varepsilon_m|$ for any n and d , so analysing the example reduces to calculating the rate of growth of the partial sums of the sequence.

To do this, we partition the integers from 1 to n according to the highest power of 3 that divides them. Let $A_{b,n}$ be the set of multiples of 3^b that are at most n and are not multiples of 3^{b+1} . Then $\sum_{m \in A_{b,n}} \varepsilon_m = 1$ if in the ternary representation of n the digit corresponding to multiples of 3^b is 1, and 0 otherwise. It follows that $\sum_{m=1}^n \varepsilon_m$ is equal to the number of ternary digits of n that are equal to 1. In particular, it has magnitude at most $\log_3 n$, which is far smaller than \sqrt{n} .

In the light of that example, it is natural to investigate the following weakening of Erdős's discrepancy conjecture, which Erdős also asked.

Conjecture 3.1. *Let $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots$ be a completely multiplicative sequence taking values in the set $\{-1, 1\}$. Then the partial sums $\sum_{m=1}^n \varepsilon_m$ are unbounded.*

Remarkably, this conjecture is also very much open. Later we shall discuss evidence that it may be more or less as hard as the discrepancy problem itself.

What about the other direction? The sequence $(1, -1, -1, 1, -1, 1, 1, -1, -1, 1, 1)$ has length 11 and has discrepancy 1 (where by "discrepancy" we mean discrepancy with respect to the set of all homogeneous arithmetic progressions). This turns out to be the longest such sequence [25]. Surprisingly, the longest sequence with discrepancy 2 is *much* longer: there are a very large number of sequences of length 1124 with discrepancy 2, and it

appears that this is the longest that such a sequence can be, though this has not yet been definitively proved. These experimental results, and almost all of the observations that follow, were discovered by the participants in Polymath5, an online collaboration that attacked the Erdős discrepancy problem in 2010 [29]. The fact that these sequences are so long may be one reason that the problem is so hard: it is difficult to imagine what a proof would be like that shows that the discrepancy of a ± 1 sequence tends to infinity with the length of the sequence, while failing to prove the false result that the discrepancy of a sequence of length 1000 is at least 3.

That is not the only reason for the problem's being hard. Another reason is that it is not easy to turn the problem into an analytic one – a technique that is extremely helpful for many other problems. It would be very nice if the result were true not because the sequence consists of 1s and -1 s but merely because it is large in some appropriate sense: for example, perhaps any sequence with values in $[-1, 1]$ such that the average magnitude of the terms is non-zero could be expanded in terms of some cleverly chosen orthonormal basis, and perhaps this would prove that its discrepancy was unbounded. But a very simple example appears to kill off this hope straight away: the discrepancy of the periodic sequence $1, -1, 0, 1, -1, 0, \dots$ is 1, and yet the average magnitude of its terms is $2/3$. Later we shall see that this example is not quite as problematic as it at first appears. Note that this example is a Dirichlet character: it is intriguing that the “difficult” examples we know of all seem to be built out of characters in simple ways.

3.2. Variants of the conjecture

Sometimes, a good way of solving a problem is to replace the statement you are trying to prove by something stronger. There are several promising strengthenings of the Erdős discrepancy conjecture. An obvious one is to replace ± 1 -valued sequences by sequences that take values in some more general set. The example presented shows that we have to be a little careful about this, but the following conjecture is a reasonable one, and is also open.

Conjecture 3.2. *Let x_1, x_2, \dots be a sequence of unit vectors in a (real or complex) Hilbert space. Then for every C there exist n, d such that $\|\sum_{m=1}^n x_{md}\| \geq C$.*

Since \mathbb{R} is a Hilbert space, this conjecture is a generalization of Erdős's conjecture. A conjecture intermediate between the two is one where the x_i are complex numbers of modulus 1.

A less obvious strengthening was formulated by Gil Kalai (one of the Polymath5 participants), and called the “modular version” of the Erdős discrepancy problem.

Conjecture 3.3. *For every prime p there exists N such that if x_1, x_2, \dots, x_N is any sequence of non-zero elements of $\mathbb{Z}/p\mathbb{Z}$, then for every $r \in \mathbb{Z}/p\mathbb{Z}$ there exist n and d with $nd \leq N$ and $\sum_{m=1}^n x_{md} \equiv r \pmod{p}$.*

If we insist that each x_i is $\pm 1 \pmod{p}$, then the conjecture becomes obviously equivalent to the original Erdős problem. However, since the problem does not involve products of the x_i , there is nothing special about the numbers ± 1 , so in this context it becomes natural to replace the set $\{-1, 1\}$ by the set of all non-zero elements. The motivation for this conjecture was the hope that the polynomial method might be applicable to it. So far this has not succeeded, but the modular version gives us a valuable new angle on the problem.

A possible generalization of the modular version to composite moduli m would be to ask that the x_i are coprime to m (which is obviously a necessary condition if we want to be able to produce all numbers r). For amusement only, we state another conjecture here. It is similar in spirit to the more general modular version, but not quite the same.

Conjecture 3.4. *Let K be a finite set of irrational numbers and let x_1, x_2, \dots be a sequence of elements of K . Then the sums $s_{n,d} = \sum_{m=1}^n x_{md}$ are dense mod 1.*

Note that the special case where K is of the form $\{\alpha, -\alpha\}$ for an irrational number α is equivalent to the original discrepancy conjecture. It is not clear whether there are any logical relationships between Conjectures 3.3 and 3.4.

3.3. Some approaches to the conjecture

Although the Erdős discrepancy problem looks very hard, there are some approaches that at least enable one to start thinking seriously about it. Here we discuss three of these approaches.

3.3.1. Completely multiplicative sequences. A close look at the very long sequences of discrepancy 2 that were produced experimentally reveals interesting multiplicative structure. The sequences are not completely multiplicative, but they appear to “want” to have multiplicative features. For example, if you look at the values of a completely multiplicative ± 1 sequence along a geometric progression, then they will either be constant or alternating. In the long sequences of discrepancy 2 we do not see that behaviour,

but we do see quasiperiodic behaviour, at least for a while: towards the end, the patterns break down. There is a natural, but speculative, interpretation of this. The sequences appear to be some kind of “projection” to the set of ± 1 sequences of highly structured sequences taking values in \mathbb{C} . Towards the end, if the structure is followed too closely, the discrepancy rises to 3, but for a while that can be countered by simply switching the signs of a few terms in the sequence. If those terms correspond to integers with not many factors, then not many homogeneous progressions are affected, so one can extend the length of the sequence by sacrificing the structure. But since it was the structure that allowed the sequence to get long in the first place, this process is eventually doomed: one has to make more and more ad hoc tweaks, and eventually it becomes impossible to continue.

This picture suggests the following line of attack. Perhaps one could attempt to show that the worst examples – that is, the ones with lowest discrepancy – have to have some kind of multiplicative structure. Then one could attempt to prove the easier (one hopes) statement that a sequence with multiplicative structure must have unbounded discrepancy.

An approach like this might seem a bit fanciful. Remarkably, however, there is a precise reduction from the Erdős discrepancy problem to a related problem about multiplicative sequences, discovered by Terence Tao (another Polymath5 participant). With the help of a few lines of Fourier analysis, he proved the following result [30].

Proposition 3.5. *Suppose that there exists an infinite ± 1 sequence of discrepancy at most C . Then there exists a completely multiplicative sequence z_1, z_2, \dots of complex numbers of modulus 1 such that the averages $N^{-1} \sum_{n=1}^N |\sum_{i=1}^n z_i|^2$ are bounded above by a constant depending on C .*

Thus, to prove the Erdős discrepancy problem, it is enough to prove the following conjecture about completely multiplicative complex-valued sequences.

Conjecture 3.6. *There exists a function $\omega : \mathbb{N} \rightarrow \mathbb{R}$ tending to infinity with the following property. Let z_1, z_2, \dots be any completely multiplicative sequence z_1, z_2, \dots of complex numbers of modulus 1. For each n let s_n be the n th partial sum of this sequence. Then $(|s_1|^2 + \dots + |s_N|^2)/N \geq \omega(N)$ for every N .*

This is not quite the same as saying that every completely multiplicative sequence has unbounded discrepancy, even if we generalize to the complex case. What it says is not just that the *worst* partial sums of such a sequence should be large, but that the *average* partial sums should be large (uniformly

over all such sequences). However, if the weaker statement is true, then it looks likely that the stronger statement will be true as well.

A pessimistic view of this reduction would be to say that it shows that the multiplicative problem is probably just as hard as the original. However, completely multiplicative sequences have so much more structure than arbitrary sequences that it is not clear that such pessimism is justified.

3.3.2. Semidefinite programming. The following very nice observation was made by Moses Charikar (yet another Polymath5 participant), which offers a way round the obstacle that the sequence $1, -1, 0, 1, -1, 0, \dots$ has bounded discrepancy.

Proposition 3.7. *Suppose that we can find non-negative coefficients $c_{m,d}$ for each pair of natural numbers m and d , and a sequence (b_n) such that $\sum_{m,d} c_{m,d} = 1$, $\sum_n b_n = \infty$, and the real quadratic form*

$$\sum_{m,d} c_{m,d} (x_d + x_{2d} + \dots + x_{md})^2 - \sum_n b_n x_n^2$$

is positive semidefinite. Then every ± 1 sequence has unbounded discrepancy.

Proof. If (ε_n) is a ± 1 sequence, then the positive semidefiniteness of the quadratic form tells us that

$$\sum_{m,d} c_{m,d} (\varepsilon_d + \varepsilon_{2d} + \dots + \varepsilon_{md})^2 \geq \sum_n b_n \varepsilon_n^2 = \sum_n b_n$$

Since $\sum_{m,d} c_{m,d} = 1$ and $\sum_n b_n = \infty$, it follows that the sums $\varepsilon_d + \dots + \varepsilon_{md}$ are unbounded. ■

The same argument shows that if $\sum_n b_n = C$ then there exist m, d such that $|\varepsilon_d + \varepsilon_{2d} + \dots + \varepsilon_{md}| \geq C^{1/2}$. It also proves the Hilbert-space version of the Erdős discrepancy conjecture, since if the x_i are vectors in a Hilbert space, then the non-negative definiteness of the quadratic form implies that

$$\sum_{m,d} c_{m,d} \|x_d + x_{2d} + \dots + x_{md}\|^2 - \sum_n b_n \|x_n\|^2$$

is non-negative (as can be seen by expanding out the norms and looking at each coordinate).

Less obviously, the existence of a quadratic form satisfying the conditions of Proposition 3.7 is actually *equivalent* to a positive solution to the Hilbert-space version of the conjecture.

Proposition 3.8. *Suppose that every infinite sequence of unit vectors in a real Hilbert space has unbounded discrepancy. Then for every C there exists N , a set of non-negative coefficients $c_{m,d}$ for each pair of natural numbers m and d with $md \leq N$, and a sequence (b_1, \dots, b_N) such that $\sum_{m,d} c_{m,d} = 1$, $\sum_{n=1}^N b_n \geq C$, and the real quadratic form*

$$\sum_{m,d} c_{m,d} (x_d + x_{2d} + \dots + x_{md})^2 - \sum_n b_n x_n^2$$

is positive semidefinite.

Proof. For each m, d with $md \leq N$ define $A_{m,d}$ to be the $N \times N$ matrix with ij th entry equal to 1 if both i and j belong to the arithmetic progression $\{d, 2d, \dots, md\}$ and 0 otherwise. Then the conclusion tells us that there exists an $N \times N$ diagonal matrix with entries adding up to at least C that can be written as a convex combination of the matrices $A_{m,d}$ minus a positive semidefinite matrix. If this cannot be done, then by the Hahn-Banach separation theorem there must be a functional that separates the convex set of diagonal matrices with entries adding up to at least C from the convex set consisting of convex combinations of the $A_{m,d}$ minus positive semidefinite matrices. Let us regard this functional as an $N \times N$ matrix B in the inner product space that consists of all $N \times N$ matrices with square-summable entries and the obvious inner product.

What properties must this matrix B have? We may suppose that $\langle D, B \rangle \geq 1$ for every diagonal matrix with entries adding up to at least C and $\langle A, B \rangle \leq 1$ whenever A is a convex combination of the matrices $A_{m,d}$ minus a positive semidefinite matrix. The first condition implies that B is constant on the diagonal and that the constant is at least C^{-1} .

The second condition implies that B has non-negative inner product with every positive semidefinite matrix, since if A were a counterexample, then we could make $\langle -\lambda A, B \rangle$ arbitrarily large and positive by taking λ sufficiently large and positive. In particular, if $x \in \mathbb{R}^N$ and we take A to be the positive semidefinite matrix $x \otimes x$ (that is, the matrix with ij th element $x_i x_j$), then $\langle x, Bx \rangle = \langle x \otimes x, B \rangle \geq 0$, so B is itself positive semidefinite. This is well known to be equivalent to the assertion that there are vectors v_1, \dots, v_N in an inner product space such that $B_{ij} = \langle v_i, v_j \rangle$ for every i, j . Since $B_{ii} = c \geq C^{-1}$ for every i , we find that each vector v_i has norm \sqrt{c} .

Finally, since the zero matrix is positive semidefinite, the second condition also implies that B must have inner product at most 1 with each $A_{m,d}$. In terms of the vectors v_i , this is precisely the statement that $\|v_d + v_{2d} + \dots + v_{md}\|^2 \leq 1$, as can be seen by expanding the left-hand side.

If we now rescale so that the v_i become unit vectors, this last inequality changes to $\|v_d + v_{2d} + \dots + v_{md}\|^2 \leq K$, for some constant $K \leq C$.

Therefore, if the conclusion fails for some constant C , we can find, for each N a sequence of N unit vectors of discrepancy at most \sqrt{C} . After applying a suitable rotation, we may assume that for each n the n th vector in this sequence is spanned by the first n standard basis vectors of \mathbb{R}^N . Therefore, an easy compactness argument gives us an infinite sequence of unit vectors with discrepancy at most \sqrt{C} , a contradiction. ■

Recall that the problem with the sequence $1, -1, 0, 1, -1, 0, \dots$ is that it is “large” in a natural sense (namely having average magnitude bounded away from zero), but has bounded discrepancy. What Proposition 3.7 tells us is that there is a chance of proving that every sequence that is large with respect to a suitable *weighted* norm – the weighted ℓ_2 -norm with weights b_n – has unbounded discrepancy. Thus, there is after all a way of making the problem analytic rather than purely combinatorial.

What can we say about a set of weights that would work? The lesson of the troublesome $1, -1, 0, 1, -1, 0, \dots$ example is that the weights should be concentrated on numbers with many factors. For example, if the sum of the b_n over all non-multiples of 3 is infinite, then the weights cannot work, since then if (x_n) is the troublesome sequence, we have $\sum_n b_n x_n^2 = \infty$ and yet the discrepancy is finite. (This does not contradict Proposition 3.7: it just means that for this choice of (b_n) we cannot find appropriate coefficients $c_{m,d}$.)

It is not easy to write down a set of weights that has any chance of working – in fact, that is worth stating as an open problem – albeit not a wholly precise one.

Problem 3.9. Find a system of weights (b_n) with $\sum_n b_n = \infty$ for which it is reasonable to conjecture that every sequence (x_n) such that $\sum_n b_n x_n^2 = \infty$ has unbounded discrepancy.

One of the things that makes Proposition 3.7 interesting is that it suggests an experimental line of attack on the Erdős discrepancy problem. First, one uses semidefinite programming to determine, for some large N , the sequence (b_1, b_2, \dots, b_N) with largest sum such that the diagonal matrix with those weights can be written as a convex combination of the matrices $A_{m,d}$ minus a positive semidefinite matrix. Next, one stares hard at the sequence and tries to spot enough patterns in it to make a guess at an infinite sequence that would work. Finally, one attempts to decompose the corresponding infinite diagonal matrix (perhaps using the experimental values of the coefficients $c_{m,d}$ as a guide).

Some efforts were made by Polymath5 participants in this direction, but so far they have not succeeded. One problem is that cutting off sharply at N appears to introduce misleading “edge-effects”. But even if one finds ways of smoothing the cutoff, the experimental data is hard to interpret, though it certainly confirms the principle that the weights b_n should be concentrated on positive integers n with many factors. Another serious difficulty is that because we already know that there are very long sequences with small discrepancy, the matrices we find experimentally will have to be extremely large if they are to give us non-trivial lower bounds for discrepancy – large enough that the semidefinite programming algorithms take a long time to run. Despite these difficulties, this still seems like a promising approach that should be explored further.

3.3.3. Representing diagonal matrices. We end by mentioning an approach based on an observation that is somewhat similar to Proposition 3.7 but that does not involve the slightly tricky concept of positive semidefiniteness. This approach was again one of the fruits of the Polymath5 discussion.

Let us define a *HAP matrix* to be a matrix A of the following form. Take two homogeneous arithmetic progressions P and Q and define A_{ij} to be 1 if $i \in P$ and $j \in Q$ and 0 otherwise. In other words, a HAP matrix is the characteristic function of a product of two homogeneous arithmetic progressions.

Proposition 3.10. *Suppose that there exists an $N \times N$ diagonal matrix of trace at least C that belongs to the symmetric convex hull of all HAP matrices. Then every ± 1 sequence of length N has discrepancy at least \sqrt{C} .*

Proof. Let the diagonal matrix D have diagonal entries b_1, \dots, b_N and suppose that it can be written as $\sum_i \lambda_i A_i$ with $\sum_i |\lambda_i| \leq 1$ and with each A_i a HAP matrix. Let $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)$ be a ± 1 sequence. Then

$$C \leq \sum_n b_n \varepsilon_n^2 = \langle \varepsilon, D\varepsilon \rangle = \sum_i \lambda_i \langle \varepsilon, A_i \varepsilon \rangle.$$

It follows that there exists i such that $|\langle \varepsilon, A_i \varepsilon \rangle| \geq C$. If P and Q are the HAPs from which A_i is built, then

$$\langle \varepsilon, A_i \varepsilon \rangle = \left(\sum_{i \in P} \varepsilon_i \right) \left(\sum_{j \in Q} \varepsilon_j \right),$$

which implies that at least one of $\sum_{i \in P} \varepsilon_i$ and $\sum_{j \in Q} \varepsilon_j$ has modulus at least \sqrt{C} . ■

Once again, the argument generalizes easily to unit vectors in a Hilbert space. And again there is an implication in the other direction.

Proposition 3.11. *Let C be a constant, let N be a positive integer, and suppose that for every $N \times N$ real matrix $A = (a_{ij})$ with 1s on the diagonal there exist homogeneous arithmetic progressions P and Q such that $|\sum_{i \in P} \sum_{j \in Q} a_{ij}| \geq C$. Then there is a diagonal matrix of trace at least C that belongs to the symmetric convex hull of all HAP matrices.*

Proof. Again we use the Hahn-Banach theorem. If no such diagonal matrix exists, then there is a linear functional, which we can represent as taking the inner product with a matrix A , that separates diagonal matrices of trace at least C from convex combinations of HAP matrices and minus HAP matrices. If $\langle D, A \rangle \geq 1$ for every diagonal matrix D of trace at least C , then A must be constant on the diagonal and the constant must be at least C^{-1} . And if $|\langle B, A \rangle| < 1$ for every HAP matrix B , then for any two homogeneous arithmetic progressions P and Q we have $|\sum_{i \in P} \sum_{j \in Q} a_{ij}| < 1$. And now if we choose λ such that λA has 1s along the diagonal, then the matrix λA contradicts our hypothesis. ■

In the light of this proposition (which is easily seen to be an equivalence) it is natural to make the following conjecture, which is yet another strengthening of the Erdős discrepancy problem.

Conjecture 3.12. *For every C there exists N such that if $A = (a_{ij})$ is any real $N \times N$ matrix with 1s on the diagonal, then there exist homogeneous arithmetic progressions P and Q such that $|\sum_{i \in P} \sum_{j \in Q} a_{ij}| \geq C$.*

If we apply that conjecture in the case where $a_{ij} = \varepsilon_i \varepsilon_j$ for some ± 1 sequence $(\varepsilon_1, \dots, \varepsilon_N)$, then the conclusion is that $|\sum_{i \in P} \varepsilon_i \sum_{j \in Q} \varepsilon_j| \geq C$, from which it follows that the sequence has discrepancy at least \sqrt{C} . Thus, the conjecture really is a strengthening of the Erdős discrepancy conjecture. Indeed, given how much weaker the condition of having 1s on the diagonal is than the condition of being a tensor product of two ± 1 sequences, it is a very considerable strengthening. And yet it still appears to have a good chance of being true.

4. CONCLUSION

The aim of this paper has been to give some idea of what is currently known about two notable conjectures of Erdős concerning arithmetic progressions.

It has therefore been more about questions than answers, but Erdős would have been the last person to mind that. I imagine him sitting with “the book” open at the relevant page, smiling at us as we struggle to find the proofs that he is now able to enjoy.

REFERENCES

- [1] F. A. Behrend, *On sets of integers which contain no three in arithmetic progression*, Proc. Nat. Acad. Sci. **23** (1946), 331–332.
- [2] V. Bergelson, A. Leibman, *Polynomial extensions of van der Waerden’s and Szemerédi’s theorems*, J. Amer. Math. Soc. **9** (1996), 725–753.
- [3] M. Bateman, N. H. Katz, *New bounds on cap sets*, J. Amer. Math. Soc. **25** (2012), 585–613.
- [4] V. Bergelson, T. C. Tao, T. Ziegler, *An inverse theorem for the uniformity seminorms associated with the action of F^ω* , Geom. Funct. Anal. **19** (2010), 1539–1596.
- [5] P. Borwein, K.-K. S. Choi, M. Coons, *Completely multiplicative functions taking values in $\{-1, 1\}$* , Trans. Amer. Math. Soc. **362** (2010), 6279–6291.
- [6] J. Bourgain, *On triples in arithmetic progression*, Geom. Funct. Anal. **9** (1999), 968–984.
- [7] J. Bourgain, *Roth’s theorem on progressions revisited*, J. Anal. Math. **104** (2008), 155–192.
- [8] P. Dodos, V. Kanellopoulos, K. Tyros, *A simple proof of the density Hales–Jewett theorem*, <http://arxiv.org/abs/1209.4986>
- [9] G. Elek, B. Szegedy, *A measure-theoretic approach to the theory of dense hypergraphs*, Adv. Math., **231** (2012), 1731–1772.
- [10] P. Erdős, P. Turán, *On some sequences of integers*, J. London Math. Soc. **11** (1936), 261–264.
- [11] H. Furstenberg, *Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.
- [12] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291.
- [13] H. Furstenberg, Y. Katznelson, *A density version of the Hales–Jewett theorem*, J. Analyse Math. **57** (1991), 64–119.
- [14] P. Frankl, V. Rödl, *The uniformity lemma for hypergraphs*, Graphs and Combinatorics **8** (1992), 309–312.
- [15] P. Frankl, V. Rödl, *Extremal problems on set systems*, Rand. Struct. Alg. **20** (2002), 131–164.
- [16] W. T. Gowers, *A new proof of Szemerédi’s theorem for arithmetic progressions of length four*, Geom. Funct. Anal. **7** (1997), 322–337.
- [17] W. T. Gowers, *A new proof of Szemerédi’s theorem*, Geom. Funct. Anal. **11** (2001), 465–588.

- [18] W. T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Annals of Math. **166** (2007), 897–946.
- [19] B. J. Green, T. C. Tao, *The primes contain arbitrarily long arithmetic progressions*, Annals of Math. **167** (2008), 481–547.
- [20] B. J. Green, T. C. Tao, *An inverse theorem for the Gowers U^3 -norm, with applications*, Proc. Edinburgh Math. Soc. **51** (2008), 71–153.
- [21] B. J. Green, T. C. Tao, *New bounds for Szemerédi’s theorem, Ia: progressions of length 4 in finite field geometries revisited*, <http://arxiv.org/abs/1205.1330>
- [22] B. J. Green, T. C. Tao, T. Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$ -norm*, Annals of Math. **176** (2012), 1231–1372.
- [23] D. R. Heath-Brown, *Integer sets containing no arithmetic progressions*, J. London Math. Soc. (2) **35** (1987), 385–394.
- [24] L. Lovász, B. Szegedy, *Szemerédi’s Lemma for the analyst*, Geom. Funct. Anal. **17** (2007), 252–270.
- [25] A. R. D. Mathias, *On a conjecture of Erdős and Čudakov*, in Combinatorics, Geometry and Probability: A tribute to Paul Erdős, P. Erdős, B. Bollobás and A. Thomason eds., Cambridge 1997, pp. 489–492.
- [26] R. Meshulam, *On subsets of finite abelian groups with no 3-term arithmetic progressions*, J. Combin. Theory Ser. A **71** (1995), 168–172.
- [27] B. Nagle, V. Rödl, M. Schacht, *The counting lemma for regular k -uniform hypergraphs*, Rand. Struct. Alg. **28** (2006), 113–179.
- [28] D. H. J. Polymath, *A new proof of the density Hales–Jewett theorem*, Annals Math. **175** (2012), 1283–1327.
- [29] Polymath5, Wiki containing observations and experimental results concerning the Erdős discrepancy problem: http://michaelnielsen.org/polymath1/index.php?title=The_Erdős_discrepancy_problem
- [30] Polymath5, Fourier reduction of Erdős discrepancy problem to a problem about completely multiplicative sequences, http://michaelnielsen.org/polymath1/index.php?title=Fourier_reduction
- [31] K. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 245–252.
- [32] I. Z. Ruzsa, E. Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, 939–945.
- [33] T. Sanders, *On certain other sets of integers*, <http://arxiv.org/abs/1007.5444>
- [34] T. Sanders, *On Roth’s theorem on progressions*, Annals of Math. **174** (2011), 619–636.
- [35] T. Schoen, I. D. Shkredov, *Roth’s theorem in many variables*, <http://arxiv.org/abs/1106.1601>
- [36] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104.
- [37] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 299–345.

- [38] E. Szemerédi, *Integer sets containing no arithmetic progressions*, Acta Math. Hungar. **56** (1990), 155–158.
- [39] T. C. Tao, *A quantitative ergodic theory proof of Szemerédi's theorem*, Electron. J. Combin. **13** (2006), Research Paper 99, 49pp.

W. Timothy Gowers

*Centre for Mathematical Sciences,
Wilberforce Road,
Cambridge,
CB3 0WB UK*

e-mail: `wtg10@dpms.cam.ac.uk`

PAUL ERDŐS AND EGYPTIAN FRACTIONS

RONALD L. GRAHAM

One of Paul Erdős' earliest mathematical interests was the study of so-called *Egyptian fractions*, that is, finite sums of distinct fractions having numerator 1. In this note we survey various results in this subject, many of which were motivated by Erdős' problems and conjectures on such sums. This note complements the excellent treatment of this topic given by A. Schinzel in 2002.¹

1. INTRODUCTION

The Rhind Papyrus of Ahmes [47] (see also [34, 63]) is one of the oldest known mathematical manuscripts, dating from around 1650 B.C. It contains among other things, a list of expansions of fractions of the form $\frac{2}{n}$ into sums of distinct *unit* fractions, that is, fractions with numerator 1. Examples of such expansions are $\frac{2}{35} = \frac{1}{30} + \frac{1}{42}$ and $\frac{2}{63} = \frac{1}{56} + \frac{1}{72}$. More generally, one can consider expansions of more general rational numbers into sums of unit fractions with distinct denominators such as:

$$\frac{10}{73} = \frac{1}{11} + \frac{1}{22} + \frac{1}{1606}, \quad \frac{67}{2012} = \frac{1}{31} + \frac{1}{960} + \frac{1}{2138469} + \frac{1}{10670447077440},$$

and

$$1 = \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \frac{1}{10} + \frac{1}{14} + \frac{1}{15} + \frac{1}{18} + \frac{1}{20} + \frac{1}{24} + \frac{1}{28} + \frac{1}{30}.$$

There are various explanations as to why the Egyptians chose to use such representations (for example, see [63]) but perhaps the most compelling is that given to the author some years ago by the legendary mathematician André Weil [62]. When I asked him why he thought the Egyptians used this

¹See [52].

method for representing fractions, he thought for a moment and then said, "It is easy to explain. *They took a wrong turn!*"

As is well known, Erdős' first major result (and first paper) was his beautiful 1932 proof [25] of Bertrand's postulate, namely that for any positive integer $n > 1$, there is always a prime between n and $2n$.² In particular, Erdős' proof was based in part on an analysis of the prime divisors of the binomial coefficients $\binom{2n}{n}$. What is perhaps less well known is that Erdős' second paper [26], also published in 1932, dealt with Egyptian fractions. In it, he generalizes an elementary result of Kürschák [41] by showing that for any choice of positive integers a, d and n , the sum $\sum_{k=1}^n \frac{1}{a+kd}$ is never an integer.³

The next paper of Erdős dealing with Egyptian fractions was his 1945 paper with I. Niven [29]. In that paper, they showed among other things that no two partial sums of the harmonic series can be equal, i.e., $\sum_{i=r}^s i^{-1} = \sum_{i=t}^u i^{-1}$ implies $r = t$ and $s = u$. In that paper they also showed that for only finitely many n can one or more of the elementary symmetric functions of $1, \frac{1}{2}, \dots, \frac{1}{n}$ be an integer. Very recently, this was strengthened in a paper of Chen and Tang [17]. In that paper, they showed that the only pairs (k, n) for which the k^{th} elementary function $S(k, n)$ of $1, \frac{1}{2}, \dots, \frac{1}{n}$ is an integer is $S(1, 1) = 1$ and $S(2, 3) = (1)(\frac{1}{2}) + (1)(\frac{1}{3}) + (\frac{1}{2})(\frac{1}{3}) = 1$. Thus, for $n \geq 4$, none of the elementary functions are integers.

Perhaps the paper of Erdős dealing with Egyptian fractions which has had the greatest impact was his 1950 paper [27]. In this seminal paper, he considers the quantity $N(a, b)$, defined for integers $1 \leq a < b$ to be least value n such that the equation $\frac{a}{b} = \sum_{k=1}^n \frac{1}{x_k}$ has a solution with $0 < x_1 < x_2 < \dots < x_n$. In particular, he shows that $N(b) = \max_{1 \leq a < b} N(a, b)$ satisfies $\log \log b \ll N(b) \ll \frac{\log b}{\log \log b}$, sharpening an earlier result of deBruijn and others. It is conjectured in [27] that $N(b) \ll \log \log b$. The best result in this direction at present is due to Vose [59] who showed that $N(b) \ll \sqrt{\log b}$.

It is also in this paper that the celebrated Erdős-Straus " $\frac{4}{n}$ conjecture" occurs, namely that $N(4, b) \leq 3$ for every $b > 2$. This will be the subject of the next section.

²This was memorialized by Leo Moser's limerick: "Chebyshev said it and I'll say it again. There is always a prime between n and $2n$."

³Interestingly, Erdős states in the German abstract of that paper: "Der Grundgedanke des Beweises besteht darin, dass ein Glied $a + kd$ angegeben wird, welches durch eine höhere Potenz einer Primzahl teilbar ist, als die übrigen Glieder. Dies ergibt sich aus der Analyse der Primteiler der Ausdrücke von $\frac{(a+d)(a+2d)\dots(a+nd)}{n!}$ und $\binom{2n}{n}$ " (The basic idea of the proof is that some term $a + kd$ is divisible by a higher power of some prime than any other terms. This follows from the analysis of the prime divisors of the expressions $\frac{(a+d)(a+2d)\dots(a+nd)}{n!}$ and $\binom{2n}{n}$).

2. THE ERDŐS-STRAUS CONJECTURE

The first proof that any positive rational $\frac{a}{b}$ has an Egyptian fraction representation:

$$(1) \quad \frac{a}{b} = \frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}, \quad 1 \leq x_1 < x_2 < \dots < x_n,$$

was given by Fibonacci (= Leonardo Pisano) in 1202 [32]. His method was to apply the *greedy algorithm*, namely always subtract the largest possible unit fraction from the current remainder so that the result is nonnegative. While this ordinarily does not produce the shortest possible representation, or the one with smallest maximum denominator, it does terminate in finitely many steps since eventually the numerator of the reduced remainder must strictly decrease at each step. In particular, for fractions of the form $\frac{2}{n}$ for $n > 1$, the greedy algorithm only needs 2 steps, and for $\frac{3}{n}$, it only needs 3 steps. While this algorithm would guarantee that for the fractions $\frac{4}{n}$, a representation with 4 unit fractions is guaranteed, Erdős and Straus [27] conjectured that in fact such a fraction always had an Egyptian fraction expansion with *at most 3* terms. It is easy to see that in order to prove this, it is enough to show that it holds for prime values of n . There have been many papers published studying various aspects of this problem (for example, see [1, 40, 48, 61, 60] and especially the references in [39]). For example, it is known that if the conjecture fails for some value n then n must be congruent to one of $1^2, 11^2, 13^2, 17^2, 19^2$ or $23^2 \pmod{840}$. From a computational perspective, the conjecture has been verified for $n \leq 10^{14}$ [57]. One of the most recent treatments is in a long paper of Elsholtz and Tao [24] (extending earlier work of Elsholtz [23]). Among their many results are the following. Let $f(n)$ denote the number of different solutions to the equation

$$(2) \quad \frac{4}{n} = \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}$$

where here the x_i are *not* assumed to be distinct or ordered by size. It is easy to see that the Erdős-Straus conjecture is that $f(n) > 0$ for $n > 1$. In [24], it is shown that :

- (i) $N \log^2 N \ll \sum_{q \leq N} f(q) \ll N \log^2 N \log \log N$ where q ranges over primes;
- (ii) For any prime q ,

$$f(q) \ll q^{\frac{3}{5} + O\left(\frac{1}{\log \log q}\right)}.$$

(iii) For infinitely many n , one has

$$f(n) \geq \exp\left((\log 3 + o(1))\frac{\log n}{\log \log n}\right).$$

In particular, it follows from this that there are relatively few solutions to (2) for most n . However, Vaughan [58] has shown that the number of $n \leq x$ for which the Erdős-Straus conjecture fails is $O(x \exp(-c(\log x)^{\frac{2}{3}}))$, $c > 0$. As of this writing, the original conjecture of Erdős and Straus is still unresolved.⁴

Motivated by the Erdős-Straus conjecture, Sierpiński [55] made the analogous conjecture⁵ for the fractions $\frac{5}{n}$, namely, that for all $n \geq 5$, there is a decomposition:

$$\frac{5}{n} = \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}, \quad 1 \leq x_1 < x_2 < x_3.$$

This has been verified for $5 \leq n \leq 1057438801$ (see [39]). More generally, Schinzel (also in [55]) conjectured that for any fraction $\frac{a}{n}$, one can express it as:

$$\frac{a}{n} = \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}, \quad 1 \leq x_1 < x_2 < x_3,$$

provided $n > n_0(a)$. Needless to say, these conjectures are currently still unsettled.

⁴As a historical note, this conjecture also occurred around the same time in a paper of Obláth [46] (submitted for publication in 1948) in which the constraint that the x_i be distinct is relaxed.

⁵It is curious why Erdős and Straus didn't make this conjecture in [27] as well.

3. DENSE EGYPTIAN FRACTIONS

In [27], Erdős also considers various questions relating to Egyptian fraction decompositions of $1 = \sum_{k=1}^n \frac{1}{x_k}$. In particular, he conjectures that we must always have $\frac{x_n}{x_1} \geq 3$, with the extreme example coming from the decomposition $1 = \frac{1}{2} + \frac{1}{3} + \frac{1}{6}$. In fact, he suggests that it may even be true that $\lim_{n \rightarrow \infty} \frac{x_n}{x_1} = \infty$. However, it is now known that this is not the case. It follows from the work of Martin [43, 44] and Croot [18, 19] that the following holds.

Theorem 1 [18]. *Suppose that $r > 0$ is a given rational number. Then for all $N > 1$, there exist integers x_1, x_2, \dots, x_k , with*

$$N < x_1 < x_2 < \dots < x_k \leq \left(e^r + O_r \left(\frac{\log \log N}{\log N} \right) \right) N$$

such that

$$r = \frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_k}.$$

Moreover, the error term $O_r \left(\frac{\log \log N}{\log N} \right)$ is best possible.

This result settled one of the many questions raised in Chapter 4 (Unit Fractions) of the booklet [28] of Erdős and the author.

Another question raised in [28] and answered by Martin [44] deals with the quantity $L_j(s)$ defined for a positive rational s by

$$L_j(s) = \left\{ \begin{array}{l} x \in \mathbf{Z}, x > s^{-1} : \text{there do not exist} \\ x_1, \dots, x_t \in \mathbf{Z}, x_1 > \dots > x_t \geq 1 \text{ with } \sum_{i=1}^t \frac{1}{x_i} = s \text{ and } x_j = x \end{array} \right\}.$$

The largest denominator in an Egyptian fraction representation of s can be a prime only if it is a prime divisor of s . Hence the set $L_1(s)$ contains most primes and it is clearly infinite. However, $L_1(s)$ must have zero density as dictated by the following result [44]:

Let $L_1(s; x)$ denote the counting function of $L_1(s)$, i.e.,

$$L_1(s, x) = |\{1 \leq n \leq x : n \in L_1(s)\}|.$$

Then for any rational $s > 0$ and any real $x \geq 3$, we have;

$$\frac{x \log \log x}{\log x} \ll_s L_1(s, x) \ll_s \frac{x \log \log x}{\log x}.$$

However, for $j \geq 2$, the situation is quite different. In fact, for any $j \geq 2$, $L_j(s)$ is finite. In particular, there are only finitely many numbers which cannot be the second-largest denominator in an Egyptian fraction representation of 1. Martin suggests that perhaps the set $\{2, 4\}$ is the complete list (of those greater than 1).

4. MORE PROBLEMS FROM *Old and New Problems and Results* [28]

(Many of the problems and results in this section are taken more or less directly from the above mentioned book. The reader can consult [28] for more details).

It is known that any positive rational $\frac{a}{2b+1}$ can be represented as a finite sum of the form $\sum_k \frac{1}{2q_k+1}$ (e.g., see [3, 9, 56]. An old question of Stein [53] asks if such a decomposition can always be accomplished by the greedy algorithm. In other words, if we start with an arbitrary positive rational $\frac{a}{2b+1}$ and repeatedly subtract the largest unit fraction $\frac{1}{2q+1}$ so that the remainder is nonnegative, must this process always terminate? No examples are known which provably do not terminate, although there are terminating rationals for which the denominators become very large. For example, starting with $\frac{5}{1444613}$, the greedy algorithm takes 37 terms to terminate, with the largest denominator having 384,122,451,172 decimal digits (see [45]). It is known [36] that a positive rational $\frac{a}{b}$ can be expressed as a finite sum of fractions of the form $\frac{1}{pk+q}$ if and only if $\left(\frac{b}{(b,(p,q))}, \frac{p}{(p,q)}\right) = 1$. One could ask here whether the greedy algorithm always terminates for this representation as well. Restricting the denominators even more, the author has shown [37] that a necessary and sufficient condition that a rational $\frac{a}{b}$ can be expressed as

$$\frac{a}{b} = \frac{1}{x_1^2} + \frac{1}{x_2^2} + \dots + \frac{1}{x_k^2} \quad \text{for positive integers } 0 < x_1 < x_2 < \dots < x_k,$$

is that

$$\frac{a}{b} \in \left[0, \frac{\pi^2}{6} - 1\right) \cup \left[1, \frac{\pi^2}{6}\right).$$

For example,

$$\frac{1}{2} = \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \frac{1}{15^2} + \frac{1}{16^2} + \frac{1}{36^2} + \frac{1}{60^2} + \frac{1}{180^2}.$$

I believe that it would be a very rare event for the greedy algorithm to succeed in this situation!⁶

In this vein, a number of questions were raised by Wilf [64] concerning what he called “reciprocal bases for the integers”. By this he meant sets S of integers so that every positive integer can be represented as a finite sum of reciprocals of integers taken from S . For example, he asked: “Is every infinite arithmetic progression a reciprocal basis?” (Yes, by [3, 36]); “Must a reciprocal basis have positive density?” (No, by [3, 36]).

More generally, one could define a reciprocal basis for the *rational*s to be a set S of positive integers so that every positive rational $\frac{p}{q}$ is a finite sum of reciprocals of elements in S . At present, we don’t know necessary and sufficient conditions for a set to be a reciprocal basis for the integers or the rational^s. However, a general theorem in this direction is the following.

For a set $T = \{t_1, t_2, \dots\}$ of positive integers, define $P(T)$ to be the set of all finite sums of elements taken from T . Also, define $T^{-1} = \{\frac{1}{t_i} : t_i \in T\}$. We will say that T is *complete* if every sufficiently large integer belongs to $P(T)$. Further, define $M(T)$ to be the set of all products $t_{i_1}t_{i_2} \dots t_{i_r}$ where $1 \leq i_1 < i_2 < \dots < i_r$ with $r = 1, 2, \dots$. Finally, let us say that a real number α is *T-accessible* if for all $\varepsilon > 0$, there is a $u \in T$ such that $0 \leq u - \alpha < \varepsilon$. In [36], the following result is proved.

Theorem 2. *Suppose $S = (s_1, s_2, \dots)$ is a sequence of positive integers so that $M(S)$ is complete and $\frac{s_{n+1}}{s_n}$ is bounded as $n \rightarrow \infty$.*

Then $\frac{p}{q} \in P(M(S))^{-1}$ (with $(p, q) = 1$) if and only if $\frac{p}{q}$ is $M(S)^{-1}$ -accessible and q divides some element of $M(S)$.

It follows from this, for example, the set consisting of the primes together with the squares forms a reciprocal basis for the rational^s. It is not known whether the condition that $\frac{s_{n+1}}{s_n}$ be bounded is needed for the conclusion of the theorem to hold.

A classical result of Curtiss [22] asserts that the closest strict under approximation R_n of 1 by a sum of n unit fractions is always given by taking $R_n = \sum_{k=1}^n \frac{1}{u_k+1}$, where u_n is defined recursively by: $u_1 = 1$, and $u_{n+1} = u_n(u_n + 1)$ for $n \geq 1$. The analogous fact is also known to hold [27]

⁶For similar results using n^{th} powers rather than squares, see [37].

⁷In fact, I don’t know of any good conjectures here.

for rationals of the form $\frac{1}{m}$. However, it does not hold for some rationals, e.g., $R_1\left(\frac{11}{24}\right) = \frac{1}{3}$ while $R_2\left(\frac{11}{24}\right) = \frac{1}{4} + \frac{1}{5}$. Perhaps it is true that for any rational it does hold eventually. In other words, is it true that for any rational $\frac{a}{b}$, the closest strict under approximation $R_n\left(\frac{a}{b}\right)$ of $\frac{a}{b}$ is given by

$$R_n\left(\frac{a}{b}\right) = R_{n-1}\left(\frac{a}{b}\right) + \frac{1}{m}$$

where m is the least denominator not yet used for which $R_n\left(\frac{a}{b}\right) < \frac{a}{b}$ provided that n is sufficiently large? In fact, as we state in [28], this behavior might even hold for all algebraic numbers.

For each n , let \mathbf{X}_n denote the set

$$\left\{ \{x_1, x_2, \dots, x_n\} : \sum_{k=1}^n \frac{1}{x_k} = 1, 0 < x_1 < x_2 < \dots < x_n \right\}$$

and let $\mathbf{X} = \cup_{n \geq 1} \mathbf{X}_n$. There are many attractive unresolved questions concerning these sets which were raised in [28], some of which I will now mention.

To begin, it would be interesting to have asymptotic formulas or even good estimates for $|\mathbf{X}_n|$. To the best of my knowledge, the best estimates currently known [50] are:

$$e^{c \frac{n^3}{\log n}} < |\mathbf{X}_n| < c_0^{(1+\varepsilon)2^{n-1}}$$

where $c_0 = \lim_{n \rightarrow \infty} u_n^{\frac{1}{2^n}} = 1.264085\dots$, with u_n defined as above (see [2]). Perhaps the lower bound can be replaced by $c_0^{2^{n(1-\varepsilon)}}$.

In view of the large number of sets in \mathbf{X} , one would suspect that the condition that the reciprocals of a set of integers sum to 1 is not really a very stringent condition (modulo some obvious modular and size restrictions, e.g., the largest element cannot be prime). For example, it has been shown in [35] that for all $m \geq 78$, there is a set $\{x_1, x_2, \dots, x_t\} \in \mathbf{X}$ with $\sum_{k=1}^t x_k = m$. Furthermore, this is not true for 77 [42]. I would conjecture that this behavior is true much more generally. Namely, it should be true that for any polynomial $p : \mathbf{Z} \rightarrow \mathbf{Z}$, there is a set $\{x_1, x_2, \dots, x_t\} \in \mathbf{X}$ with $\sum_{k=1}^t p(x_k) = m$, for all sufficiently large m , provided p satisfies the obvious necessary conditions:

- (i) The leading coefficient of p is positive;
- (ii) $\gcd(p(1), p(2), \dots) = 1$.

It is known [15] that these conditions are sufficient for expressing every sufficiently large integer as a sum $\sum_{a_i \text{ distinct}} p(a_i)$.

How many integers $x_k < n$ can occur as an element of $\{x_1, x_2, \dots, x_n\} \in \mathbf{X}_n$? Are there $o(n)$, cn or $n - o(n)$?

What is the least integer $v(n) > 1$ which does not occur as an x_k , k variable, for $\{x_1, x_2, \dots, x_n\} \in \mathbf{X}_n$? It is easy to see that $v(n) > cn!$ by results in [6, 7, 8]. It may be that $v(n)$ actually grows more like $2^{2^{\sqrt{n}}}$ or even $2^{2^{n(1-\varepsilon)}}$.

Denote by $k_r(n)$ the least integer which does not occur as x_r in any $\{x_1, x_2, \dots, x_t\} \in \mathbf{X}_n$ with $x_1 < x_2 < \dots < x_t \leq n$. It is not hard to show

$$k_1(n) < \frac{cn}{\log n}.$$

We have no idea of the true value of $k_r(n)$ or even $k_1(n)$.

As a related problem, suppose we define $K(n)$ to be the least integer which does not occur as x_i for any i in any $\{x_1, x_2, \dots, x_t\} \in \mathbf{X}_n$ with $x_1 < x_2 < \dots < x_t \leq n$. Again,

$$K(n) < \frac{cn}{\log n}$$

is easy but at present we do not even know if $k_1(n) < K(n)$.

How many disjoint sets $S_i \in \mathbf{X}$, $1 \leq i \leq k$, can we find so that $S_i \subseteq \{1, 2, \dots, n\}$? As C. Sándor notes [51], applying the results of Theorem 1 iteratively, we should be able to achieve $k = (1 + o(1)) \log n$. More generally, how many disjoint sets $T_i \subseteq \{1, 2, \dots, n\}$ are there so that all the sums $\sum_{t \in T_i} \frac{1}{t}$ are equal. By using strong Δ -systems [30], it can be shown that there are at least $\frac{n}{e^{c\sqrt{\log n}}}$ such T_i . Is this the right order of magnitude? One could also ask how many disjoint sets $\{x_1, x_2, \dots, x_n\} \in \mathbf{X}_n$ are possible. It is probably true that there are only $o(\log n)$ such sets.

Another set of attractive questions concerns what might be called *Ramsey* properties of the \mathbf{X}_n . It was asked in [28] whether for any partition of $\{2, 3, 4, \dots\}$ into finitely many blocks, some block must contain an element of \mathbf{X} . Put another way, is it true that if the integers greater than 1 are arbitrarily r -colored, then at least one of the color classes contains a finite set of integers whose reciprocals sum to 1? Erdős and I liked this problem so much that we posted a reward \$500 for its solution. As it turned out, the problem was settled in the affirmative by a beautiful argument of Ernie Croot [20].⁸

⁸As it happened, Erdős did not live to see the solution. When I asked Ernie whether he would like a check for the \$500 signed by Erdős, he said he would be pleased to be paid this

A stronger conjecture is that any sequence $x_1 < x_2 < \dots$ of positive upper density contains a subset whose reciprocals sum to 1. Perhaps this can be proved if we assume that the differences $x_{k+1} - x_k$ are bounded. It is not enough to just assume that $\sum_k \frac{1}{x_k}$ is unbounded as the set of primes shows. (The letter in Figure 1 from Erdős' mathematical notebook from 1963 shows our interest in these questions going back some 50 years. In the appendix, we show some additional notes of Erdős on these problems). However, perhaps the sum $\sum_{k=1}^n \frac{1}{x_k}$ cannot grow much faster than this (i.e., $\log \log n$) for the x_k to fail to form some $\bar{x} \in \mathbf{X}$.

Let $A(n)$ denote the largest value of $|S|$ such that $S \subseteq \{1, 2, \dots, n\}$ contains no set in \mathbf{X} . Probably $A(n) = n - o(n)$ but this is not known. A related question is this. What is the smallest set $S' \subseteq \{1, 2, \dots, n\}$ which contains no set in \mathbf{X} and which is maximal in this respect. Very little is known here. More generally, one could ask for the largest subset $S_n^* \subseteq \{1, 2, \dots, n\}$ so that for any distinct elements $s, s_1, s_2, \dots, s_m \in S_n^*$, we have $\frac{1}{s} \neq \sum_{k=1}^m \frac{1}{s_k}$ where $m > 1$? We can certainly have $|S_n^*| > cn$ as the set $\{i : \frac{n}{2} < i < n\}$ shows. Can $|S_n^*| > cn$ for $c > \frac{1}{2}$? Is it true that if $S \subseteq \{1, 2, \dots, n\}$ with $|S| > cn$ then S contains x, y, z with $\frac{1}{x} + \frac{1}{y} = \frac{1}{z}$? It has been shown by Brown and Rödl [10] that the partition version of this question holds, i.e., for any partition of \mathbf{Z} into finitely many classes and for any fixed value of n , one of the classes must contain a solution to $\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} = \frac{1}{z}$.

There are many interesting unresolved questions which involve restricting the denominators of the elements in the \mathbf{S}_n . For example, Burshtein [11] gives an example of $\{x_1, x_2, \dots, x_n\} \in \mathbf{X}_n$ with no x_i dividing any other x_j . Even more striking, Barbeau [5] finds an example in which each x_i is the product of exactly 2 distinct primes. A smaller such example was given by Burshtein [12, 13], The smallest such example known is that of Allan Johnson (see [39]) with the denominators shown in the table below.

way. (I kept a number of checks pre-signed by Erdős for just such contingencies.) After sending Ernie the Erdős check, I subsequently sent Ernie a *real* check for \$500, which he certainly earned. However, unknown to me, Ernie *cash*ed the Erdős check. That is, it was sent to my bank and it was honored. This was unexpected since Erdős never had an account at my bank! I am guessing that the bank tellers were so used to seeing Erdős' checks countersigned by me that they just assumed this was one of those and they cashed it. When I discovered this, I wrote to Ernie that he owed me \$500. He agreed to send back the \$500 overpayment but on the condition that I send him back the canceled Erdős check (which I did).

(Graham) Legyen $a_1 < \dots$, $a_{k+1} - a_k < c$. Igar e hogy $1 = \sum \frac{1}{a_i}$
 megoldható? Ugyanez kidecketo' karsak azt temilk fel, hogy
 $\lim_{k \rightarrow \infty} a_k/k < \infty$.
 63 III 29 (Graham, Kraus) Is kidecs mamol felbonthato' raga'rot
 pl ha $m > m_0$ es a mamokat m -ig k ronse ontul lisonyira
 $\sum x_i = m$ az egyik rontra megoldható, (jó kato' m -re?).
 Ha két ronse ontul akkor $m = x + y$ vagy $m = x + y + z$ is már
 megoldható len. Ha a rat mamokat két ronse ontul akkor minden
 rat mam ~~egy intora~~ előállitható mint egy ontáglól vagy hüllöml.
 mam omese - ugyanez igar len k ronse a valós mamokra is, de
 R_0 ronse nem igar a valós mamokra / legalább is ha $c = \frac{1}{2}$

Fig. 1. A page from Erdős' 1963 notebook

6	21	34	46	58	77	87	114	155	215	287	391
10	22	35	51	62	82	91	119	187	221	299	689
14	26	38	55	65	85	93	123	203	247	319	731
15	33	39	57	69	86	95	133	209	265	323	901

Table 1. Denominators for Johnson's decomposition of 1

However, as Barbeau notes in [4], it is not known if 1 can be represented as the product of two sums of the form $\frac{1}{q_1} + \frac{1}{q_2} + \dots + \frac{1}{q_r}$ where the q_i are distinct primes. Perhaps this can be done if we just assume that the q_i are pairwise relatively prime. (Related results can be found in [33].) In a (still) unfinished manuscript of Erdős and the author⁹, it is shown that any integer can be represented as a sum of reciprocals of distinct numbers which each have exactly three prime factors (see [39]). Whether this can be accomplished with just two prime factors is not clear.

In [54], Shparlinski answers a question of Erdős and the author by proving the following result.

⁹I'm still working on it!

Theorem 3. For any $\varepsilon > 0$ there is a $k(\varepsilon)$ such that for any prime p and any integer c there exist $k \leq k(\varepsilon)$ pairwise distinct integers x_i with $1 \leq x_i \leq p^\varepsilon$, and such that

$$\sum_{i=1}^k \frac{1}{x_i} \equiv c \pmod{p}.$$

(Here, the reciprocals are taken modulo p). This has been generalized by Croot [21] to the case when the denominators are all of the form x_i^k for a general positive integer k .

5. THE STORY OF AN INCORRECT CONJECTURE

Naturally, not every conjecture of Erdős and the author in [28] was correct. Here is an example of one such conjecture and some of the subsequent developments. In [28], the following question was raised.

Suppose that a_k are positive integers satisfying

$$(3) \quad 1 < a_1 < a_2 < \dots < a_t.$$

Is it true that if $\sum_{k=1}^t \frac{1}{a_k} < 2$, then there exist $\varepsilon_k = 0$ or 1 so that

$$\sum_{k=1}^t \frac{\varepsilon_k}{a_k} < 1 \quad \text{and} \quad \sum_{k=1}^t \frac{1 - \varepsilon_k}{a_k} < 1?$$

As noted in [28], this is not true if we just assume that

$$(4) \quad 1 < a_1 \leq a_2 \leq \dots \leq a_t$$

as the sequence 2, 3, 3, 5, 5, 5, 5 shows. However, it was pointed out by Sándor [49] that our conjecture was too optimistic since the sequence consisting of the divisors of 120 with the exception of 1 and 120 provides a counterexample. In fact, Sándor proved the more general result that for every $n \geq 2$, there exist integers a_k satisfying (3) such that $\sum_{k=1}^t \frac{1}{a_k} < n$ and that this sum cannot be split into n parts so that all the partial sums are ≤ 1 . However, he also shows that for such a sequence the sum cannot be too much less than n . Specifically, Sándor proves:

Theorem 4. Suppose $n \geq 2$. If $1 < a_1 < a_2 < \dots < a_t$ are integers and

$$\sum_{k=1}^t \frac{1}{a_k} < n - \frac{n}{e^{n-1}}$$

then this sum can be decomposed into n parts so that all partial sums are ≤ 1 .

It was however conjectured by Erdős, Spencer and the author that if the a_k satisfy (4), as well as the stronger condition

$$(5) \quad \sum_{k=1}^t \frac{1}{a_k} < n - \frac{1}{30},$$

then the a_k can be split into n sequences $a_k^{(i)}, 1 \leq i \leq n$, so that

$$\sum_k \frac{1}{a_k^{(i)}} \leq 1$$

for all i . The reason that the bound $n - \frac{1}{30}$ was chosen was because of the example $a_1 = 2, a_2 = a_3 = 3, a_4 = a_5 = \dots = a_{5n-3} = 5$. Put another way, define $\alpha(n)$ to be the least real number so that if the a_k satisfy (4) and

$$(6) \quad \sum_{k=1}^t \frac{1}{a_k} < n - \alpha(n)$$

then the a_k can be split into n sequences $a_k^{(i)}, 1 \leq i \leq n$, so that

$$\sum_k \frac{1}{a_k^{(i)}} \leq 1$$

for all i . Thus, the conjecture in [28] was that $\alpha(n) = \frac{1}{30}$. In [49] it was shown by Sándor that $\alpha(n) \leq \frac{1}{2}$. This was improved by Chen [16] who shows that $\alpha(n) \leq \frac{1}{3}$. This in turn was followed by the paper of Fang and Chen [31] who prove that $\alpha(n) \leq \frac{2}{7}$. However, the original conjecture that $\alpha(n) = \frac{1}{30}$ was finally disproved by Guo [38] who showed that $\alpha(n) \geq \frac{5}{132} > \frac{1}{30}$. He shows that for the sequence $a_1 = 2, a_2 = 3, a_4 = 4, a_5 = \dots = a_{11n-12} = 11$,

$$\sum_{k=1}^{11n-12} \frac{1}{a_k} = n - \frac{5}{132},$$

but for any partition of $\{1, 2, \dots, 11n - 12\} = \cup_{j=1}^n A_j$, there exists a j such that $\sum_{k \in A_j} \frac{1}{a_k} > 1$. At present, we have no guess as to what the truth is for this problem.

6. CONCLUDING REMARKS

We have tried to give a sample of the very many interesting questions and results that were inspired by Paul Erdős' interest in Egyptian fractions. Of course, this list is far from complete, and in fact the subject is still quite dynamic. For further references, the reader can consult [39], [28], [52] or [14], for example, and the references therein.

REFERENCES

- [1] M. H. Ahmadi and M. N. Bleicher, On the conjectures of Erdős and Straus, and Sierpiński on Egyptian fractions, *J. Math. Stat. Sci.*, **7** (1998), 169–185.
- [2] A. V. Aho and N. J. A. Sloane, Some doubly exponential sequences, *Fib. Quart.* **11** (1973), 429–438.
- [3] P. J. van Albada and J. H. van Lint, Reciprocal bases for the integers, *Amer. Math. Monthly* **70** (1963), 170–174.
- [4] E. J. Barbeau, Compute challenge corner: Problem 477: A brute force program, *J. Rec. Math.*, **9** (1976/77), p. 30.
- [5] E. J. Barbeau, Expressing one as a sum of odd reciprocals: comments and a bibliography, *Cruz Mathematicorum* **3** (1977), 178–181.
- [6] M. N. Bleicher and P. Erdős, The number of distinct subsums of $\sum_{i=1}^N \frac{1}{i}$, *Math. Comp* **29** (1975), 29–42.
- [7] M. N. Bleicher and P. Erdős, Denominators of unit fractions, *J. Number Th.*, **8** (1976), 157–168.
- [8] M. N. Bleicher and P. Erdős, Denominators of unit fractions II, *Illinois J. of Math.* **20** (1976), 598–613.
- [9] Robert Breusch, A special case of Egyptian fractions, solution to Advanced Problem 4512, *Amer. Math. Monthly* **61**, (1954), 200–201.
- [10] T. C. Brown and V. Rödl, Monochromatic solutions to equations with unit fractions, *Bull. Australian Math. Soc.* **43** (1991), 387–392.
- [11] N. Burshtein, On distinct unit fractions whose sum equals 1, *Discrete Math.* **5** (1973), 201–206.
- [12] N. Burshtein, Improving solutions of $\sum_{i=1}^k \frac{1}{x_i} = 1$ with restrictions as required by Barbeau respectively by Johnson, *Discrete Math.* **306** (2006), 1438–1439.
- [13] N. Burshtein, An improved solution of $\sum_{i=1}^k \frac{1}{x_i} = 1$ in distinct integers when x_i doesn't divide x_j for $i \neq j$. *NNTDM* **16** (2010), 1–4.
- [14] Paul Campbell, Bibliography of algorithms for Egyptian fractions (preprint), Beloit Coll., Beloit WI53511, U.S.A.
- [15] J. W. S. Cassels, On the representation of integers as the sums of distinct summands taken from a fixed set, *Acta Sci. Math. Szeged* **21** (1960), 111–124.

- [16] Y.-G. Chen, On a conjecture of Erdős, Graham and Spencer, *J. Number Th.* **119** (2006), 307–314.
- [17] Y.-G. Chen and M. Tang, On the elementary symmetric functions of $1, \frac{1}{2}, \dots, \frac{1}{n}$, *Amer. Math. Monthly* **119** (2012), 862–867.
- [18] E. S. Croot III, On unit fractions with denominators from short intervals, *Acta Arith.* **99**, (2001), 99–114.
- [19] E. S. Croot III, On some questions of Erdős and Graham about Egyptian fractions, *Mathematika* **46** (1999), 359–372.
- [20] E. S. Croot III, On a coloring conjecture about unit fractions, *Ann. of Math.* **157** (2003), 545–556.
- [21] E. S. Croot III, Sums of the form $\frac{1}{x_1^k} + \dots + \frac{1}{x_n^k}$ modulo a prime, *Integers* **4**, (2004), A20 6.
- [22] D. R. Curtiss, On Kellogg's Diophantine problem, *Amer. Math. Monthly* **29** (1922), 380–387.
- [23] C. Elsholtz, Sums of k unit fractions, *Trans. Amer. Math. Soc.* **353** (2001), 3209–3227.
- [24] C. Elsholtz and T. Tao, Counting the number of solutions to the Erdős-Straus equation on unit fractions, arXiv:1107.1010 (53 pages).
- [25] P. Erdős, Beweis eines Satzes von Tschebyschef (in German), *Acta Litt. Sci. Szeged* **5**, (1932), 194–198.
- [26] P. Erdős, Egy Kürschák-féle elemi számelméleti tétel általánosítása (Generalization of an elementary number-theoretic theorem of Kürschák, in Hungarian), *Mat. Fiz. Lapok* **39** (1932), 17–24.
- [27] P. Erdős, Az $\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} = \frac{a}{b}$ egyenlet egész számú megoldásairól (On a Diophantine equation, in Hungarian), *Mat. Lapok* **1** (1950) 192–210.
- [28] P. Erdős and R. L. Graham, Old and New Problems and Results in Combinatorial Number Theory, Mono. No. 28 de L'Enseignement Math., Univ. Geneva (1980) 128 pp.
- [29] P. Erdős and I. Niven, On certain variations of the harmonic series, *Bull. Amer. Math. Soc.*, **51** (1945), 433–436.
- [30] P. Erdős and R. Rado, Intersection theorems for systems of sets, *J. London Math. Soc.* **35** (1960), 85–90.
- [31] J.-H. Fang and Y.-G. Chen, On a conjecture of Erdős, Graham and Spencer II, *Disc. Appl. Math.* **156** (2008), 2950–2958.
- [32] Leonardo Fibonacci, *Liber Abaci*, translated by L. E. Sigler, Springer, New York, 2003 (first published in 1202).
- [33] C. Friedman, Sums of divisors and Egyptian fractions, *J. Num. Th.* **44** (1993), 328–339.
- [34] Richard J. Gillings, *Mathematics in the Time of the Pharaohs*, 1972, MIT Press, Dover reprint ISBN 0-486-24315-X.
- [35] R. L. Graham, A theorem on partitions, *J. Australian Math. Soc.* **4** (1963), 435–441.

- [36] R. L. Graham, On finite sums of unit fractions, *Proc. London Math. Soc.* **14** (1964), 193–207.
- [37] R. L. Graham, On finite sums of reciprocals of distinct n^{th} powers, *Pacific J. Math.* **14** (1964), 85–92.
- [38] S. Guo, A counterexample to a conjecture of Erdős, Graham and Spencer, *The Electronic Journal of Combinatorics* 15.43 (2008): 1.
- [39] R. K. Guy, *Unsolved Problems in Number Theory*, Third edition, Springer, New York, 2004, 437 pp.
- [40] Chao Ko, Chi Sun and S. J. Chang, On equations $\frac{4}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$, *Acta Sci. Natur. Szechuanensis*, **2** (1964), 21–35.
- [41] J. Kürschák. A harmonikus sorról (On the harmonic series, in Hungarian), *Mat. Fiz, Lapok* **27** (1918), 288–300.
- [42] D. H. Lehmer, (personal communication).
- [43] Greg Martin, Dense Egyptian fractions, *Trans. Amer. Math. Soc.* **351** (1999), 3641–3657.
- [44] Greg Martin, Denser Egyptian fractions, *Acta Arith.* **95** (2000), 231–260.
- [45] R. Nowakowski, Unsolved Problems, 1989–1999, *Amer. Math. Monthly*, **106**, 959–962.
- [46] M. R. Obláth, Sur l’équation diophantienne $\frac{4}{n} = \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}$, *Mathesis* **59** (1950), 308–316.
- [47] Gay Robins and Charles Shute, *The Rhind Mathematical Papyrus: an ancient Egyptian text*, Dover, New York, 1987, 88pp.
- [48] J. Sander, On $\frac{4}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$ and Rosser’s sieve, *Acta Arith.*, **59** (1991), 183–204.
- [49] C. Sándor, On a problem of Erdős, *J. Number Th.* **63** (1997), 203–210.
- [50] C. Sándor, On the number of solutions of the Diophantine equation $\sum_{i=1}^n \frac{1}{x_i} = 1$, *Period. Math. Hungar.* **47** (2003), no. 1–2, 215–219.
- [51] C. Sándor, (personal communication).
- [52] A. Schinzel, Erdős’ work on finite sums of unit fractions, in **Paul Erdős and His Mathematics**, vol. I, Springer, Berlin, 2002, pp. 629–636.
- [53] Sherman Stein, (personal communication).
- [54] I. Shparlinski, On a question of Erdős and Graham, *Archiv der Math.* **78** (2002), 445–448.
- [55] W. Sierpiński, Sur les décompositions de nombres rationnels en fractions primaires, *Mathesis*, **65** (1956), 16–32.
- [56] B. M. Stewart, Sums of distinct divisors, *Amer. J. Math.* **76** (1954), 779–785.
- [57] A. Swett, <http://math.uindy.edu/swett/esc.htm> (accessed on 12/8/12).
- [58] R. C. Vaughan, On a problem of Erdős, Straus and Schinzel, *Mathematica.* **17** (1970), 193–198.
- [59] M. D. Vose, Egyptian fractions, *Bull. London Math. Soc.*, **17**, (1985), 21–24.
- [60] W. A. Webb, Rationals not expressible as the sum of three unit fractions, *Elemente der Math.* **29**, (1974), 1–6.

- [61] W. A. Webb, On $\frac{4}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$, *Proc. Amer. Math. Soc.* **25** (1970), 578–584.
- [62] André Weil (personal communication).
- [63] Wikipedia entry on the Rhind Papyrus.
- [64] H. S. Wilf, Reciprocal bases for the integers, *Bull. Amer. Math. Soc.* **67** (1961), p. 456.

Ronald L. Graham

*Department of Computer Science and
Engineering,
University of California,
San Diego,
La Jolla, CA 92093-0404,
USA*

e-mail: graham@ucsd.edu

7. APPENDIX: SOME (UNDATED) NOTES OF ERDŐS ON EGYPTIAN FRACTIONS

Graham and I proved some years ago the following question: Color the integers by k colors. Is it true

that
 (1) $\sum \frac{1}{x_i} = 1, x_1 < x_2 < \dots$

is monochromatically solvable? The sum (1) is of course supposed to be finite, but the number of summand can be as large as we please.

Let $f(m)$ be the largest integer for which there is a sequence $x_1 < x_2 < \dots < x_t \leq m, t = f(m)$ which does not contain a solution of (1). Trivially

$$f(m) > m(1 + \frac{1}{2} - \epsilon)$$

but perhaps $f(m) = m + o(m)$. We could not get

non trivial upper or lower bounds for $f(m)$.

Fig. 2. Some notes of Erdős on Egyptian fractions

Let $x_1 < x_2 < \dots$ have positive upper density doesn't then follow that it contains a solution of (1).

Also if

$$(3) \quad \sum_{x_i < m} \frac{1}{x_i} > g(m)$$

when does (3) force that $\sum \varepsilon_i/x_i = 1$, $\varepsilon_i = 0$ or 1 should be solvable.

Perhaps the following related problems are not without interest: Denote by $g_n(m)$ resp $g_\infty(m)$ the largest set of integers $x_1 < x_2 < \dots < x_k \leq m$ for which all the sums

$$\sum_{i=1}^k \varepsilon_i/x_i \quad (\text{resp } \sum \varepsilon_i/x_i)$$

are all distinct. Estimate $g_n(m)$ resp $g_\infty(m)$ as

accurately as possible. Also if $x_1 < x_2 < \dots$ is an infinite sequence and all the sums $\sum_{i=1}^k \varepsilon_i/x_i$ resp $\sum \varepsilon_i/x_i$ are all different how large can $g_n(m)$ resp $g_\infty(m)$ be for all n ?

Fig. 3. More notes of Erdős on Egyptian fractions



AT&T Bell Laboratories

600 Mountain Avenue
Murray Hill, NJ 07974-2070
908-582-3000

Graham and I conjectured several years ago that if we color the integers by k colors then

$$(1) \quad \sum \frac{1}{x_i}, \quad x_1 < x_2 < \dots$$

is solvable monochromatically. (In (1) the sum is of course finite).

More generally we conjectured that if we divide the integers into k classes there is $S_r, 1 \leq r \leq k$, there is an r so that every rational $\frac{a}{b}$ can be written in the form

$$(2) \quad \frac{a}{b} = \sum \frac{1}{x_i}, \quad x_1 < x_2 < \dots, \quad x_i \in S_r$$

and the sum in (2) is of course finite. The slightly weaker conjecture: (2) is solvable, monochromatically, is also open. These attractive conjectures have perhaps ^{been} undeservedly neglected. We formulated several related problems which we feel are also interesting. Let $f(r, m)$ be the smallest integer for which if we divide the integer $z = t \leq f(r, m)$ into r classes then (1) is not monochromatically. Prove that $f(r, m)$ exists for every r and estimate $f(r, m)$ as well as possible. A further complication can be added: Denote by $f(r, t)$ the smallest integer (if it exists) for which if we divide the integer $z = t \leq f(r, t)$ into r classes then

$$(3) \quad 1 = \sum \frac{1}{x_i}, \quad z \leq x_1 < x_2 < \dots < x_n = f(r, t, m), \quad 3 \leq n \leq t$$

is solvable monochromatically. Clearly for small t $f(r, t, m)$ does not exist, e.g. for $r=2$. Perhaps one could try to determine the smallest t for which $f(2, t)$ exists.

Determine or estimate the smallest $g(m)$ for which in every set of $g(m)$ integers $z \leq x_1 < x_2 < \dots < x_t \leq m, t = g(m)$ (1) is solvable. Trivially

$$(4) \quad g(m) > m(1 - \frac{1}{e}) + O(1)$$

Fig. 4. Notes of Erdős on Egyptian fractions (while visiting Bell Labs)

We have no non-trivial upper or lower bound for $g(m)$ and could not decide if $g(m) > m - \sigma(m)$ holds.

The following Greiner type result could perhaps hold:

Let $x_1 < x_2 < \dots$ be a sequence of positive lower density. Is it then true that

$$(5) \quad 1 = \sum_i \frac{\epsilon_i}{x_i}, \quad \epsilon_i = 0 \text{ or } 1 \quad (\text{finite sum})$$

is always solvable. The primes show that \neq the divergence of $\sum \frac{1}{x_i}$ is not enough for the solvability of (1), but perhaps if

$$\frac{1}{\log \log m} \sum_{x_i \leq m} \frac{1}{x_i} \rightarrow \infty$$

then (5) is always solvable. In fact let $h(m)$ be the smallest number for which if

$$\sum_{x_i \leq m} \frac{1}{x_i} > h(m)$$

then (5) is solvable. Estimate $h(m)$ as well as possible from above and below.

Several further related questions could be posed, for further details see our book Old and new problems and results in combinatorial number theory, Monographie N° 28 de G. G. & L'Enseignement Math. also see some papers of Bleicher and Galis which are referenced in our book

Let m be the smallest integer for which if we divide the proper divisors of m into k classes m is the monochromatic sum of numbers of the same class. I can not even prove that m is odd.

Fig. 5. Notes of Erdős on Egyptian fractions (while visiting Bell Labs)

PERFECT POWERS IN PRODUCTS WITH CONSECUTIVE TERMS FROM ARITHMETIC PROGRESSIONS, II

KÁLMÁN GYÖRY*

1. INTRODUCTION

There is a very rich literature on perfect powers or almost perfect powers in products of the form $m(m+d)\dots(m+(k-1)d)$, where m, d are coprime positive integers and $k \geq 3$. By a conjecture, such a product is never a perfect n -th power if $k > 3$, $n \geq 2$ or $k = 3$, $n > 2$. In the classical case $d = 1$ the conjecture has been proved by Erdős and Selfridge [11]. The general case $d \geq 1$ seems to be very hard, then there are only partial results; for survey papers on results obtained before 2006 we refer to Tijdeman [46]–[48], Shorey and Tijdeman [43, 44], Shorey [38]–[42] and Győry [15, 16].

Since 2006, considerable progress has been made in the general situation. In this paper which may be considered as a continuation of Győry [16] we give an overview of the most important recent results. We restrict ourselves to those results which provide, for a fixed k , all perfect or almost perfect powers in products of the above type.

In Section 2, a brief survey is given on the classical case and a related problem concerning binomial coefficients. In Section 3, we present some results of Hirata-Kohno, Laishram, Shorey and Tijdeman [23] and Tengely [45] for $n = 2$, Hajdu, Tengely and Tijdeman [21] for $n = 3$, Győry, Hajdu and Pintér [17] for $n \geq 5$ and Hajdu and Kovács [20] for $n = 5$. These results confirmed the above-mentioned conjecture for $k < 35$. In Section 4, we deal with an application from [17] to rational solutions of a related superelliptic equation. Finally, in Section 5, the basic ideas and the main tools of the proofs are discussed. As will be seen, different techniques are needed for $n = 2, 3, 5$ and $n \geq 7$. The case $n \geq 7$ requires the complete solution of a

*Research was supported by the OTKA grants T67580, K75566, NK104208 and K100339.

number of ternary equations by means of the theory of Galois representations and modular forms.

2. PRODUCT OF CONSECUTIVE INTEGERS

After a lot of special results, Erdős and Selfridge [11] proved in 1975 the following remarkable theorem.

Theorem A. *The equation*

$$(2.1) \quad m(m+1)\dots(m+k-1) = y^n$$

has no solutions in positive integers m, k, y, n with $k \geq 2, n \geq 2$.

In other words, the product of consecutive positive integers is never a perfect power. The proof is elementary but complicated and ingenious.

Saradha and Shorey [34, 36] determined all the solutions of (2.1) in the case when one or two of the factors $(m+i)$ on the left hand side are omitted.

A related equation is

$$(2.2) \quad \binom{m+k-1}{k} = y^n,$$

where m, k, y, n are integers with $k, y, n \geq 2$ and $m \geq k+1$. When $k = n = 2$, this yields a Pell equation, having infinitely many solutions. For $(k, n) = (3, 2)$, Meyl [27] and Watson [49] proved that $(m, y) = (48, 140)$ is the only solution of (2.2).

Using his elementary method applied to the equation (2.1), Erdős [10] proved in 1951 that for $k \geq 4$, equation (2.2) has no solution. For $k < 4$, his approach does not work. The case $k = 2$ of the next theorem is a consequence of a result of Darmon and Merel [7], while the case $k = 3, n > 2$ is due to the present author [13].

Theorem B. *Apart from the case $k = n = 2, (m, k, y, n) = (48, 3, 140, 2)$ is the only solution of (2.2).*

In the case $k = 2, 3, n > 2$, the proofs depend on some deep results on generalized Fermat's equations.

For some further interesting related results, we refer to [51], [53] and [26].

Denote by $P(b)$ the greatest prime factor of an integer $b > 1$, and write $P(1) = 1$. As a common generalization of equations (2.1) and (2.2) consider the equation

$$(2.3) \quad m(m+1)\dots(m+k-1) = by^n,$$

where m , k , b , y , n are unknown positive integers with $k \geq 2$, $n \geq 2$, $P(b) \leq k$.

In (2.3), (m, y) yields a solution with $P(y) \leq k$ if and only if $m \in \{1, 2, \dots, p^{(k)} - k\}$, where $p^{(k)}$ denotes the least prime with $p^{(k)} > k$; cf. Györy [14]. Such solutions are called *trivial*. For given k , the trivial solutions can easily be found. Hence it suffices to deal with non-trivial solutions. Further, to make b and y uniquely determined in (2.3), we may assume that b is n th power free.

For $P(b) < k$, the following theorem is due to Erdős and Selfridge [11], for $k \geq 4$ to Saradha [32], while for $k < 4$ to Györy [14].

Theorem C. *Apart from the case $(k, b, n) = (2, 2, 2)$, the only non-trivial solution of equation (2.3) is $(m, k, b, y, n) = (48, 3, 6, 140, 2)$.*

The proof of the case $k \geq 4$ is based on a refinement of Erdős' elementary method, while the cases $k = 2, 3$ involve some profound results on generalized Fermat's equations.

For $b = 1$, (2.3) is just equation (2.1), while for $b = k!$ it reduces to equation (2.2). Hence Theorem C gives Theorem A and Theorem B as special cases.

Clearly, Theorem C remains valid also with $P(b) < p^{(k)}$. Later, Theorem C has been refined by Saradha [32] for $k \geq 9$, Hanrot, Saradha and Shorey [22] for $6 \leq k \leq 8$ and Bennett [1] for $3 \leq k \leq 5$. They proved that for $n \geq 3$ and $P(b) \leq p^{(k)}$, equation (2.3) has no non-trivial solutions. Moreover, Pintér and the author [19] showed that in the case $3 \leq k \leq 5$, $n > 2$, equation (2.3) has no non-trivial solution even for $P(b) \leq p_k$, where p_k denotes the k th prime. Obviously, $p_k > p^{(k)}$ if $k \geq 4$.

In [19], the following conjecture is proposed.

Conjecture 1. *For $k \geq 3$ and $n > 2$, (2.3) has no non-trivial solutions with $P(b) \leq p_k$.*

3. PRODUCTS OF CONSECUTIVE TERMS IN ARITHMETIC PROGRESSION

This section is devoted to the more general equations

$$(3.1) \quad m(m+d) \dots (m+(k-1)d) = y^n$$

and

$$(3.2) \quad m(m+d) \dots (m+(k-1)d) = by^n,$$

where m, k, d, b, y, n are unknown positive integers such that $k \geq 3$, $n \geq 2$, $\gcd(m, d) = 1$ and $P(b) \leq k$. As was seen above, these equations have been solved in the special case $d = 1$. Hence we shall concentrate on the case $d > 1$. Further, we restrict ourselves to the case when k is fixed.

It is easy to show that equation (3.1) has infinitely many solutions both for $k = 2$ and for $(k, n) = (3, 2)$. It was proved by Euler that for $(k, n) = (4, 2)$, equation (3.1) is impossible. This result was extended by Obláth [28, 29] to the cases $(k, n) = (3, 3), (3, 4), (3, 5)$ and $(5, 2)$.

For arbitrary $n > 2$, the first result in this direction was obtained by the author [15] who proved that, for $k = 3$, $n > 2$ and $P(b) \leq 2$, equation (3.2) has no solution. This implies at once that for $k = 3$, $n > 2$, equation (3.1) is also impossible. The assumption $P(b) \leq 2$ cannot be relaxed to $P(b) \leq 3$ because (3.2) has infinitely many solutions with $P(b) = 3$; see Tijdeman [46]. The proof of Győry's result [15] is based on theorems of Wiles [50], Ribet [30] and Darmon and Merel [7] on Fermat's type equations.

In 2004, Győry, Hajdu and Saradha [18] showed that equation (3.2) has no solution for $k = 4, 5$ and $P(b) \leq 2$. This was extended by Bennett, Bruin, Győry and Hajdu [3] to the cases

$$\begin{aligned} k = 6, & \quad P(b) \leq 2 \\ 7 \leq k \leq 10, & \quad P(b) \leq 3 \\ k = 11, & \quad P(b) \leq 5. \end{aligned}$$

Later, for $k = 5, 6$ and $n \geq 7$, Bennett [2] improved the earlier bound on $P(b)$ to $P(b) \leq 3$. Clearly, the above results imply that, for $4 \leq k \leq 11$, equation (3.1) is impossible.

As will be pointed out in Section 5, the proofs required different methods for $n = 2, 3, 5$ and $n \geq 7$.

Since 2006, considerable progress has been made. For $n = 2$, Hirata-Kohno, Laishram, Shorey and Tijdeman [23] have achieved a significant

extension of Euler's theorem. They established their result for the more general equation (3.2), but were not able to handle the situation for some exceptional values of $b > 1$; for these values, (3.2) was solved by Tengely [45]. The results of [23] and [45] together give the following.

Theorem 1a. *Equation (3.2) with $n = 2$ and $5 \leq k \leq 100$, $d > 1$ has no solution.*

In the case $b = 1$, the authors of [23] proved even more.

Theorem 1b. *Equation (3.1) with $n = 2$ and $4 \leq k \leq 109$ is not possible.*

The following two theorems concern the case $n = 3$. They are due to Hajdu, Tengely and Tijdeman [21].

Theorem 2a. *Equation (3.2) with $n = 3$, $d > 1$, $3 \leq k < 32$ and with $P(b) < k$ if $k = 3$ or $k \geq 13$, is not possible.*

In the special case $b = 1$, the range of k 's has been further augmented.

Theorem 2b. *Equation (3.1) with $n = 3$ and $3 \leq k < 39$ has no solution.*

For the case $n > 3$, Györy, Hajdu and Pintér [17] considerably extended the results of [15], [18] and [3] by proving the following. We may assume that n is a prime.

Theorem 3a. *Equation (3.2) has no solution*

(i) *if $n \geq 7$ prime and*

$$\begin{aligned} 12 \leq k \leq 22, \quad P(b) \leq 7, \\ 22 < k < 35, \quad P(b) \leq \frac{k-1}{2}, \end{aligned}$$

(ii) *or if $n = 5$, $d > 1$ and*

$$\begin{aligned} 8 \leq k \leq 22, \quad P(b) \leq 7, \\ 22 < k < 35, \quad P(b) \leq \frac{k-1}{2}. \end{aligned}$$

It is clear that Theorem 3a remains valid if in (3.2) n is not necessarily prime but has a prime factor ≥ 5 .

For $n = 5$, $8 \leq k \leq 11$, Theorem 3a gives an improvement of the corresponding result of [3].

The above results on equation (3.1) can be summarized as follows.

Theorem 3b. *If $3 \leq k < 35$, then equation (3.1) has no solution with $(k, n) \neq (3, 2)$.*

When $n \leq 3$ or $k \leq 11$, Theorem 3b follows from Theorems 1b, 2b and the above-mentioned results of [15], [18] and [3]. The case $n > 3$, $11 < k < 35$ is an immediate consequence of Theorem 3a.

Theorem 3b suggests the following conjecture which is a more precise version of an earlier conjecture of Erdős.

Conjecture 2. *For $k \geq 3$ and $(k, n) \neq (3, 2)$, equation (3.1) has no solution.*

Similarly, the results concerning (3.2) suggest the following.

Conjecture 3. *If $k \geq 3$, $(k, n) \neq (3, 2)$ and $P(b) \leq 2$, equation (3.2) has no solution.*

As is shown by the examples $2 \cdot 9 \cdot 16 = 2^5 \cdot 3^2$ and $1 \cdot 2 \cdot 3 \cdot 4 = 2^3 \cdot 3$, for $k = 3$ and 4 the assumption $P(b) \leq 2$ cannot be replaced by $P(b) \leq 3$. It is likely, however, that for $k \geq 5$ this assumption can be weakened.

In the case $n = 5$, Hajdu and Kovács [20] have recently obtained a further extension.

Theorem 4a. *For $n = 5$ and $3 \leq k \leq 36$, equation (3.2) has the only solution $(m, k, d) = (2, 3, 7)$.*

For the equation (3.1), even more has been proved in [20].

Theorem 4b. *If $n = 5$ and $3 \leq k \leq 54$, equation (3.1) has no solution.*

We note that the authors of [15], [18], [3] and Theorems 1a to 4a extended their results mentioned above to the case when in (3.2) m and b are non-zero integers but not necessarily positive. Further, we mention the interesting papers of Saradha and Shorey [35], Laishram, Shorey and Tengely [24], Yang, Togbé and He [52], Saradha [33], and Laishram and Shorey [25].

4. AN APPLICATION OF THEOREM 3b

We present a consequence of Theorem 3b for the superelliptic equation

$$(4.1) \quad x(x+1) \dots (x+k-1) = w^n,$$

where x, k, w, n are unknowns with integers $k, n \geq 2$ and positive rationals x, w .

Sander [31] proved that if $2 \leq k \leq 4$ and $(k, n) \neq (2, 2)$, then equation (4.1) has no solutions. Further, he conjectured that (4.1) has no solution if $(k, n) \neq (2, 2)$.

By putting $x = m/d$ and $w = y/u$ with positive integers m, d, y, u such that $\gcd(m, d) = 1$ and $\gcd(y, u) = 1$ we see that (4.1) reduces to the equation

$$m(m+d)\dots(m+(k-1)d) = y^n, \quad d^k = u^n.$$

The following corollary is a straightforward consequence of Theorem 3b.

Corollary to Theorem 3b. *Suppose that $1 < k < 35$, $n \geq 2$ and $(k, n) \neq (2, 2)$. Then equation (4.1) has no solution in positive rational numbers x, w .*

For $k \leq 11$, this was proved in [3].

5. BASIC IDEAS AND MAIN TOOLS IN THE PROOFS

We outline the basic ideas and main tools in the proofs of Theorems 1a to 4a. Fix $k \geq 3$, and assume that the equation

$$(3.2) \quad m(m+d)\dots(m+(k-1)d) = by^n$$

has a solution in positive integers m, d, b, y, n with $n \geq 2$, $\gcd(m, d) = 1$ and $P(b) \leq k$. We may assume that n is a prime. From (3.2) one can then deduce that

$$(5.1) \quad m + id = a_i x_i^n, \quad P(a_i) \leq k,$$

with some positive integers $a_i, x_i, i = 0, \dots, k-1$. Clearly, (5.1) implies (3.2), i.e. (3.2) and (5.1) are equivalent. The a_i, x_i can be chosen so that a_i is n th power free. There are only finitely many and effectively determinable such k -tuples $(a_0, a_1, \dots, a_{k-1})$.

If there are i, j with $0 \leq i, j < k-1$ such that $P(a_i a_{i+1} \dots a_{i+j}) \leq j+1$, then we can reduce (3.2) to the case when k is replaced by $j+1 < k$. However, this is not the case in general.

The equation (3.2) can be reduced to ternary equations. There are two possibilities:

1) For distinct integers $0 \leq p, q, r \leq k-1$, it is easy to find non-zero integers $\lambda_p, \lambda_q, \lambda_r$ with absolute values $\leq k$ such that

$$\lambda_p(m + pd) + \lambda_q(m + qd) = \lambda_r(m + rd).$$

Consequently, using (5.1) we obtain an equation of the form

$$(5.2) \quad AX^n + BY^n = CZ^n \quad \text{in coprime non-zero integers } X, Y, Z,$$

where A, B, C are relatively prime non-zero integers with $P(ABC) \leq k$.

2) For integers $0 \leq p < q \leq r < s \leq k - 1$ with $p + s = q + r$, we have the identity

$$(m + qd)(m + rd) - (m + pd)(m + sd) = (qr - ps)d^2.$$

Thus, in view of (5.1) we get an equation of the shape

$$(5.3) \quad AX^n + BY^n = CZ^2 \quad \text{in coprime non-zero integers } X, Y, Z,$$

where A, B, C are relatively prime non-zero integers with $P(AB) \leq k$ and $|C| \leq (k - 1)^2$.

In (5.2) and (5.3) it suffices to study the coprime non-zero solutions X, Y, Z with $XYZ \neq \pm 1$. Such solutions will be called *non-trivial*.

We arrived at complicated systems of equations which consist of equations of the shapes (5.2) and (5.3). When $2 \leq n \leq 7$, for certain choices of the a_i one can use local methods, showing that at least one of the equations (5.2) or (5.3) involved is not solvable (mod p) for an appropriate prime p . In particular, quadratic, cubic and more generally n th power residues are successfully applied.

In general, several other methods are also needed to solve the equations under consideration. Different techniques must be used depending on the exponents $n = 2, 3, 5$ and ≥ 7 .

The case $n = 2$. Here the main ingredients are quadratic residues and elliptic curves. In many cases, for $n = 2$ or 3 , equation (3.2) may be reduced to finding the rational (torsion) points on certain rank 0 elliptic curves over \mathbb{Q} , and then one can use the program package MAGMA [5] to find all rational points. In a number of situations, however, this approach proves to be inadequate to deduce the desired result. Then, instead, one can use explicit Chabauty techniques due to Bruin and Flynn [6]. Hirata-Kohno, Laishram, Shorey and Tijdeman [23] provide a method for solving equation (3.2) for $n = 2$ and for any given value of k , unless $(a_0, a_1, \dots, a_{k-1})$ belongs to a finite and effectively determinable set of tuples. For $k \leq 100$, the exceptional cases $(a_0, a_1, \dots, a_{k-1})$ occurring in [23] have been settled by Tengely [45] using the elliptic Chabauty method.

The case $n = 3$. In a number of cases classical results of Selmer [37] and others on cubic equations $AX^3 + BY^3 + CZ^3 = 0$ can be used to prove that at least one of the equations (5.2) arising from (3.2) has no non-trivial solution. A further important tool is reducing equation (3.2) to elliptic curves and applying the Chabauty method to solve the corresponding equations. In Hajdu, Tengely and Tijdeman [21], the above-mentioned methods are combined with the approach of [23] developed for the case $n = 2$.

The case $n = 5$. The methods applied in the cases $n = 2$ and 3 do not work for $n = 5$. On the other hand, for $n = 5$ hardly new information is available through the theory of “general” modular forms which, as will be seen below, is the main tool in the case $n \geq 7$. For $n = 5$, one can make use of some classical and new results of Dirichlet, Lebesgue, Maillet (see e.g. [9]), Dénes [8], Győry [12] and Bennett, Bruin, Győry and Hajdu [3] on equations of the form $AX^5 + BY^5 = CZ^5$, the proofs of which involve cyclotomic and local considerations. In the case $n = 5$, Hajdu and Kovács [20] considerably improved and extended the previous results on (3.1) and (3.2) by using genus 2 curves and Chabauty method (both the classical and the elliptic version). They solved a large number of genus 2 equations by Chabauty method, and then built a kind of sieve system based upon them.

The case $n \geq 7$. For $n \geq 7$, the main tool is the application of the modular method to ternary equations of the form (5.2) and (5.3) or, more precisely, the use of the approach based on the theory of Frey curves, Galois representations and modular forms.

The following **ternary equations** were used in our proofs in Győry [15], Győry, Hajdu and Saradha [18], Bennett, Bruin, Győry and Hajdu [3] and Győry, Hajdu and Pintér [17].

Case $3 \leq k \leq 34$. In [15], [18], [3] and [17] it was used that the equation

$$X^n + Y^n = 2^\alpha Z^n, \quad \alpha \geq 0,$$

has no non-trivial solution. For $\alpha = 0$, this is Wiles’ [50] famous theorem on the Fermat equation. For $\alpha = 1$, the result is due to Darmon and Merel [7], while for $1 < \alpha < n$, to Ribet [30].

Case $4 \leq k \leq 34$. In [18], [3] and [17], some results of Bennett and Skinner [4] were utilized which say that for coprime A, B and non-negative α, β , the equations

$$X^n + 2^\alpha Y^n = 3^\beta Z^2, \quad \alpha \neq 1,$$

$$X^n + Y^n = CZ^2, \quad C \in \{2, 6\}$$

$$X^n + 5^\alpha Y^n = 2Z^2 \quad \text{with } n \geq 11 \text{ if } \alpha > 0,$$

$$AX^n + BY^n = Z^2, \quad AB = 2^\alpha p^\beta, \quad \alpha \neq 1, \quad p \in \{11, 19\}$$

have no non-trivial solutions.

Case $6 \leq k \leq 11$. In [3], the authors obtained as auxiliary results that for coprime A, B and non-negative α, β the equations

$$X^n + 2^\alpha Y^n = Z^2 \quad \text{with } p \mid XY \text{ for } p \in \{3, 5, 7\},$$

$$X^n + 2^\alpha Y^n = 3Z^2 \quad \text{with } p \mid XY \text{ for } p \in \{5, 7\},$$

$$X^n + 3^\alpha Y^n = 2Z^2 \quad \text{with } p \mid XY \text{ for } p \in \{5, 7\}, \quad n \geq 11,$$

$$AX^n + BY^n = Z^2, \quad AB = 2^\alpha p^\beta, \quad \alpha \geq 6, \quad p \in \{3, 5, 13\},$$

$$AX^n + BY^n = Z^2, \quad P(AB) \leq 3, \quad p \mid XY \text{ for } p \in \{5, 7\},$$

$$AX^n + BY^n = Z^2, \quad P(AB) \leq 5, \quad 7 \mid XY, \quad n \geq 11$$

have no non-trivial solutions. In these statements p always denotes a prime with $p < n$.

To extend the results concerning (3.1) and (3.2) from $k \leq 11$ to $12 \leq k \leq 34$, several new ternary equations had to be solved in [17]. Denote by

$$\text{rad}(m) = \prod_{p \mid m} p, \quad \text{rad}(1) = 1$$

the *radical* of a positive integer m , where the product is taken over all distinct prime factors p of m . Consider the set

$$\begin{aligned} I := \{ & (2, 1), (2, 3), (2, 5), (2, 7), (6, 1), (6, 5), (10, 1), (10, 3), (14, 1), (14, 3), \\ & (22, 1), (26, 1), (30, 1), (34, 1), (38, 1), (42, 1), (46, 1), (66, 1), (70, 1), \\ & (78, 1), (102, 1), (114, 1), (130, 1), (138, 1), (3, 1), (3, 5), (5, 1), (5, 3), \\ & (7, 1), (13, 1), (15, 1), (17, 1), (21, 1), (23, 1), (33, 1), (35, 1), (39, 1), \\ & (51, 1), (57, 1), (69, 1), (165, 1), (3, 2), (5, 6), (7, 2), (11, 2), (13, 2), \\ & (15, 2), (17, 2), (19, 2), (21, 2), (23, 2), (33, 2), (35, 2), (39, 2)\}. \end{aligned}$$

In [17], Győry, Hajdu and Pintér proved the following.

Theorem 5. *Let $n > 31$ be a prime, A, B, C coprime positive integers with $(\text{rad}(AB), C) \in I$ and p a prime such that $11 \leq p \leq 31$ and $p \nmid AB$. Then the equation*

$$AX^n + BY^n = CZ^2$$

has no non-trivial solutions X, Y, Z with $p \mid XY$, unless, possibly, for 60 tuples $(n, \text{rad}(AB), C, p)$ (which are listed explicitly in [17]).

We note that in the exceptional tuples, $37 \leq n \leq 239$.

To solve these equations, we combined the Frey curve and Galois representation approach with local and cyclotomic considerations. For the most part, our results concerning ternary equations, which may be of independent interest, do not follow from straightforward application of the modularity of Galois representations attached to Frey curves; it is also necessary to understand the reduction types of these curves at certain small primes.

In the proof of Theorems 3a and 3b, one of the main difficulties is that the number of systems of equations, that is, the number of arising tuples $(a_0, a_1, \dots, a_{k-1})$ grows so rapidly with k that in [17], for $k \geq 12$, it was practically impossible to handle the different cases as before for $k \leq 5$ in [18] and $k \leq 11$ in [3]. The main novelty in [17] lies in the development of an algorithm for our proofs, which enabled us to use a computer. We applied an efficient, iterated combination of our procedure for solving the arising new ternary equations with several “sieves” based on ternary equations already solved. This made it possible to exclude in each step the solvability of enormous number of systems of equations under consideration. Our general algorithm seems to work for larger values of k as well, although there are, of course, limits in computation of modular forms of higher and higher level and in the computational time itself.

Acknowledgements. The author is indebted to Professors L. Hajdu and Á. Pintér and Dr. Sz. Tengely for their useful remarks.

REFERENCES

- [1] M. A. Bennett, Products of consecutive integers, *Bull. London Math. Soc.*, **36** (2004), 683–694.
- [2] M. A. Bennett, Powers from five terms in arithmetic progression, in: *Diophantine Equations*, Narosa Publ. House, New Delhi, 2008, pp. 53–57.
- [3] M. A. Bennett, N. Bruin, K. Györy, and L. Hajdu, Powers from products of consecutive terms in arithmetic progression, *Proc. London Math. Soc.* **92** (2006), 273–306.

- [4] M. A. Bennett and C. M. Skinner, Ternary Diophantine equations via Galois representations and modular forms, *Canad. J. Math.*, **56** (2004), 23–54.
- [5] W. Bosma, J. Cannon, and C. Playoust, The Magma algebra system. I. The user language, *J. Symbolic Comput.* **24** (1997), 235–265.
- [6] N. Bruin and E. V. Flynn, Towers of 2-covers of hyperelliptic curves, *Trans. Amer. Math. Soc.*, **357** (2005), 4329–4347.
- [7] H. Darmon and L. Merel, Winding quotients and some variants of Fermat’s last theorem, *J. Reine Angew. Math.*, **490** (1997), 81–100.
- [8] P. Dénes, Über die diophantische Gleichung $x^l + y^l = cz^l$, *Acta Math.*, **88** (1952), 241–251.
- [9] L. E. Dickson, History of the Theory of Numbers, Vol. II, Carnegie Inst., Washington DC, 1919.
- [10] P. Erdős, On a diophantine equation, *J. London Math. Soc.*, **26** (1951), 176–178.
- [11] P. Erdős and J. L. Selfridge, The product of consecutive integers is never a power, *Illinois J. Math.*, **19** (1975), 292–301.
- [12] K. Győry, Über die diophantische Gleichung $x^p + y^p = cz^p$, *Publ. Math. Debrecen*, **13** (1966), 301–305.
- [13] K. Győry, On the diophantine equation $\binom{n}{k} = x^l$, *Acta Arith.*, **80** (1997), 289–295.
- [14] K. Győry, On the diophantine equation $n(n+1)\dots(n+k-1) = bx^l$, *Acta Arith.*, **83** (1998), 87–92.
- [15] K. Győry, Power values of products of consecutive integers and binomial coefficients, in: Number Theory and Its Applications, Kluwer Acad. Publ., 1999, pp. 145–156.
- [16] K. Győry, Perfect powers in products with consecutive terms from arithmetic progressions, in: More Sets, Graphs and Numbers, Springer and Bolyai Society, Budapest, 2006, pp. 143–155.
- [17] K. Győry, L. Hajdu and Á. Pintér, Perfect powers from products of consecutive terms in arithmetic progression, *Compos. Math.*, **145** (2009), 845–864.
- [18] K. Győry, L. Hajdu and N. Saradha, On the diophantine equation $n(n+d)\dots(n+(k-1)d) = by^l$, *Canad. Math. Bull.*, **47** (2004), 373–388.
- [19] K. Győry and Á. Pintér, Almost perfect powers in products of consecutive integers, *Monatsh. Math.*, **145** (2005), 19–33.
- [20] L. Hajdu and T. Kovács, Almost fifth powers in arithmetic progression, *J. Number Theory*, **131** (2011), 1912–1923.
- [21] L. Hajdu, Sz. Tengely and R. Tijdeman, Cubes in products of terms in arithmetic progression, *Publ. Math. Debrecen*, **74** (2009), 215–232.
- [22] G. Hanrot, N. Saradha and T. N. Shorey, Almost perfect powers in consecutive integers, *Acta Arith.*, **99** (2001), 13–25.
- [23] N. Hirata-Kohno, S. Laishram, T. N. Shorey and R. Tijdeman, An extension of a theorem of Euler, *Acta Arith.*, **129** (2007), 71–102.
- [24] S. Laishram, T. N. Shorey and Sz. Tengely, Squares in products in arithmetic progression with at most one term omitted and common difference a prime power, *Acta Arith.*, **135** (2008), 143–158.

- [25] S. Laishram and T. N. Shorey, Baker's explicit abc-conjecture and applications, *Acta Arith.*, **155** (2012), 419–429.
- [26] F. Luca, Perfect powers in q -binomial coefficients, *Acta Arith.*, **151** (2012), 279–292.
- [27] A. J. J. Meyl, Question 1194, *Nouv. Ann. Math.*, **17** (1878), 464–467.
- [28] R. Obláth, Über das Produkt fünf aufeinander folgender Zahlen in einer arithmetischen Reihe, *Publ. Math. Debrecen*, **1** (1950), 222–226.
- [29] R. Obláth, Eine Bemerkung über Produkte aufeinander folgender Zahlen, *J. Indian Math. Soc.*, **15** (1951), 135–139.
- [30] K. A. Ribet, On the equation $a^p + 2^\alpha b^p + c^p = 0$, *Acta Arith.*, **79** (1997), 7–16.
- [31] J. W. Sander, Rational points on a class of superelliptic curves, *J. London Math. Soc.*, **59** (1999), 422–434.
- [32] N. Saradha, On perfect powers in products with terms from arithmetic progressions, *Acta Arith.*, **82** (1997), 147–172.
- [33] N. Saradha, Application of the explicit abc-conjecture to two Diophantine equations, *Acta Arith.*, **151** (2012), 401–419.
- [34] N. Saradha and T. N. Shorey, Almost perfect powers in arithmetic progression, *Acta Arith.*, **99** (2001), 363–388.
- [35] N. Saradha and T. N. Shorey, On the equation $n(n+d)\dots(n+(i_0-1)d)(n+(i_0+1)d)\dots(n+(k-1)d) = y^l$ with $0 < i_0 < k-1$, *Acta Arith.*, **129** (2007), 1–21.
- [36] N. Saradha and T. N. Shorey, Almost perfect powers in consecutive integers, III, *Indag. Math.*, **19** (2008), 649–658.
- [37] E. Selmer, The diophantine equation $ax^3 + by^3 + cz^3 = 0$, *Acta Math.*, **85** (1951), 203–362.
- [38] T. N. Shorey, Exponential diophantine equations involving products of consecutive integers and related equations, in: *Number Theory*, Hindustan Book Agency, 1999, pp. 463–495.
- [39] T. N. Shorey, Mathematical contributions, *Bombay Math. Coll.*, **15** (1999), 1–19.
- [40] T. N. Shorey, Powers in arithmetic progression, in: *A Panorama in Number Theory*, Cambridge, 2002, pp. 325–336.
- [41] T. N. Shorey, Powers in arithmetic progression, II, in: *New Aspects of Analytic Number Theory*, Kyoto, 2002, pp. 202–214.
- [42] T. N. Shorey, Powers in arithmetic progression, III, in: *The Riemann Zeta Function and Related Themes*, 2006, pp. 131–140.
- [43] T. N. Shorey and R. Tijdeman, On the greatest prime factor of an arithmetic progression, in: *A Tribute to Paul Erdős*, Cambridge, 1990, pp. 385–389.
- [44] T. N. Shorey and R. Tijdeman, Some methods of Erdős applied to finite arithmetic progressions, in: *The Mathematics of Paul Erdős, I*, Springer, 1997, pp. 251–267.
- [45] Sz. Tengely, Note on the paper “An extension of a theorem of Euler” by Hirata-Kohno et al., *Acta Arith.*, **134** (2008), 329–335.
- [46] R. Tijdeman, Diophantine equations and Diophantine approximations, in: *Number Theory and Applications*, Kluwer Acad. Press, 1989, pp. 215–243.

- [47] R. Tijdeman, Exponential diophantine equations 1986–1996, in: *Number Theory*, de Gruyter, 1998, pp. 523–539.
- [48] R. Tijdeman, Highlights in the research work of T. N. Shorey, in: *Diophantine Equations*, Tata Inst. Fund. Research, Narosa Publ. House, New Delhi, 2008, pp. 1–18.
- [49] G. N. Watson, The problem of the square pyramid, *Messenger Math.*, **48** (1919), 1–22.
- [50] A. Wiles, Modular elliptic curves and Fermat’s last theorem, *Ann. of Math. (2)*, **141** (1995), 443–551.
- [51] B. Xia and T. Cai, A new kind of diophantine equations, *Acta Arith.*, **147** (2011), 245–251.
- [52] S. Yang, A. Togbé and A. B. He, Diophantine equations with products of consecutive values of a quadratic polynomial, *J. Number Theory*, **131** (2011), 1840–1851.
- [53] P. Yang and T. Cai, On the Diophantine equation $\binom{n}{k_1, \dots, k_s} = x^l$, *Acta Arith.*, **151** (2012), 7–9.

Kálmán Győry

*University of Debrecen,
Institute of Mathematics,
H-4010 Debrecen,
P.O. Box 12,
Hungary*

e-mail: gyory@science.unideb.hu

ERDŐS'S WORK ON INFINITE GRAPHS*

PÉTER KOMJÁTH

The theory of infinite graphs was one of Erdős's favorite topics, and it is no exaggeration to state that the major results and notions were created by him and his collaborators. As one of the few persons equally versed in finite as well as in infinite sets, upon hearing a result on finite graphs, he always eagerly checked if it has a reasonable counterpart for infinite graphs.

Here we give an overview of his work on this topic, describing the later developments.

András Hajnal wrote his recollections on Erdős's work in set theory in [34].

1. Erdős's earliest work in the topic of infinite graphs was the infinite generalization of Menger's theorem. He learned Menger's theorem as a freshman at the university, in the class of Dénes König, in 1931. This states that if A and B are two vertices in a finite graph, which are not joined, then the minimal number of vertices separating A and B equals the maximal number of $A - B$ paths, pairwise vertex disjoint, except at their extremities. König asked if this equality held for infinite graphs, as well, and Erdős proved this overnight. The proof was then included in König's 1936 monograph on graphs ([54]).

Several years after Erdős, who was not satisfied with the above result, found the 'right' form of the generalization of Menger's theorem. The conjecture, one of Erdős's finest, states the following. If X is a (finite or infinite) graph, A and B are two vertices which are not joined, then there are a system \mathcal{P} of disjoint $A - B$ paths, and a set S which separates A and B , further, each vertex in S is on a path in \mathcal{P} , and each path in \mathcal{P} meets S in exactly one element. In other words, incidence gives a bijection between \mathcal{P} and S . It is easy to see that for X finite this is equivalent to Menger's original theorem. The first occurrence of the conjecture seems to be [14].

*Research supported by the Hungarian National Research Grant OTKA K 81121.

The proof of the conjecture was the work of Ron Aharoni, who devoted several decades to it. First he proved the general form of König's following theorem: in each bipartite graph there are a set I of independent edges and a covering set C of vertices such that each edge in I meets C in exactly one vertex ([1]). This is easily seen to be a special case of Erdős's general conjecture. The fairly involved proof uses the result of Aharoni et al. on the matching of infinite bipartite graphs. Then, in 1987, he showed the general Menger theorem for countable graphs ([2]). The full proof of the conjecture required another 20 years' work until 2009, when Aharoni with his young collaborator, Eli Berger, proved it with a sophisticated, and hard, argument ([3]).

2. Another early result of Erdős and co. gives an infinite generalization of Euler's famous result giving a sufficient and necessary condition for the existence of an Euler line, i.e., a cycle passing through all edges exactly once.

In his recollections, Endre Vázsonyi described, how quickly Erdős found the right argument.

Theorem (Erdős–Gallai–Vázsonyi, 1936, [16], [17]). *An infinite graph X possesses an Euler-line exactly if*

- (a) X is connected;
- (b) X is countable;
- (c) no vertex has odd degree;
- (d) if A is a finite set of vertices, then $X - A$ has at most two infinite components;
- (e) if A is a finite set of vertices, such that in $X|A$ all vertices have even degree, then the graph obtained from X by removing the edges in A has exactly one infinite component.

3. Also an early, nice result is the Erdős-Kakutani theorem: the complete graph on \aleph_1 vertices (K_{\aleph_1} , that is) is the union of countably many forests, i.e., circuitless graphs ([26]). When first heard, this statement looks trivially false. The proof is not especially hard, the result is one more variant to the theme of 'large sets are the union of a few small sets' (Sierpiński's decomposition of the plane, the reals can be the union of countably many sets, independent over the rationals, the latter was also proved in the Erdős-Kakutani paper).

4. A further result, still from the early period of Erdős's career is the theorem obtained with de Bruijn, stating that the fact that a graph can be good colored with k colors depends on the finite subgraphs (k is finite).

Theorem (de Bruijn–Erdős [7]). *If k is a natural number, then an infinite graph can be good colored with k colors if and only if each of its finite subgraphs can.*

This reduces the theory of infinite graphs with finite chromatic number to the theory of finite graphs, as the chromatic number of some graph X is k , in short, $\text{Chr}(X) = k$, if and only if k is the maximum of the chromatic number of finite subgraphs of X . The result has several proofs; well ordering the vertex set and coloring the vertices by transfinite recursion, with the Teichmüller–Tukey lemma, with the Zorn lemma (Gabriel Dirac and Lajos Pósa, cf. [56], Problem 9.14.). De Bruijn and Erdős used Tychonoff's compactness theorem on the product of topological spaces. These are not that surprising as the theorem is a special case of Gödel's compactness theorem. As Rado and others pointed out, several other statements can be proved with similar so called compactness arguments: if k is finite, an infinite graph can be directed so that each out-degree is at most k iff this holds for all finite subgraphs, if the edges of every finite subgraph can be 2-colored with no monochromatic triangle, then the whole graph can similarly colored, etc.

5. An extremal graph theory problem on countably infinite graphs was raised by Czipser, Erdős, and Hajnal in [6]. Let X be a graph on the set ω of natural numbers which does not contain an increasing path of length k (i.e., one with k edges and $k + 1$ vertices, $k \geq 2$). Let $e_X(n)$ denote the number of edges of X on the first n natural numbers, and set

$$p(X) = \liminf_{n \rightarrow \infty} \frac{e_X(n)}{n^2}.$$

Finally, let $p(k) = \sup p(X)$, where the supremum is taken for all graphs without increasing paths of length k . The definition of $p(\infty)$ is analogous but with forbidding infinite increasing paths. They investigated the values of $p(k)$ for various values of k . It is easily seen that $p(2) \leq p(3) \leq \dots \leq p(\infty)$ and Turán's theorem implies

$$p(k) \leq \frac{1}{2} \left(1 - \frac{1}{k} \right)$$

for $k \geq 2$. They proved that $p(2) = \frac{1}{8}$, $p(3) = \frac{1}{6}$, and conjectured the general equality

$$p(k) = \frac{1}{4} \left(1 - \frac{1}{k} \right).$$

They also proved the bounds

$$\frac{1}{4} + \frac{1}{36} \leq p(\infty) \leq \frac{1}{4} + \frac{3}{16}.$$

The baton was taken up some 50 years later, when Dudek and Rödl proved several surprising results concerning this question in [11]. They disproved the Czipser-Erdős-Hajnal conjecture by establishing

$$p(16) > \frac{1}{4} \left(1 - \frac{1}{16} \right).$$

and

$$p(k) > \frac{1}{4} + \frac{1}{20} \quad (k \geq 162).$$

They also proved the upper bound $p(k) \leq \frac{1}{3}$ for every $k \geq 2$ and noticed that $p(k) < p(\infty)$ holds for any finite k .

6. Several important discoveries of Erdős stemmed out from the following observation of Tutte, Zykov, and Ungar: there are arbitrarily large chromatic finite triangle-free graphs. Erdős extended this with his signature method, random graphs, to showing that for every s and k there is a finite graph, whose chromatic number is k and omits C_3, C_4, \dots, C_s , that is, all short circuits ([13]). In fact, this was one of the first spectacular applications of the random methods. This is much harder than the corresponding result for triangle-free graphs: it took several years until the first explicit construction was given by Lovász.

As I have already mentioned, Erdős tried to extend all finite graph theory results to the infinite and this result was no exception. With Richard Rado they constructed for any infinite cardinal κ a triangle-free graph whose chromatic number was larger than κ (and of cardinality 2^κ , [27]). For a while, Erdős tried, in vain, to find the common generalization of these results, until, with András Hajnal, they discovered the surprising fact that uncountably chromatic graphs contain C_4 , the four-circuit, even all finite bipartite graphs.

The proof led Erdős and Hajnal to the invention of the notion of coloring number. The *coloring number*, $\text{Col}(X)$ of a graph X is the least cardinal μ such that there is a well ordering of the vertex set in which each vertex admits $< \mu$ edges going down. Notice that the chromatic number, $\text{Chr}(X)$ is likewise a minimum: the least number of colors needed to good color the vertices of X .

If $\text{Col}(X) = \mu$, then a good coloring can be defined along the well ordering witnessing this, by transfinite recursion. This shows $\text{Chr}(X) \leq \text{Col}(X)$.

Erdős and Hajnal proved that if $\text{Col}(X) > \omega$ then X contains all finite bipartite graphs. As there are bipartite graphs with arbitrarily large coloring number, this method cannot be used to give more on the finite subgraphs of large chromatic graphs.

The coloring number of a graph is close to the number of forests needed to cover the edges. In particular, if $\text{Col}(X) \leq \kappa^+$ then X is the union of κ forests where κ can be finite or infinite. For infinite κ this can be reversed, for finite values we have the following: if a graph is the union of n forests (n finite), then its coloring number is at most $2n$ and this is sharp ([19]).

7. After this discovery, Erdős and Hajnal could easily find and prove the correct uncountable version of the theorem on the existence of large chromatic finite graphs omitting short circuits: for each finite n there are arbitrarily large (infinite) chromatic graphs omitting $C_3, C_5, \dots, C_{2n+1}$.

They found a simple, but powerful construction: the edge-graph. For this, let κ be an infinite cardinal and consider an ordered set $(A, <)$ of cardinality $(2^\kappa)^+$. The vertices of the edge-graph $X(A, <)$ are the pairs $\{x, y\}$ of the elements of A , and join $\{x, y\}$ with $\{y, z\}$ exactly if $x < y < z$. It is readily seen that $X(A, <)$ is triangle-free. A coloring of the vertices of $X(A, <)$ with κ is a coloring of the pairs of A , and by the Erdős-Rado theorem there is a monochromatic triangle, that is, there are $x < y < z$ such that $\{x, y\}$, $\{x, z\}$, and $\{y, z\}$ get the same color, and so $\{x, y\}$, $\{y, z\}$ are two vertices of $X(A, <)$ which are joined and get the same color. We obtained that the chromatic number of $X(A, <)$ is greater than κ . (This very example, using the finite Ramsey theorem, can be used to give very simple examples of finite, large chromatic, triangle-free graphs: the edge-graph of a sufficiently large finite ordered set.)

Erdős and Hajnal generalized this construction to obtain large chromatic graphs omitting $C_3, C_5, \dots, C_{2n+1}$, that is, all small odd circuits. Their construction can be described as follows. Let $X = (V(X), E(X))$ be a graph on the ordered set $V(X)$. We construct the following graph $X' = (V(X'), E(X'))$. The vertex set $V(X')$ of X' is $E(X)$ and we join the successive edges, that is, $\{x, y\}$ and $\{y, z\}$ are joined if $x < y < z$. It is easy to see that if X does not contain $C_3, C_5, \dots, C_{2n-1}$, then X' does not have $C_3, C_5, \dots, C_{2n+1}$, further, as was observed by Galvin ([31]), the chromatic number of X is greater than (2^κ) iff the chromatic number of X' is greater than κ .

This way, if X is the complete graph of cardinality $\lambda = (2^{2^{\dots^\kappa}})^+$, then $X^{\dots'}$ is a graph omitting $C_3, C_5, \dots, C_{2n+1}$, with chromatic number larger

than κ (n powers and primes, respectively) ([20]). It is easy to show, that the graph so obtained is the so called n -shift graph: the vertices of $\text{Sh}_n(\lambda)$ are the $(n+1)$ -tuples of a well ordered set of cardinality λ , the edges are the consecutive $(n+1)$ -sets: those pairs of the form

$$\{\{x_0, \dots, x_n\}, \{x_1, \dots, x_{n+1}\}\}$$

where $x_0 < x_1 < \dots < x_{n+1}$.

Similar to the shift graphs, but smaller are the so called *Specker graphs*. They are defined as follows. Let κ be an uncountable cardinal. The vertex set will be $[\kappa]^n$, that is, the set of all n -element subsets of κ .

We join two (increasingly enumerated) sets, $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_n\}$ exactly if they are disjoint and they are in a prescribed interlacing pattern. For example, a possibility for $n = 3$:

$$x_1 < x_2 < y_1 < x_3 < y_2 < y_3$$

(Specker's original example).

It can be proved that the chromatic number is κ and with an appropriate choice of the interlacing pattern we can exclude various finite graphs, for example, all odd circuits up to a certain bound. Notice that here the cardinality of the graph equals the chromatic number: both are κ ([18]).

8. Rado asked if the de Bruijn-Erdős phenomenon holds for the coloring number. Surprisingly, this is not true: as Erdős and Hajnal showed, there is a countable graph with coloring number 4, all whose finite subgraphs have coloring number at most 3. In general, if the finite subgraphs have coloring number at most k , then the graph has coloring number at most $2k - 2$, and this is sharp (Erdős-Hajnal, [18]).

Eric Milner noticed that the above statement makes the following conjecture nontrivial: if k is finite, and X is a graph with $\text{Col}(X) = k + 1$, then there is a subgraph Y with $\text{Col}(Y) = k$. This was then proved in [44].

In a different sense, however, the coloring number does satisfy compactness: if $\lambda > \mu$ are cardinals, λ is a singular cardinal, X is a graph of cardinality λ , all whose smaller subgraphs have coloring number at most μ , then so does X . This was first observed in some cases by Erdős and Hajnal, then fully proved by Shelah. ([62]). This makes possible to give a particularly simple characterization of graphs with large coloring number with applications as the full description of graphs obligatory for graphs with uncountable coloring number or the proof of the consistency that every graph with uncountable coloring number contains a subgraph with size and coloring number \aleph_1 (Komjáth, [39]).

Shelah's result is more general: he axiomatized those structure classes for which a similar singular cardinal compactness holds (he was mainly interested in what uncountable cardinals λ are there Abelian groups of cardinality λ which are not free but every smaller subgroups are).

I succeeded in proving that this fails for the chromatic number: it is consistent that there is a graph X of cardinality \aleph_{ω_1} which is uncountably chromatic yet all smaller subgraphs are countably chromatic (necessarily $\text{Chr}(X) = \aleph_1$, [40]). With an elegant forcing argument Shelah proved that this may be possible under GCH, and also that there are examples in L ([64]).

9. Erdős and Hajnal therefore determined those finite graphs which necessarily appear in all uncountably chromatic graphs—these are the finite bipartite graphs. On the other hand, some odd circuits must appear (as otherwise the chromatic number would be ≤ 2). Erdős and Hajnal then started to investigate what can be said on the classes of finite graphs which must be contained in uncountably chromatic graphs. For example, they conjectured that all sufficiently long odd circuits must appear. This was later proved by Erdős-Hajnal-Shelah ([24]) and Thomassen [67], independently.

If X is an uncountably chromatic graph, let $\mathcal{F}(X)$ denote the family of finite subgraphs of X . A reasonable conjecture was that if X is uncountably chromatic, then there are arbitrarily large chromatic graphs Y with $\mathcal{F}(Y) = \mathcal{F}(X)$ (Taylor's conjecture, cf. [24]). A very bold claim even stated that $\mathcal{F}(X)$ is contained in $\mathcal{F}(\text{Sh}_2(\omega))$ for some n . The latter statement was never seriously believed and we disproved it without much ado in [35].

Later I disproved Taylor's conjecture, too. In [53] I gave a forcing model with a graph X with $\text{Chr}(X) = \aleph_1$ (and of cardinality \aleph_1) such that if $\mathcal{F}(Y) = \mathcal{F}(X)$, then $\text{Chr}(Y) \leq \aleph_2$. In a different model, the following holds. If $\text{Chr}(X) > \aleph_2$ then for every cardinal λ there is a graph Y with $\text{Chr}(Y) > \lambda$ such that $\mathcal{F}(Y) = \mathcal{F}(X)$.

10. Erdős invented two more nice conjectures on the finite subgraphs of uncountably chromatic graphs.

If X is a graph, construct the following function mapping natural numbers to natural numbers: $f_X(n)$ is the maximum of the chromatic number of n -vertex induced subgraphs. Clearly, $f_X(n) \leq n$ and f_X is weakly increasing. Further, the de Bruijn-Erdős theorem implies that $f_X(n) \rightarrow \infty$. What Erdős asked if this divergence can be arbitrarily slow for uncountably chromatic X . Erdős, Hajnal, and Szemerédi proved in [25] that for the shift graph $\text{Sh}_k(\lambda)$ $f_X(n)$ is essentially the $(k - 1)$ -fold iterated logarithm.

The consistency of Erdős's conjecture was finally proved by Shelah using an extraordinarily clever argument ([53]).

Erdős, Hajnal, and Szemerédi investigated several variants of f_X in their paper [25]. For example, they showed, that in a shift graph $\text{Sh}_k(\lambda)$ every n -element vertex set contains a bipartite graph containing $(1 - \frac{2}{k})n$ vertices. Consequently, if we define $f_X^1(n)$ as the largest size of an independent set that can be found in any n -vertex subset of the graph X , then for every $\varepsilon > 0$ there are arbitrarily large chromatic graphs, for which $f_X^1(n) > (\frac{1}{2} - \varepsilon)n$ holds. In the other direction, it is easy to see that for every uncountably chromatic graph X there is some $\varepsilon > 0$, such that for all sufficiently large n , $f_X^1(n) < (\frac{1}{2} - \varepsilon)n$ holds, as there is some m , that there are infinitely many vertex-disjoint C_m 's and if take k of them, then the largest independent set has at most km vertices. There remains the question if there exists a graph X of size and chromatic number \aleph_1 , for which $f_X^1(n) > cn$ holds for some appropriate $c > 0$. (The shift graph has cardinality $(2^{\aleph_0})^+$.) The Specker graphs are no good, either, for them

$$f_X^1(n) = O\left(\frac{n \log \log n}{\log n}\right)$$

holds.

In [25] they also investigated the following function $g_X(n)$. If X is an uncountably chromatic graph, n is finite, let $g_X(n)$ be the least number of edges, whose removal makes any n -vertex set bipartite. They proved that if X is the edge graph, then $g_X(n) \leq 2n^{3/2}$ and with the help of the general shift graphs, for every $\varepsilon > 0$ there is an uncountably chromatic graph X such that $g_X(n) = O(n^{1+\varepsilon})$.

11. The other conjecture is the following. Any two uncountably chromatic graphs have a common 4-chromatic subgraph. The analogous claim for 3-chromatic common subgraphs follows easily from the above mentioned Erdős-Hajnal-Shelah-Thomassen result—both graphs contain all sufficiently long odd circuits. What is particularly interesting about this conjecture is that it states a hard claim without telling what 4-chromatic subgraphs must be contained in uncountably chromatic graphs, a question we do not even have a reasonable guess about.

A related topic is the the chromatic number of Cartesian products of graphs. Here, if $X_0 = (V_0, E_0)$ and $X_1 = (V_1, E_1)$ are two graphs we define their Cartesian product $X_0 \times X_1$ as the graph with vertex set $V_0 \times V_1$ and edge set

$$\{(x_0, x_1), (y_0, y_1)\} : \{x_0, y_0\} \in E_0, \{x_1, y_1\} \in E_1\}.$$

It is immediately seen that

$$\text{Chr}(X_0 \times X_1) \leq \min(\text{Chr}(X_0), \text{Chr}(X_1)).$$

Hajnal ([32]) proved that we have equality here, if $\text{Chr}(X_0)$ is finite and $\text{Chr}(X_1)$ is infinite. He also showed the surprising result that there are graphs X_0, X_1 with $\text{Chr}(X_0) = \text{Chr}(X_1) = \aleph_1$, but $\text{Chr}(X_0 \times X_1) = \aleph_0$. Later Soukup proved that it is consistent with GCH that there exist graphs X_0, X_1 with $\text{Chr}(X_0) = \text{Chr}(X_1) = \aleph_2$ and $\text{Chr}(X_0 \times X_1) = \aleph_0$ ([65]).

12. An even harder question is which countable graphs must be contained in uncountably chromatic graphs, in other words, which are the countable obligatory graphs. Erdős and Hajnal proved in [18], that each graph with uncountable chromatic number, or even coloring number contains K_{n, \aleph_1} for each finite n .

Hajnal later showed that the ‘halfgraph’ must be contained, where the halfgraph is the graph with vertex set $\{x_i, y_i : i < \omega\}$ where x_i is joined to y_j if $i < j$ iff $i < j$. Earlier Hajnal proved that there is an uncountably chromatic graph omitting $K_{\omega, \omega}$, the complete countable bipartite graph.

These results were somewhat extended in [35]:

Theorem (Hajnal–Komjáth, [35]). *Each uncountably chromatic graph contains X_0 but not necessarily X_1 , where the vertex set of X_0 is $\{x_i, y_i, z : i < \omega\}$, with y_i joined to x_j ($j < i$) and z is joined to each x_j . X_1 is similar, but with two vertices joined to every x_j .*

This eventually led to the complete description of all graphs appearing in every graph with uncountable coloring number ([39]), but the corresponding problem for the chromatic number stays open.

Unsolved remains the following nice conjecture of Erdős. If X is an uncountably chromatic graph then X contains an ω -connected subgraph. This was motivated by the result of Erdős and Hajnal that uncountably chromatic graphs contain the complete bipartite graph $K_{n, n}$, which, as it is easily seen, is n -connected ($n < \omega$). I showed that even an n -connected uncountably chromatic subgraph can be found ([38]), but the statement that \aleph_1 -chromatic graphs of cardinality \aleph_1 contain ω -connected \aleph_1 -chromatic subgraphs is consistent as well as independent ([40], [48]).

13. The following Erdős conjecture also concerns finite subgraphs. Having learned that there existed an uncountably chromatic graph omitting $K_{\omega, \omega}$, Erdős promptly asked if it can simultaneously omit triangles. This was later proved by Hajnal. Erdős then noticed that it would also follow from the following statement. Every uncountably chromatic graph has

an uncountably chromatic, triangle free subgraph. Considering the above statements, Erdős had stronger conjectures: if the chromatic number of the graph X is $\kappa > \aleph_0$, then for every finite n it has an uncountable chromatic subgraphs omitting $C_3, C_5, \dots, C_{2n+1}$, if, however the graph has chromatic number \aleph_0 , then there is a subgraph with infinite chromatic number, that omits $C_3, C_4, C_5, \dots, C_n$. The simplest case of the latter conjecture, omitting only triangles, was established by Rödl ([61]), the general case stays unresolved. The claim for the uncountably chromatic graphs was disproved by Shelah: consistently there is a graph with size and chromatic number \aleph_1 , all whose triangle free subgraphs are countably chromatic. I discovered some extensions of this: for example, the large graph can omit K_4 ([51]).

14. A problem, which was not discovered by Erdős, but he liked it very much, asks if the chromatic number satisfies the Darboux property. He repeatedly asked it in his problem papers. Fred Galvin observed that if $\kappa < \lambda$ are finite then every λ -chromatic graph contains a κ -chromatic subgraph, an easy corollary of the de Bruijn-Erdős theorem. This, however, is far from being obvious if κ and λ are infinite. More exactly, the case $\kappa = \aleph_0 < \lambda$ can also be deduced from the de Bruijn-Erdős theorem, so the first open case is $\kappa = \aleph_1, \lambda = \aleph_2$. In his paper [31], Galvin proved that the answer is consistently false, at least, if we restrict to induced subgraphs. If $2^{\aleph_0} = 2^{\aleph_1} < 2^{\aleph_2}$ holds (this can be arranged by Cohen's method of forcing), then the edge-graph on an ordered set of cardinality 2^{\aleph_2} satisfies the following: its chromatic number is \aleph_2 , all its induced subgraphs are of the form X' for some edge-graph X , and $\text{Chr}(X') \leq \kappa$ iff $\text{Chr}(X) \leq 2^\kappa$. If, therefore, $\text{Chr}(X') \leq \aleph_1$, then $\text{Chr}(X) \leq 2^{\aleph_1} = 2^{\aleph_0}$, then $\text{Chr}(X') \leq \aleph_0$, and so no induced subgraph exists whose exact chromatic number is \aleph_1 .

Later I succeeded in showing that it is consistent that there is a graph of size and chromatic number \aleph_2 , which does not contain a subgraph of exact chromatic number \aleph_1 (induced or not, [40]).

This has been further extended in [46]. If X is a graph with uncountable chromatic number, then let $S(X)$ denote the set of chromatic numbers of the subgraphs of X . Similarly, let $I(X)$ denote the set of chromatic numbers of all induced subgraphs of X . We remove the natural numbers and \aleph_0 from both sets as they are of no interest for us. Clearly, both $I(X)$ and $S(X)$ are sets of uncountable cardinals, both with $\text{Chr}(X)$ as the maximal element. Are there other conditions? In [46] we prove that $I(X)$ is closed under taking limits, and if $\lambda \in I(X)$ is a singular cardinal, then it is a limit point of $I(X)$. Conversely, if A is a nonempty set of uncountable cardinals satisfying these properties then there is a cardinal preserving forcing that adds a graph X with $I(X) = A$. For $S(X)$, we have that if $\lambda \in S(X)$ is a singular cardinal, then λ is a limit point of $S(X)$, and $S(X)$ contains its

limit points which are singular. However, a complicated forcing argument adds a graph X such that $S(X)$ is not closed at a regular cardinal.

15. Erdős and Hajnal observed the following. If the continuum hypothesis holds, then there is an \aleph_1 -chromatic graph X of cardinality \aleph_2 , such that every subgraph of X of cardinality \aleph_1 is countably chromatic. X is the edge-graph of an ordered set of cardinality \aleph_2 . They immediately asked (and repeated their question in their famous problem paper, [21]) if the chromatic number of X can be \aleph_2 ? This question remained unanswered for quite a while, then the consistency of both directions was shown*. Baumgartner proved ([4]) that consistently there is such a graph, while Foreman and Laver proved in [30] that if the existence of a so called huge cardinal is consistent, then so is that no graph as above exists. Both proofs are involved. Later Shelah deduced the existence of such a graph from the axiom of constructibility ([64]).

Erdős and Hajnal made the following interesting observation concerning the first construction ([20]). Let $G(\omega_2, \omega)$ be the following graph: its vertex set consists of all $\omega_2 \rightarrow \omega$ functions with f and g joined if they eventually differ, that is, if $f(\alpha) \neq g(\alpha)$ holds for all sufficiently large $\alpha < \omega_2$. They proved the following properties:

- (a) every subgraph of $G(\omega_2, \omega)$ of cardinality at most \aleph_1 is countably chromatic;
- (b) if X is a graph of cardinality \aleph_2 , all whose subgraphs of cardinality \aleph_1 are countably chromatic, then X embeds into $G(\omega_2, \omega)$.

If the continuum hypothesis holds, then, by the above remark, we do have a graph of cardinality \aleph_2 and chromatic number \aleph_1 , all whose smaller graphs are countably chromatic. By (b), it embeds into $G(\omega_2, \omega)$, consequently the chromatic number of $G(\omega_2, \omega)$ is uncountable (its cardinality is $\aleph_0^{\aleph_2} = 2^{\aleph_2}$). What, exactly, is the value of $\text{Chr}(G(\omega_2, \omega))$? I reached some partial answers in [42]: in different models of the generalized continuum hypothesis it can be both \aleph_2 and \aleph_3 . Foreman deduced from the consistency of the existence of a huge cardinal, that it is consistent that $\text{Chr}(G(\omega_2, \omega)) = \aleph_1$ ([29]). Finally, Todorčević proved that $G(\omega_2, \omega)$ is always uncountably chromatic ([69], [70]).

16. Ramsey's theorem and the investigation of Ramsey type phenomena was always in the focus of Erdős' work. The simplest case of Ramsey's theorem states that if all pairs of a six-element set are colored with two colors then there is a monochromatic triangle. With the usual notation this is denoted as $6 \rightarrow (3)_2^2$. Here, on the right hand side of the formula, the top

*Independence raised its ugly head, as Erdős frequently remarked.

2 stands for the size of the sets colored, the down 2 is the number of colors. That the statement fails for 5 is denoted by $5 \not\rightarrow (3)_2^2$.

The infinite Ramsey theorem is the following. If r, n are positive integers, then $\aleph_0 \rightarrow (\aleph_0)_n^r$ holds, that is, if the r -tuples of an infinite set are colored with n colors, then there is an infinite monochromatic (homogeneous) set.

The infinite versions of this statement were considered by Erdős, Hajnal, and Richard Rado in the fifties and sixties, giving rise to the so called *partition calculus*. The case $r = 2$ can be considered a purely set theoretical result: if the edges of a countably infinite graph are colored with 2 color, then there is an infinite homogeneous subgraph. With his collaborators, Erdős passionately investigated the graph theoretic generalizations of Ramsey's theorem. Already in 1942 he settled the simplest question: if we are given a sequence $\langle X_\alpha : \alpha < \kappa \rangle$ of graphs, each of cardinality at most κ^+ , then there is a graph Y with the property that if the edges are colored with κ colors, then for some color α there is a copy of X_α all whose edges are colored by the α 'th color. This follows from the following partition result of Erdős (usually called the Erdős-Rado theorem, although it appeared in [12]):

$$(2^\kappa)^+ \rightarrow (\kappa^+)_\kappa^2.$$

This proves the result for complete graphs, the general case follows.

17. In order to bypass trivial arguments, some restrictive versions were sought, for which it is not true that if X is a subgraph of Y , then a positive answer for Y gives also a positive answer for X (as then it suffices to consider complete graphs). One such version originates from Walter Deuber, who required induced target graphs. We denote the relevant statement by $Y \mapsto (X_\alpha : \alpha < \kappa)^2$ and if all X_α 's are equal to X , then we write $Y \mapsto (X)_\kappa^2$. Here even the simplest statement is very hard to prove: if X is a finite graph, then there is a finite graph Y such that $Y \mapsto (X)_2^2$ holds, that is, if the edges of Y are colored with two colors, then there is an induced monochromatic copy of X . This was independently proved by W. Deuber [8], [9], V. Rödl [60], and Erdős-Hajnal-Pósa [22]. Deuber and Rödl used sophisticated induction. Erdős and his coauthors, however, proved much more.

Theorem (Erdős-Hajnal-Pósa, [22]).

- (a) If X_0, X_1 are countable graphs and X_0 is locally finite, then there is a countable Y , for which $Y \mapsto (X_0, X_1)^2$ holds.
- (b) If Y is a countable graph, then $Y \not\mapsto (K_{\omega, \omega})^2$.
- (c) If X_0, X_1, \dots, X_k are finitely many countable graphs, then there is a graph Y of cardinality continuum for which $Y \mapsto (X_0, \dots, X_k)^2$ holds.

As concerning infinite graphs, it was naturally conjectured that the full Ramsey property holds, i.e., whenever X is a graph, κ a cardinal, then there is a graph Y satisfying the relation $Y \mapsto (X)_\kappa^2$. With Hajnal we gave a surprising negative answer in [36]: it is consistent that there is a graph X of cardinality \aleph_1 such that for no graph Y does $Y \mapsto (X)_2^2$ hold. Shelah was quick to add the consistency of the positive direction; consistently for every graph X and every cardinal κ there is some graph Y with $Y \mapsto (X)_\kappa^2$ (in fact, the general Ramsey theorem holds for arbitrary structures, [63]). I later showed that it is no accident that Shelah uses class forcing—in every model obtained by a nontrivial set forcing there are a graph and a cardinal κ such that $Y \not\mapsto (X)_\kappa^2$ holds for every graph Y ([43]). Finally Hajnal proved the following deep and very satisfactory theorem. If X is a finite graph and κ is an infinite cardinal, then there is a graph Y , for which $Y \mapsto (X)_\kappa^2$ holds ([33]).

A notorious problem of this subtopic if Hajnal's theorem can be extended to countable graphs. The conjecture is that if X is a countable graph and κ is a cardinal, then there is a graph Y such that $Y \mapsto (X)_\kappa^2$ holds. As by a theorem of Rado's there is a universal countable graph, i.e., one that includes all other as induced graphs, it suffices to prove the conjecture for that graph only. Further, as the Rado graph cannot be changed by forcing, it is hopeless to use the methods of [36] to force a counterexample as the target graph.

18. A different possibility to make harder the edge coloring problem is the following. Start with the simplest statement, that is, that if we color the edges of the complete graph K_6 with two colors, then there must be a monochromatic triangle, that is, $K_6 \rightarrow (K_3)_2^2$. Obviously, this property holds for every graph containing K_6 as a subgraph: if $K_6 \leq X$, then $X \rightarrow (K_3)_2^2$. Erdős and Hajnal asked in 1967, if there is a graph omitting K_6 for which $X \rightarrow (K_3)_2^2$ holds. This was quickly answered in the positive by G. L. Cherlin, R. Graham, van Lint, and Lajos Pósa. Pósa's example did not contain a K_5 , either, but Erdős's next question, namely, if we can omit K_4 , turned out to be very hard, and it was finally answered in the positive with a deep and complicated construction by Jon Folkman [28].

Folkman's example was gigantic, with more than

$$10^{10^{10^{10^{10^{10}}}}}$$

vertices, and Erdős was always curious if this can be lowered to some more "human" bound, say to 10^{10} . Frankl and Rödl constructed a graph with some 10^{11} vertices, this was fine tuned by Spencer to an example with a

few hundred million vertices ([66]). In the years after 2000 this was lowered to around 10,000 by L. Lu, then Dudek and Rödl ([10]) gave the surprising upper bound 941, this has recently been improved to 786 ([55]).

Nešetřil and Rödl proved the corresponding general partition theorem; if X is a finite graph, containing no clique K_p , $k \geq 2$ is a natural number, then there is a graph Y , with no clique K_p , satisfying $Y \twoheadrightarrow (X)_k^2$ ([57], [58]).

Erdős and Hajnal methodically investigated the infinite cases. They proved the following results.

Theorem (Erdős–Hajnal [19]).

- (a) *If κ is infinite, n is finite, then there is a graph of cardinality κ , omitting K_{n+1} which has the property that if its vertices are colored with κ colors, then there is a monochromatic K_n .*
- (b) *If κ, λ are infinite cardinals, then there is a graph of cardinality κ^λ that omits K_{λ^+} and whenever its vertices are colored with κ colors, then there is a monochromatic K_λ .*
- (c) *If κ is an infinite cardinal, then there is a graph of cardinality*

$$(2^{(2^\kappa)^+})^+$$

omitting K_{\aleph_0} , such that every edge-coloring with κ colors contains a monochromatic K_n , for every finite n .

- (d) *if κ is an infinite cardinal, then there is a graph of cardinality $(2^\kappa)^+$, omitting $K_{(2^\kappa)^+}$, such that each edge-coloring of the edges with κ colors gives rise to a monochromatic K_{κ^+} .*

The proof of (d) is based on the partition relations $(2^\kappa)^+ \twoheadrightarrow ((2^\kappa)^+, (2^\kappa)^+)^2$ and $(2^\kappa)^+ \rightarrow ((2^\kappa)^+, (\kappa^+)^\kappa)^2$.

We can slightly improve the claim with the price of increasing the size. For concreteness' sake assume that $\kappa = \aleph_0$. There is a cardinal λ satisfying $\lambda = \lambda^{\aleph_0} < \lambda^{\aleph_1}$. Partition theory gives $\lambda^+ \rightarrow (\lambda^+, (\omega_1)_\omega)^2$ and $\lambda^{\aleph_1} \twoheadrightarrow (\lambda^+, \omega_2)^2$, and therefore $\lambda^+ \twoheadrightarrow (\lambda^+, \omega_2)^2$. One class of the latter coloring is a graph without K_{\aleph_2} , with no independent λ^+ , and, by another theorem, all colorings of it by countably many colors contain a monochromatic K_{\aleph_1} .

A simpler proof of (c) with the weaker bound

$$\lambda = (2^{2^{2^\kappa}})^+$$

can be obtained as follows. Choose the set of pairs of λ as the vertex set of the graphs, join $\{x, y\}$ and $\{x', y'\}$ exactly if $x < x' < y' < y$. There is

no K_{\aleph_0} as it would lead to an infinite decreasing sequence of ordinals, the coloring statement follows from the partition relation $\lambda \rightarrow (2n)_\kappa^4$ (Erdős-Rado theorem).

The most important open questions are the following. Does there exist a graph omitting K_4 , which, when edge-colored with countably many colors, always contains a monochromatic triangle, that is, $K_4 \not\rightarrow X \rightarrow (K_3)_{\aleph_0}^2$?

This was one of Erdős' favorite problems, he regularly mentioned it in his lectures, problem papers. He promised \$ 250 for the solution. Shelah in [63] established the consistency of this statement, i.e., that forcing can give a graph as required. His result has an interesting corollary. If $K_4 \not\rightarrow X$ and $X \rightarrow (K_3)_{\aleph_0}^2$, then X satisfies the weaker relation $K_4 \not\rightarrow X$ and $X \rightarrow (K_3)_2^2$, as well. Using the compactness principle mentioned at the de Bruijn-Erdős theorem, it follows, that X contains a finite subgraph Y satisfying $K_4 \not\rightarrow Y$ and $Y \rightarrow (K_3)_2^2$. As forcing does not add new finite graphs, we obtain that each countable model of set theory contains a graph like that, which gives, using Gödel's completeness theorem, that outright there is a Folkman-type graph. Now what is it? There does not seem to be any way of transforming this proof into a construction of such a graph.

Curiously, there is another, different set theoretical proof of the existence of a finite graph X with $K_4 \not\rightarrow X$ and $X \rightarrow (3)_2^2$. In their paper [5] Baumgartner and Hajnal proved the ordinary partition theorems $\omega_1^2 \rightarrow (\omega_1\omega, 4)^2$ and $\omega_1 \rightarrow (\omega_1\omega, 3, 3)^2$, the first under the continuum hypothesis. The former can be interpreted as showing the existence of a graph X on a ground set of ordinal ω_1^2 such that X contains neither a K_4 , nor an independent set of ordinal $\omega_1\omega$. The latter can be interpreted as the statement that if the edges of a graph on ω_1^2 with no independent set of ordinal $\omega_1\omega$ are two colored, then there is a monochromatic triangle. Putting together, we obtain a K_4 -free graph, with a monocolored triangle in every 2-coloring of the edges—if the continuum hypothesis holds. But CH can be obtained by forcing, so we can conclude as in the previous argument.

Another, still unsolved, question of the paper of Erdős and Hajnal [19], if there exists a graph X containing no K_{\aleph_1} satisfying $X \rightarrow (K_{\aleph_0})_{\aleph_0}^2$.

Extending the above mentioned method of Shelah, we proved in [52] that the full Ramsey theorem for classes omitting cliques is consistent: if X is a graph, μ is a cardinality, and X does not contain a K_α , then there is a graph Y , still omitting K_α such that $Y \rightarrow (X)_\mu^2$ holds. This, as we have shown with Hajnal, is not outright true: it is consistent that there is a triangle-free graph X , such that if $Y \rightarrow (X)_{\aleph_0}^2$ holds for some graph Y , then Y contains a K_{\aleph_0} ([36]).

The vertex coloring version is, however, much easier: given any graph X omitting K_α , and a cardinal μ , then there exists, without any extra set

theoretic assumption, a graph Y , omitting K_α , such that $Y \twoheadrightarrow (X)_\mu^1$ holds, that is, when the vertices of Y are colored with μ colors, then there is a monochromatic induced copy of X ([36]).

19. The swinging sixties was the golden period of hypergraph theory, in finite combinatorics, that is. Not surprisingly, Erdős and his collaborators started generalizing the theory of infinite graphs to infinite hypergraphs, when we consider a system \mathcal{H} of n -element subsets of some ground set S for some finite $n \geq 3$. In this case, the chromatic number of \mathcal{H} is the least number of colors required to a coloring of S without a monocolored member of \mathcal{H} . As even this particular case is extremely hard, most results and questions have been formulated for $n = 3$, i.e., for systems of triples.

In order to formulate some results, let me introduce the ad hoc notion of *romboid*. The system $\{A, B\}$ of two triples is a romboid, if $|A \cap B| = 2$. Already in [18] Erdős and Hajnal proved that if \mathcal{H} is a triple system of cardinality \aleph_1 which omits the romboid, then \mathcal{H} is countably chromatic. This, however, does not mean that all romboid-free triple systems are countably chromatic, as it was pointed out by Erdős, Hajnal, and Bruce Rothschild in [23]. The system they gave is particularly simple: its underlying set consists of all pairs of a set S of cardinality $(2^{\aleph_0})^+$ and the triplets are those of the form $\{\{x, y\}, \{y, z\}, \{x, z\}\}$. It is immediately seen that the system does not contain a romboid, and if the vertices, i.e., the pairs are colored with countably many colors, then there is a monochromatic triangle by the Erdős-Rado theorem

$$(2^{\aleph_0})^+ \rightarrow (3)_{\aleph_0}^2.$$

This result initiated an intensive research period. In [15] Erdős, Galvin, and Hajnal report the result of longer than one year's work. They investigated, for example, how large must a romboid-free, uncountably chromatic triple system be. At least of cardinality \aleph_2 , by the above mentioned Erdős-Hajnal result. It must have more than κ vertices, if Martin's axiom MA_κ holds. In the other direction we find the above $(2^{\aleph_0})^+$ example of Erdős, Hajnal, and Rothschild.

They also proved that if the square bracket partition relation[†] $\omega_1 \twoheadrightarrow [\omega_1]_{\aleph_1}^2$ holds, then there is an example of cardinality 2^{\aleph_1} . That $\omega_1 \twoheadrightarrow [\omega_1]_{\aleph_1}^2$ holds, was originally proved by Erdős, Hajnal, and Rado under the continuum hypothesis. For a while it was an open question if it can be

[†]The relation $\lambda \twoheadrightarrow [\kappa]_\mu^2$ means that there is a coloring of the pairs of a set S of cardinality λ with μ colors such that each subset of S of cardinality κ contains every color. The statements $\lambda \twoheadrightarrow [\kappa]_2^2$ and $\lambda \twoheadrightarrow (\kappa)_2^2$ are therefore equivalent.

proved without any extra assumption, it was long after the writing of [15] that Todorčević proved this using a breakthrough argument ([68]).

As for the general case, Erdős, Galvin, and Hajnal proved that if \mathcal{H} is a set of n -element sets, in which any two members have at most i elements in common and $mi + 2 \leq n$, then, if $|\mathcal{H}| \leq \kappa^{+m}$, then the chromatic number of \mathcal{H} is at most κ . Assuming the Generalized Continuum Hypothesis, this is sharp in the sense, that in any other case there is a counterexample.

The 90-page long paper [15] is densely packed with theorems, estimates, constructions, still several fundamental questions are raised in it. What are the finite triple systems that occur in every uncountable chromatic triple system? We do not even have a conjecture for the answer. Another basic problem if we obtain the same finite systems if here “uncountably chromatic” is changed to “with chromatic number $> \lambda$ ” for any cardinal $\lambda > \aleph_0$? Yet another problem raised in the Erdős-Galvin-Hajnal paper is the following. What is the least cardinal κ for which the following holds: if \mathcal{F} is a finite triple system such that there is a triple system omitting \mathcal{F} , then there is such a system of cardinality $< \kappa$. It is easy to see that such a cardinal κ exists and the above results concerning the omission of the romboid show that the value of κ is not the least possible, i.e., \aleph_2 .

There has been progress, albeit slow, on the topic. In [45] I proved that all finite obligatory triple systems are tripartite, i.e., if \mathcal{F} is a finite obligatory triple system, then there are disjoint finite sets A , B , and C , such that each triple in \mathcal{F} contains exactly one point of A , B , and C . With Hajnal, we proved that consistently there are finite triple systems \mathcal{S}_0 and \mathcal{S}_1 such that each uncountable chromatic triple system contains either \mathcal{S}_0 or \mathcal{S}_1 , but either system can be omitted ([37]). This was one of the problems of [15]. We also proved that every uncountable chromatic, romboid-free triple system contains the odd circuits C_7, C_9, C_{11}, \dots ([37]) and this is sharp in the sense that consistently there is an uncountable chromatic, romboid-free triple system omitting C_3 and C_5 ([47]). In [49] I describe a forcing extension in which it can be determined, for what finite triple systems \mathcal{F} does the partition relation $\omega_1 \rightarrow (\mathcal{F}, \omega_1)^3$ hold, this solves another problem of [15].

Not surprisingly, one can also define the notion of coloring number for (not necessarily uniform) systems of finite sets. Many of the arguments used for the coloring number of graphs can be adapted and even a very weak and cumbersome result can be proved on obligatory subsystems. This, however, is sufficient to solve another problem of Erdős: if $2 \leq n < \omega$, V is a vector space over the rationals with $|V| \geq \aleph_n$, then there is a subset $W \subseteq V$, $|W| = \aleph_n$ which is not the union of countably many linearly independent sets, but every subset $W' \subseteq W$ with $|W'| < |W|$ is ([50]).

I now hope that the reader has been fully convinced not only that the theory of infinite graphs is interesting, but that Erdős's deep results and challenging conjectures were most influential in its formation.

REFERENCES

- [1] R. Aharoni: König's duality theorem for infinite bipartite graphs, *Journal of London Mathematical Society*, **29** (1984), 1–12.
- [2] R. Aharoni: Menger's theorem for countable graphs, *Journal of Combinatorial Theory (B)*, **43** (1987), 303–313.
- [3] R. Aharoni, E. Berger: Menger's theorem for infinite graphs, *Inventiones Math.*, **176** (2009), 1–62.
- [4] J. E. Baumgartner: Generic graph construction, *Journal of Symbolic Logic*, **49** (1984), 234–240.
- [5] J. E. Baumgartner, A. Hajnal: A remark on partition relations for infinite ordinals with an application to finite combinatorics, in: *Logic and combinatorics*, Contemporary Mathematics, **65**, Amer. Math. Soc., 1987, 157–167.
- [6] J. Czipser, P. Erdős, A. Hajnal: Some extremal problems on infinite graphs, *Publ. Math. Inst. Hung. Acad.Sci.*, **7** (1962), 441–456.
- [7] N.G. de Bruijn, P. Erdős: A colour problem for infinite graphs and a problem in the theory of relations, *Proc. Konink. Nederl. Akad. Wetensch. Amsterdam*, **54** (1951), 371–373.
- [8] W. Deuber: A generalization of Ramsey's theorem, *Infinite and finite sets*, (Colloq. Keszthely 1973; dedicated to P. Erdős on his 60th birthday), Vol. I. Colloq. Math. Soc. J. Bolyai, Vol. **10**, North Holland, Amsterdam, 1975,
- [9] W. Deuber: Partitionstheoreme für Graphen, *Math. Helv.*, **50** (1975), 311–320.
- [10] A. Dudek, V. Rödl: On the Folkman number $f(2, 3, 4)$, *Exp. Math.*, **17** (2008), 63–67.
- [11] A. Dudek, V. Rödl: On the Turán properties of infinite graphs, *Elect. Journal of Combinatorics*, **15** (2008), R47, pp 14.
- [12] P. Erdős: Some set-theoretical properties of graphs, *Revista de la Univ. Nac. de Tucumán, Ser. A. Mat. y Fis. Teór.* **3** (1942), 363–367.
- [13] P. Erdős: Graph theory and probability, *Canad. J. Math.* **11** (1959), 34–38.
- [14] P. Erdős: Problem 8, in: *Theory of graphs and its applications*, Proceedings of the Symposium held in Smolenice, June 1963, Czechoslovak Acad. Sci. Prague, 1964, p. 159.
- [15] P. Erdős, F. Galvin, A. Hajnal: On set-systems having large chromatic number and not containing prescribed subsystems, *Infinite and finite sets*, (Colloq. Keszthely 1973; dedicated to P. Erdős on his 60th birthday), Vol. I. Colloq. Math. Soc. J. Bolyai, Vol. **10**, North Holland, Amsterdam, 1975, 425–513.
- [16] Erdős Pál, Grünwald Tibor, Weiszfeld Endre: Végtelen gráfok Euler vonalairól, *Mat. Fiz. Lapok*, **43** (1936), 129–141.

- [17] P. Erdős, T. Grünwald, E. Vázsonyi: Über Euler-Linien unendlicher Graphen, *J. Math. Physics*, **17** (1938), 59–75.
- [18] P. Erdős, A. Hajnal: On chromatic number of graphs and set-systems, *Acta. Math. Hungar.*, **17** (1966), 61–99.
- [19] P. Erdős, A. Hajnal: On decomposition of graphs, *Acta. Math. Hungar.*, **18** (1967), 359–377.
- [20] P. Erdős, A. Hajnal: On chromatic number of infinite graphs, in: *Theory of graphs*, Proc. of the Coll. held at Tihany 1966, Hungary, (ed. P. Erdős and G. Katona), Akadémiai Kiadó, Budapest -Academic Press, New York, 1968, 83–89.
- [21] P. Erdős, A. Hajnal: Unsolved problems in set theory, in: *Axiomatic Set Theory* (Proc. Symp. Pure Math. **XIII**, Part I, Univ. Calif. Los Angeles, Calif. 1967), Amer. Math. Soc., Providence, R.I., 1971, 17–48.
- [22] P. Erdős, A. Hajnal, L. Pósa: Strong embeddings of graphs into colored graphs, in: *Infinite and finite sets*, (Colloq. Keszthely 1973; dedicated to P. Erdős on his 60th birthday), Vol. I. Colloq. Math. Soc. J. Bolyai, Vol. **10**, North Holland, Amsterdam, 1975, 585–595
- [23] P. Erdős, A. Hajnal, B. L. Rothschild: On chromatic number of graphs and set-systems, in: *Cambridge School in Mathematical Logic (Cambridge, England, 1971)*, *Lecture Notes in Mathematics*, Vol. **337**, Springer, Berlin, 1973, 531–538.
- [24] P. Erdős, A. Hajnal, S. Shelah: On some general properties of chromatic numbers, in: *Topics in topology (Proc. Colloq. Keszthely, 1972)*, *Colloq. Math. Soc. J. Bolyai*, Vol. **8**. North Holland, Amsterdam, 1974, 243–255.
- [25] P. Erdős, A. Hajnal, E. Szemerédi: On almost bipartite large chromatic graphs, *Annals of Discrete Math.*, **12** (1982), 117–123.
- [26] P. Erdős, S. Kakutani: On non-denumerable graphs, *Bull. Amer. Math. Soc.* **49** (1943), 457–461.
- [27] P. Erdős, R. Rado: A construction of graphs without triangles having pre-assigned order and chromatic number, *J. London Math. Soc.*, **35** (1960), 445–448.
- [28] J. Folkman: Graphs with monochromatic complete subgraphs in every edge coloring, *SIAM Journ. of Applied Math.*, **18** (1970), 19–24.
- [29] M. Foreman: An \aleph_1 -dense ideal on \aleph_2 , *Israel Journ. Math.*, **108** (1998), 253–290.
- [30] M. Foreman, R. Laver: Some downward transfer properties for \aleph_2 , *Advances in Mathematics*, **67** (1988), 230–238.
- [31] F. Galvin: Chromatic numbers of subgraphs, *Periodica Math. Hung.*, **4** (1973), 117–119.
- [32] A. Hajnal: The chromatic number of the product of two \aleph_1 -chromatic graphs can be countable, *Combinatorica*, **5** (1985), 137–140.
- [33] A. Hajnal: Embedding finite graphs into graphs colored with infinitely many colors, *Israel Journal of Math.*, **73** (1991), 309–319.
- [34] A. Hajnal: Paul Erdős' set theory, in: *The Mathematics of Paul Erdős*, (R. Graham, J. Nešetřil, eds.), Springer, 1997, 352–393.
- [35] A. Hajnal, P. Komjáth: What must and what need not be contained in a graph of uncountable chromatic number? *Combinatorica*, **4** (1984), 47–52

- [36] A. Hajnal, P. Komjáth: Embedding graphs into colored graphs, *Trans. Amer. Math. Soc.*, **307** (1988), 395–409.
- [37] A. Hajnal, P. Komjáth: Obligatory subsystems of triple systems, *Acta Math. Hung.*, **119** (2008), 1–13.
- [38] P. Komjáth: Connectivity and chromatic number of infinite graphs, *Israel Journal of Mathematics*, **56** (1986), 257–266.
- [39] P. Komjáth: The colouring number, *Proc. London Math. Soc.*, **54** (1987), 1–14.
- [40] P. Komjáth: Consistency results on infinite graphs, *Israel Journal of Mathematics*, **61** (1988), 285–294.
- [41] P. Komjáth: Third note on Hajnal-Máté graphs, *Periodica Math. Hung.*, **24** (1989), 403–406.
- [42] P. Komjáth: The chromatic number of some uncountable graphs, *Coll. Math. Soc. János Bolyai*, **60**, *Sets, graphs, and numbers*, Budapest (Hungary), 1991, 439–444.
- [43] P. Komjáth: Ramsey-theory and forcing extensions, *Proc. Amer. Math. Soc.* **121**, (1994), 217–219.
- [44] P. Komjáth: Two remarks on the coloring number, *Journal of Combinatorial Theory, (B)*, **70** (1997), 301–305.
- [45] P. Komjáth: Some remarks on obligatory subsystems of uncountably chromatic triple systems, *Combinatorica*, **21** (2001), 233–238.
- [46] P. Komjáth: Subgraph chromatic number, DIMACS Series in Discrete Mathematics and Computer Science, **58**, 2002, 99–106.
- [47] P. Komjáth: An uncountably chromatic triple system, *Acta Math. Hung.*, **121** (2008), 79–92.
- [48] P. Komjáth: A note on chromatic number and connectivity of infinite graphs, *Israel Journal of mathematics*, to appear.
- [49] P. Komjáth: On a problem of Erdős-Galvin-Hajnal, manuscript.
- [50] P. Komjáth: The coloring number of a system of finite sets, and a problem of Erdős, manuscript.
- [51] P. Komjáth, S. Shelah: Forcing constructions for uncountably chromatic graphs, *Journal of Symbolic Logic*, **53** (1988), 696–707.
- [52] P. Komjáth, S. Shelah: A consistent partition theorem for infinite graphs, *Acta Math. Hung.*, **61** (1993), 115–120.
- [53] P. Komjáth, S. Shelah: Finite subgraphs of uncountably chromatic graphs, *Journal of Graph Theory*, **49** (2005), 28–38.
- [54] D. König: *Theorie der endlichen und unendlichen Graphen*, Akademische Verlagsgesellschaft, MBG, Leipzig, 1936.
- [55] A. R. Lange, S. P. Radziszowski, X. Xu: Use of MAX-CUT for Ramsey arrowing triangles,
http://www.cs.rit.edu/~arl9577/is/folkman/paper/fe334_mc.pdf
- [56] L. Lovász: *Combinatorial Problems and Exercises*, North-Holland, 1973.
- [57] J. Nešetřil, V. Rödl: Type theory of partition properties of graphs, in: *Recent Advances in Graph Theory*, (ed. M. Fiedler), Academia, Prague, 1975, 183–192.

- [58] J. Nešetřil, V. Rödl: Ramsey properties of graphs with forbidden complete subgraphs, *Journ. Comb. Th.*, (B), **20** (1976), 243–249.
- [59] J. Nešetřil, V. Rödl: A short proof of the existence of restricted Ramsey graphs by means of a partite construction, *Combinatorica*, **1** (1981), 199–202.
- [60] V. Rödl: M.Sc. Thesis, Charles University, Prague, 1973.
- [61] V. Rödl: On the chromatic number of subgraphs of a given graph, *Proc. Amer. Math. Soc.*, **64** (1977), 370–371.
- [62] S. Shelah: Infinite abelian groups, Whitehead problem, and some constructions, *Israel Journal of Mathematics*, **18** (1974), 243–256.
- [63] S. Shelah: Consistency of positive partition theorems for graphs and models, *Set theory and applications* (J. Steprāns, S. Watson, eds.), Lecture Notes in Math., **1401**, 167–193.
- [64] S. Shelah: Incompactness for chromatic numbers of graphs, *A tribute to P. Erdős* (A. Baker, B. Bollobás, A. Hajnal, eds) Camb. Univ. Press, (1990), 361–371.
- [65] L. Soukup: On chromatic number of product of graphs, *Comm. Math. Univ. Carol.*, **29** (1988), 1–12.
- [66] J. Spencer: Three hundred million points suffice, *Journ. Comb. Th.*, (A), **49** (1988), 210–217.
- [67] C. Thomassen: Cycles in graphs of uncountable chromatic number, *Combinatorica*, **3** (1983), 133–134.
- [68] S. Todorcevic: Coloring pairs of countable ordinals, *Acta Math.*, **159** (1987), 261–294.
- [69] S. Todorcevic: Comparing the continuum with the first two uncountable cardinals. in: *Logic and Scientific Methods*, (eds. M. L. Dalla Chiara et al.), Kluwer, 1997, 145–155.
- [70] S. Todorcevic: Combinatorial dichotomies in set theory, *Bull. of the Symbolic Logic*, **17** (2011), 1–72.

Péter Komjáth

*Institute of Mathematics,
Eötvös University,
Budapest, P.O.Box 120,
1518, Hungary*

e-mail: kope@cs.elte.hu

THE IMPACT OF PAUL ERDŐS ON SET THEORY

KENNETH KUNEN

1. INTRODUCTION

This is a brief survey of some areas in set theory where the impact of Paul Erdős is strongly felt today. We omit topics in partition theory and graph theory, which are covered in the article by Péter Komjáth in this volume.

Two themes will emerge in this survey. First, besides proving many first-rate results himself, Erdős always seemed to know the right questions to ask, and he frequently inspired important work by other people. We shall point out some questions that he asked in writing; we cannot mention the many questions that he asked informally in person and in his lectures as he traveled around the world.

Second, modern work in set theory makes frequent use of concepts from logic; this is clear in forcing; but also, large cardinals are studied using elementary embeddings, and elementary submodels are used to prove set-theoretic results. Erdős himself did not employ methods from logic, but his work naturally suggested the use of logic once these methods were developed.

Many readers of this survey will have learned set theory not by reading the papers of Erdős, but by reading a modern text, such as [26, 29, 30, 38]. The notation has changed quite a bit over the years, and readers may be surprised to learn how much of what they know goes back to Erdős.

2. LARGE CARDINALS

The work of Erdős here centers around “medium-size” large cardinals, such as weakly compact, Ramsey, and measurable cardinals. Inaccessible and Mahlo cardinals were already known before Erdős started working in mathematics.

Some of the basic facts about large cardinals are contained in the seminal papers [24] (Erdős and Tarski, 1961) and [8, 11] (Erdős and Hajnal, 1958 and 1962). In particular, [24] considers the properties P_1, P_2, P_3, P_4, Q, R that a cardinal λ may have. We list these below, always assuming that $\lambda > \omega$. We have translated the definitions from [24] into modern terminology, but we have followed [24] in defining the *negation* of the property to be the large cardinal property; for example, $\neg P_3$ is now the standard definition of “measurable cardinal”, and $\neg P_2$ is now one of the standard equivalents of “weakly compact cardinal”.

P_1 : There is a total order \triangleleft on the set λ such that \triangleleft has no increasing or decreasing λ -sequences.

P_2 : $\lambda \rightarrow (\lambda)_2^2$.

P_3 : There is no λ -complete non-principal ultrafilter on λ .

P_4 : There is a λ -complete and λ -distributive boolean algebra \mathcal{B} that is not isomorphic to any λ -complete set algebra.

Q : There exists a λ -Aronszajn tree.

R : There is a λ -complete and λ -distributive boolean algebra \mathcal{B} that is not isomorphic to any λ -complete set algebra *and* \mathcal{B} is λ -generated by a set of size λ .

Paper [24] shows that each P_m implies P_{m+1} ($m = 1, 2, 3$). For $m = 1$, their argument is really a variant of Sierpiński’s example showing that $\mathfrak{c} \rightarrow (\omega_1)_2^2$. They also show that $\neg P_1$ (and hence each $\neg P_m$) implies that λ is strongly inaccessible. It was already clear from Ulam [53] (1930) that the stronger property $\neg P_3$ (measurability) implies strong inaccessibility. The paper [8] (1958), which was earlier than [24], mentions as plausible hypotheses:

- (*) The Generalized Continuum Hypothesis.
- (**) Every strongly inaccessible cardinal is measurable.

Of course, (*) had already been shown to be consistent by Gödel. Ulam [53] did not refute (**), although it was refuted soon after [8] appeared by Tarski [52] in 1960 (assuming that strong inaccessibles exist), with some further proofs in [11] (Erdős and Hajnal, 1962); in modern language, we would say that the first strong inaccessible is not even weakly compact.

With the benefit of hindsight, it is now well-known that for strongly inaccessible λ , each of $\neg P_1, \neg P_2, \neg Q, \neg R$ is equivalent to weak compactness, whereas $\neg P_3$ is equivalent to measurability and $\neg P_4$ is equivalent to strong compactness. Some of the implications are done in [24], and some are left as open questions that were later solved by others. Their paper proves $P_1 \rightarrow P_2$ but not $P_2 \rightarrow P_1$.

Also, they state that they do not know whether $\neg Q$ or $\neg R$ imply strong inaccessibility. For $\neg Q$, this is now known to be false. In fact, by a well-known result of Mitchell in 1970 (see [42]), even \aleph_2 can satisfy $\neg Q$.

The papers [24, 8, 11] mentioned above were all within a period of about ten years which included a flurry of activity on large cardinals by quite a number of people. Erdős was at the center of this activity, both in his writing and in person.

This activity led naturally to some characterizations of weak compactness by other people using logic, in terms of elementary end extensions of $R(\lambda)$, and in terms of Π_1^1 indescribability. In particular, there is the well-known 1964 paper of Keisler and Tarki [33], which mentions [8] (Erdős and Hajnal) and [24] (Erdős and Tarski). Also, the 1962 paper of Erdős and Hajnal [11] mentions the 1960 paper of Keisler [31] on the applications of model theory to set theory.

Also, Scott [45] in 1961 used ultraproducts by countably complete ultrafilters to show that a measurable cardinal implies $V \neq L$. Scott mentions that it was already known that “most” strong inaccessibles are not measurable, referring to, among others, [11] (Erdős and Hajnal, 1962).

Regarding $\neg P_3 \rightarrow \neg P_2$ (measurable cardinals satisfy $\lambda \not\rightarrow (\lambda)_2^2$): Although this is done in [24], where the properties are defined, the earlier [8] (Theorem 9a) contains what is essentially the stronger result that measurable cardinals λ are Ramsey (satisfy $\lambda \rightarrow (\lambda)_2^{<\omega}$). Curiously, the proof in [8] shows that assuming (**), every strong inaccessible is Ramsey, but, as they point out in [11], by which time they knew that (**) was false, the proof only required that the particular cardinal in question be measurable.

This work led to other types of “Erdős cardinal”, such as $\kappa \rightarrow (\omega_1)_2^{<\omega}$, and then these partition relations led naturally to the theory of $0^\#$ due to Silver and Solovay, applying the Ehrenfeucht-Mostowski method of indiscernibles from logic. Silver’s thesis appeared 1966 and Solovay’s paper [49] was published in 1967.

An important contribution to set theory occurs in the paper [12] (Erdős and Hajnal, 1966). The paper is mainly about Jónsson cardinals. These are infinite cardinals λ such that every first-order structure of size λ for a countable language has a proper elementary substructure of the same size; for example, every Ramsey cardinal is a Jónsson cardinal, and the existence of such a cardinal implies that $0^\#$ exists. But [12] points out that if one allows the structure to contain ω -ary operations, then the property becomes inconsistent. That is, there is always a map $f : \lambda^\omega \rightarrow \lambda$ such that for all $A \in [\lambda]^\lambda$, $f(A^\omega)$ is all of λ .

As Kunen [35] (1971) pointed out, it follows very easily from this that there cannot be a nontrivial embedding from V into V ; so this puts a

limit on how “huge” a cardinal can be. More precisely, say $j : V \rightarrow M$ is an elementary embedding first moving cardinal κ , where M is a transitive class. Let $\lambda = \sup\{j^n(\kappa) : n \in \omega\}$. Then, using the Erdős-Hajnal result, M cannot contain all subsets of λ (e.g., $j^{\omega} \lambda \notin M$). There has been much work on related weaker assumptions (e.g., $\mathcal{P}(j^{\omega}(\kappa)) \subset M$); these are not known to be inconsistent.

This is another example illustrating the fact that although Erdős did not work in logic, the addition of a little logic to his results by other people led to some important consequences.

Paper [13] (Erdős and Hajnal, 1974) gives an interesting equivalent of weak compactness with a variant of the free set lemma (see Section 8). By 1974, the term “weakly compact” had become standard terminology, and the various equivalents to weak compactness described above were well-known. For infinite cardinals $\lambda \leq \kappa$, say that $P(\kappa, \lambda)$ holds iff whenever $\mathcal{F} \subseteq [\kappa]^{<\kappa}$ and $|\mathcal{F}| = \kappa$ and $x \not\subseteq y$ for all distinct $x, y \in \mathcal{F}$, there is an $\mathcal{F}' \subseteq \mathcal{F}$ with $|\mathcal{F}'| = \kappa$ such that $|\kappa \setminus \bigcup \mathcal{F}'| \geq \lambda$. They show that $P(\kappa, \kappa)$ holds iff $\kappa = \omega$ or κ is weakly compact. They also discussed $P(\kappa, \lambda)$ for $\lambda < \kappa$, but this is not related to large cardinals.

For the \rightarrow direction of the “iff”: They use what was by then a standard argument to get an $A \subseteq \kappa$ that is a “limit point” of \mathcal{F} in the sense that there are $A_\xi \in \mathcal{F}$ for $\xi < \kappa$ such that $A \cap \xi = A_\eta$ for all $\eta \geq \xi$. Then $|\kappa \setminus A| = \kappa$ (otherwise there is an easy contradiction), and then it is easy to get the desired \mathcal{F}' as a subset of $\{A_\xi : \xi < \kappa\}$.

The proof of the \leftarrow direction of the “iff” splits into three clever arguments: one if κ is singular, another if $\exists \theta < \kappa [2^\theta \geq \kappa]$, and a third if κ is strongly inaccessible and there is a κ -Aronszajn tree. In the case of the Aronszajn tree T , they got $\mathcal{F} \subseteq [T]^{<\kappa}$. The elements of \mathcal{F} were sets of the form $T_\gamma \setminus C_\gamma$, where $T_\gamma = \{x \in T : \text{ht}(x) < \gamma\}$ and C_γ is some maximal chain in T of order type γ (so C_γ is a path through T_γ).

3. CHAIN CONDITIONS IN FORCING

Erdős didn’t do forcing, but a number of his results on chain conditions are often quoted in the forcing literature. If \mathbb{P} is a forcing poset, then the *Suslin number*, $S(\mathbb{P})$, is the least κ such that there is no antichain in \mathbb{P} of size κ ; so, \mathbb{P} is ccc iff $S(\mathbb{P}) \leq \aleph_1$. An important fact, due to Erdős and Tarski [23] (Theorem 1) in 1943, is that $S(\mathbb{P})$ must be regular, and cannot be \aleph_0 . They also showed that given any regular uncountable κ , one can find \mathbb{P} with $S(\mathbb{P}) = \kappa$. This is trivial if κ is a successor. If κ is a regular limit (i.e., weakly inaccessible), the example they gave is essentially the same as what

is now called the Lévy collapse of κ (see [29, 38]). We remark that although they used the letter “ $\mathfrak{d}(\mathbb{P})$ ” where we now use “ $S(\mathbb{P})$ ”, their definition of the notion in terms of partially ordered sets is very much like the one given in modern forcing textbooks (e.g., [38]).

Results on chain conditions are also related to topology. If X is a topological space, then one frequently studies $S(X)$, which is defined to be $S(\mathbb{P})$, where \mathbb{P} is the family of non-empty open subsets of X , ordered by \subseteq .

The Erdős – Tarski theorem was proved long before forcing was conceived of, although the ccc was already a well-known property in topology, following Suslin [50] in 1920. But, Erdős continued to do important work on chain conditions in the era of forcing. In particular, in the 1970s he gave a very clever argument that under CH, ω_1 is not a pre-caliber for random real forcing; by then, it was well-known that $\text{MA}(\aleph_1)$ implies that ω_1 is a pre-caliber for *all* ccc posets; his proof is described in the 1979 paper [39] of Kunen and Tall.

4. SET-THEORETIC TOPOLOGY

The above results on chain conditions are related to topology, but in this section we mention two results on products of spaces, $\prod_{i \in I} X_i$. The first one uses the *box topology*; this has as a base all $\prod_{i \in I} U_i$, where U_i is open in X_i . The second uses the standard Tychonov topology, where the basic sets are only those $\prod_{i \in I} U_i$ for which $U_i = X_i$ for all but finitely many i .

There are still many open questions about when a countable box product $\prod_{n \in \omega} X_n$ is normal. This is not always clear even when all the X_n are ordinals; we write this product as $\prod_n \alpha_n$. If all the α_n are successor ordinals or limits of countable cofinality, then under CH, the product is normal (and in fact paracompact) by M. E. Rudin [43], but this is still open in ZFC, even when all $\alpha_n = \omega + 1$.

A well-known paper of Erdős and Rudin [22] (in the volume honoring the 60th birthday of Erdős in 1973) is a contribution to that problem by showing that $\prod_n \alpha_n$ is not normal if $\alpha_n = \omega + 1$ for $n > 0$ and $\alpha_0 = \kappa$, where κ is regular and is the order type of a scale in ω^ω , provided that $\kappa > \omega_1$. The proviso “ $\kappa > \omega_1$ ” seems a little strange and the case $\kappa = \omega_1$ (which happens under CH) was handled by Kunen [36] (published in the same volume), who refined their method slightly. So, this is a good example of Erdős causing the right question to be asked, even if someone else settled it.

Returning to products $\prod_{i \in I} X_i$, using the Tychonov topology: The Tychonov Theorem (1935) says that any product $\prod_i X_i$ of compact spaces

is compact. Is there a similar theorem when the X_i are only Lindelöf? An important paper of Erdős and Hajnal [9] (1961), when combined with a 1959 paper of Łoś [40], shows that this is a deep question, and that the “obvious” conjectures are false.

Call a space X κ -compact iff every open cover has a subcover of size less than κ ; so \aleph_0 -compact is compact and \aleph_1 -compact is Lindelöf. By [40], if κ is below the first measurable cardinal, then there is a product of Lindelöf spaces that fails to be κ -compact; in fact the spaces are all just discrete and countable. But, by [9], if κ is measurable then the space ω^κ is κ -compact; actually, they say that the (inconsistent) hypothesis $(**)$ (see Section 2) implies that ω^κ is κ -compact for all strongly inaccessible κ , but their proof is fine when κ really *is* measurable. In modern language, the proof is similar to the ultrafilter proof of the compactness of 2^κ in ZFC, but now the ultrafilter is countably complete. Although they don’t state this explicitly, their methods show that if θ is strongly compact, then every product of θ -compact spaces is θ -compact; in fact, this (plus $\theta > \omega$) is sometimes taken as a definition of “strongly compact”; see Tall’s survey [51].

5. QUESTIONS TO BE ANSWERED BY FORCING

Although some of the later work of Erdős made explicit mention of MA and other forcing results, some of his earlier work, which predates MA, suggested natural questions to be attacked by forcing once the method was developed.

The paper of Erdős and Hajnal [9] (1961), cited above for κ -compactness, also has some important results regarding the Bernstein set construction in various models of set theory.

Following E. W. Miller [41] (1937), we say that a family \mathcal{F} of sets has *property* \mathfrak{B} iff there is a set B such that $F \cap B \neq \emptyset$ and $F \not\subseteq B$ for all $F \in \mathcal{F}$. So, if \mathcal{F} is the family of perfect subsets of \mathbb{R} , then \mathcal{F} has property \mathfrak{B} , and the set B is the famous set constructed by Bernstein in 1908.

As is typical in writings of Erdős, properties of cardinals are expressed with arrows (as in the familiar $\omega \rightarrow (\omega)_2^2$). Here, we use a modification of the arrow notation of [9], following [27] (Hajnal, Juhász, and Shelah, 2000). Call a family \mathcal{F} of sets μ -almost disjoint or μ -a.d. iff $|X \cap Y| < \mu$ whenever X, Y are two distinct members of \mathcal{F} . Then if $\mathcal{F} \subseteq [\lambda]^\kappa$, \mathcal{F} is trivially κ^+ -a.d., and \mathcal{F} is κ -a.d. iff \mathcal{F} is an almost disjoint family in the usual sense.

For infinite cardinals κ, λ, μ with $\kappa \leq \lambda$: $M(\lambda, \kappa, \mu) \rightarrow \mathfrak{B}$ denotes the assertion that whenever $\mathcal{F} \subseteq [\lambda]^\kappa$ with $|\mathcal{F}| \leq \lambda$ and \mathcal{F} is μ -a.d., then \mathcal{F} has property \mathfrak{B} .

Theorem 2 of [9] is $M(\kappa, \kappa, \kappa^+) \rightarrow \mathfrak{B}$; here, \mathcal{F} is a family of κ sets, each of size κ with no almost disjointness assumption. They rightly attribute this to Bernstein, since the Bernstein set argument, with $\kappa = \mathfrak{c}$, clearly works for all κ .

Note that $M(2^\kappa, \kappa, \kappa^+) \dashv\rightarrow \mathfrak{B}$ is trivial, taking \mathcal{F} to be all of $[2^\kappa]^\kappa$. Their Theorem 3 is $M(2^\kappa, \kappa, \kappa) \dashv\rightarrow \mathfrak{B}$ (so the counter-example \mathcal{F} is an almost disjoint family); for $\kappa = \aleph_0$, this is due to Miller [41] (or see [26], pages 163 (Exercise 13) and 275), and they say that the proof is essentially the same for any κ . Note that, when discussing almost disjoint sets of size κ , we are *not* assuming that $|\bigcup \mathcal{F}| = \kappa$, as is common in discussions of almost disjoint sets, MAD families, etc. In fact, by a result of Baumgartner, when $\kappa = \omega_1$, one can't in ZFC produce any almost disjoint family \mathcal{F} of size 2^κ with $|\bigcup \mathcal{F}| = \kappa$ (see [38], Exercise IV.7.51). Of course, for the easier example showing $M(2^\kappa, \kappa, \kappa^+) \dashv\rightarrow \mathfrak{B}$, we could get $|\bigcup \mathcal{F}| = \kappa$, letting $\mathcal{F} = [\kappa]^\kappa$.

Under CH, Theorem 3 (or Miller's result) yields $M(\aleph_1, \aleph_0, \aleph_0) \dashv\rightarrow \mathfrak{B}$. They ask (Problem 1) what happens if \neg CH. Note that their paper, in 1961, was just before the advent of forcing made it easy to answer this. Even the stronger $M(\aleph_1, \aleph_0, \aleph_1) \rightarrow \mathfrak{B}$ follows from MA(\aleph_1) (or even $\mathfrak{p} > \aleph_1$). Decoding this: $|\mathcal{F}| = \aleph_1$ and the sets in \mathcal{F} are countably infinite and not necessarily almost disjoint. Then property \mathfrak{B} asks for a set B such that $F \cap B \neq \emptyset$ and $F \not\subseteq B$ for all $F \in \mathcal{F}$. WLOG, $\bigcup \mathcal{F} \subseteq \omega_1$, and then the poset $\text{Fn}(\omega_1, 2)$ supplies the required B .

After this problem, they ask (essentially) whether perhaps \neg CH alone is sufficient here. The answer is “no” by another well-known (now) forcing result. In fact, it is consistent with \neg CH that even the weaker $M(\aleph_1, \aleph_0, \aleph_0) \rightarrow \mathfrak{B}$ fails and here we even get $|\bigcup \mathcal{F}| = \aleph_0$. To see this, assume that there is a non-principal ultrafilter \mathcal{U} on ω that is generated by \aleph_1 sets $\mathcal{E} = \{E_\alpha : \alpha < \omega_1\} \subseteq \mathcal{U}$; that is $X \in \mathcal{U} \leftrightarrow \exists \alpha [E_\alpha \subseteq X]$; this is consistent with \mathfrak{c} being arbitrarily large (see [38], Lemma V.4.27). Now, if $E_\alpha \cap B \neq \emptyset$ for all α then $\omega \setminus B \notin \mathcal{U}$ so $B \in \mathcal{U}$ so $E_\alpha \subseteq B$ for some B . This refutes $M(\aleph_1, \aleph_0, \aleph_1) \rightarrow \mathfrak{B}$. Of course, the E_α are not almost disjoint. To refute $M(\aleph_1, \aleph_0, \aleph_0) \rightarrow \mathfrak{B}$, use $\mathcal{F} = \{F_\alpha : \alpha < \omega_1\}$, where each F_α is an infinite subset of E_α chosen recursively so that $F_\alpha \notin \mathcal{U}$ and $F_\alpha \cap F_\xi$ is finite for all $\xi < \alpha$.

These independence results are fairly standard applications of the forcing method (now that the method has been developed), but Erdős and co-workers get some credit for pointing out the interesting questions to ask. A much deeper result arose from another question in [9], which demonstrates that there are non-trivial questions in this area even under GCH. The paper [27] (Hajnal, Juhász, Shelah, 2000) mentioned above shows that assuming the consistency of a supercompact cardinal, GCH plus

$M(\aleph_{\omega+1}, \aleph_1, \aleph_0) \dashv\vdash \mathfrak{B}$ is consistent. So, here we are talking about families of sets of size \aleph_1 with pairwise *finite* intersection. Note that they cannot be pairwise disjoint; otherwise a simple transversal would establish property \mathfrak{B} . [9] showed that CH implies that $M(\lambda, \aleph_1, \aleph_0) \rightarrow \mathfrak{B}$ holds for $\lambda \leq \aleph_\omega$. Here, one can prove $M(\aleph_{n+1}, \aleph_1, \aleph_0) \rightarrow \mathfrak{B}$ by induction on n (using $(\aleph_n)^{\aleph_0} = \aleph_n$), and then it is an easy step to $M(\aleph_\omega, \aleph_1, \aleph_0) \rightarrow \mathfrak{B}$. Also, work by others before 2000 (see [27] for references) showed that $M(\lambda, \aleph_1, \aleph_0) \rightarrow \mathfrak{B}$ holds for all λ assuming GCH plus suitable \square principles (such as hold in L); this shows that some large cardinal is required for the consistency of $\text{GCH} + M(\aleph_{\omega+1}, \aleph_1, \aleph_0) \dashv\vdash \mathfrak{B}$.

The older paper [3] (1943) of Erdős also has some material that begs a use of MA. One result in it improves on a result of Sierpiński. In \mathbb{R} , let null denote the null ideal and meager the meager ideal. Assuming CH, Sierpiński had shown that there is a bijection f of \mathbb{R} onto \mathbb{R} such that $X \in \text{null} \leftrightarrow f(X) \in \text{meager}$ for all $X \subseteq \mathbb{R}$. In answer to a question of Sierpiński, Erdős shows that one can get f with the additional property that $X \in \text{meager} \leftrightarrow f(X) \in \text{null}$. He also gets $f = f^{-1}$. It is easy to see that the same proof works under MA, or just under $\text{add}(\text{null}) = \text{add}(\text{meager}) = \mathfrak{c}$.

It is still not clear exactly what is needed to get this result. It cannot be proved in ZFC because even the result of Sierpiński implies that $\text{add}(\text{null}) = \text{add}(\text{meager})$ and $\text{cov}(\text{null}) = \text{cov}(\text{meager})$ and $\text{non}(\text{null}) = \text{non}(\text{meager})$.

The short three page paper [20] of Erdős and Makkai (1966) is notable also for raising some interesting questions:

Some notation: If $\mathcal{G} \subseteq \mathcal{P}(A)$ and $f \in A^\omega$, then \mathcal{G} *strongly cuts* f iff for all $n \in \omega$ there is an X_n in \mathcal{G} such that for all $i \in \omega$, $f(i) \in X_n$ iff $i < n$. So, X_0 contains none of the $f(i)$, X_1 contains only $f(0)$, X_2 contains only $f(0), f(1)$, etc. If $S \subseteq A$, let $S \setminus \mathcal{G} = \{S \setminus X : X \in \mathcal{G}\}$.

They prove that if $\kappa \geq \aleph_0$ and $\mathcal{G} \subseteq \mathcal{P}(\kappa)$ with $|\mathcal{G}| > \kappa$, then there is an $f \in \kappa^\omega$ that either \mathcal{G} or $\kappa \setminus \mathcal{G}$ strongly cuts f .

A simple example occurs when $\kappa = \omega$ and \mathcal{G} is an uncountable almost disjoint family. Then $\omega \setminus \mathcal{G}$ cannot strongly cut any f . To get an f strongly cut by \mathcal{G} : Let $\mathcal{G}_0 = \mathcal{G}$ and choose $X_0 \in \mathcal{G}_0$ arbitrarily. Then choose $f(0) \notin X_0$ such that $\mathcal{G}_1 := \{X \in \mathcal{G}_0 : f(0) \in X\}$ is uncountable. Then choose $X_1 \in \mathcal{G}_1$ arbitrarily. Then choose $f(1) \notin X_0 \cup X_1$ such that $\mathcal{G}_2 := \{X \in \mathcal{G}_1 : f(1) \in X\}$ is uncountable. Etc.

Their theorem raises a number of natural questions; we mention two of them here:

Their *Problem 1* asks whether one can sometimes dispense with the second alternative and simply prove that there is an $f \in \kappa^\omega$ that \mathcal{G} strongly cuts f ? Following Shelah [48] (1972), let $P3(\theta, \lambda)$ denote the assertion that

there exist A of size λ and $\mathcal{G} \subseteq \mathcal{P}(A)$ of size θ such that no ω -sequence from A is strongly cut by \mathcal{G} . Problem 1 asks whether $P3(\lambda^+, \lambda)$ holds for every infinite λ , and Shelah proved that this is indeed true. Of course, for $\lambda = \aleph_0$, the result is in [20], using $\mathcal{G} = \omega \setminus \mathcal{H}$, where \mathcal{H} is an almost disjoint family. Since this is a ZFC theorem, forcing is not involved. Shelah actually shows that $\mu < \lambda < \text{Ded}(\mu)$ implies that $P3(\lambda, \mu)$ is false; here, $\text{Ded}(\mu)$ is the least cardinal $\delta > \mu$ such that no ordered set of size δ has a dense subset of size μ . He uses a tree argument – one can view λ as the set of branches through a tree of size μ .

Their *Problem 3* asks: Suppose that $\mathcal{G} \subseteq \mathcal{P}(\omega)$ is uncountable. Does there exist an $f \in \omega^\omega$ such that either: f is strongly cut by $\omega \setminus \mathcal{G}$, or both f is strongly cut by \mathcal{G} and $\text{ran}(f)$ is a subset of some member of \mathcal{G} ? So, in the example with the almost disjoint family, one would need $\text{ran}(f)$ to be a subset of some member of the family. But, they mention that Máté had already shown that an almost disjoint family cannot be a counter-example. We remark that if there is any counter-example \mathcal{G} , then by a standard absoluteness argument, it remains a counter-example in every forcing extension of the universe that preserves ω_1 .

Besides [48], the papers Shelah [47] (1971) and Keisler [32] (1976) make use of the above mentioned result from [20]. All three papers are primarily in model theory, and study the possible values of the stability function. In particular, [47] uses this result to show that unstability implies the order property. Keisler [32] (1976) carries the analysis further and shows that there are exactly six possibilities for the stability function.

6. ORDER TYPES

Paper [10] (Erdős and Hajnal, 1962) gives a complete analysis of the countable (total) order types. It is roughly analogous to the Cantor-Bendixson analysis in topology.

As usual, η denotes the order type of the rationals. Of course, they knew, by Cantor, that the only countable dense order types are η , $1 + \eta$, $\eta + 1$, and $1 + \eta + 1$. An order type is called *discrete* or *scattered* iff no subset of it is densely ordered.

First, they describe all the countable discrete order types. These are obtained by the following process: Let $\mathcal{O}_0 = \{0, 1\}$; that is, the empty order and the one-element order. Let \mathcal{O}_α be the class of all ω -sums and ω^* -sums of order types taken from $\bigcup\{\mathcal{O}_\delta : \delta < \alpha\}$. This defines \mathcal{O}_α for all ordinals α , but it is easily seen that the process closes off at stage ω_1 , so that $\mathcal{O}_{\omega_1} = \bigcup\{\mathcal{O}_\delta : \delta < \omega_1\}$. Let $\mathcal{O} = \mathcal{O}_{\omega_1}$. It is easily seen by induction

that all types in each \mathcal{O}_α , and hence all types in \mathcal{O} , are discrete. But in fact, they show that \mathcal{O} contains all countable discrete order types. They seemed to be unaware of the fact that this part of their paper was known earlier; it was published by Hausdorff [28] in 1908.

But furthermore, they show that every non-discrete countable order type is a sum of the form $\sum_{d \in D} \Theta_d$, where D is a densely ordered set and each Θ_d is a non-empty discrete order type.

They apply their analysis of order types to partition relations. This is more the subject of the paper by Komjáth in this volume, but briefly: The relation $\Theta \rightarrow (\Theta, \aleph_0)^2$ asserts that for each partition of the pairs from a set of order type Θ into {red, blue}, there is either a subset of order type Θ all of whose pairs are colored red, or a subset of cardinality \aleph_0 all of whose pairs are colored blue. An earlier paper [21] (Erdős and Rado, 1956) showed that $\eta \rightarrow (\eta, \aleph_0)^2$. It follows immediately that $\Theta \rightarrow (\Theta, \aleph_0)^2$ holds for every non-discrete countable type (since Θ both contains a copy of η and embeds into η). Also, $\omega \rightarrow (\omega, \aleph_0)^2$ and $\omega^* \rightarrow (\omega^*, \aleph_0)^2$ are immediate from Ramsey's Theorem. They show in this paper that these are the only examples; that is if Θ is a countably infinite discrete type other than ω and ω^* , then $\Theta \not\rightarrow (\Theta, \aleph_0)^2$. This is non-trivial and makes essential use of the inductive construction of the discrete types.

7. GEOMETRY

Erdős contributed extensively to geometry. We discuss here only some *set-theoretic* questions about Euclidean space. But, we include a few facts that mix the geometry with measure theory and/or linear algebra.

An example of a result on linear structure is his paper with Kakutani [17] (1943), which shows that CH is equivalent to the statement that \mathbb{R} is a countable union of rationally independent sets. This continues in the spirit of Sierpiński's book [46] (1934) on CH, to which they refer. Like many of the results [46], this theorem of [17] can be re-phrased as a ZFC theorem: If V is a vector space over a countable field (such as \mathbb{Q}), then V is a countable union of linearly independent sets iff $|V| \leq \aleph_1$. In any case, the \leftarrow direction is an easy exercise but the \rightarrow direction is non-trivial.

Some more purely geometric results occur in his paper [4] (1950). This shows in ZFC that if S is an infinite subset of \mathbb{R}^k , then there is an $S' \subseteq S$ with $|S'| = |S|$ with *distinct distances*; that is $d(x, y) \neq d(z, t)$ unless $\{x, y\} = \{z, t\}$. He points out that this is a theorem about finite dimensional geometry, not metric spaces, since it is false for infinite dimensional Hilbert spaces.

He then raises the natural question of whether one could partition \mathbb{R}^k as $\bigcup_n S_n$, where each S_n has distinct distances. By [17], this is false for all k under $\neg\text{CH}$, and true for $k = 1$ under CH . For $k = 2$, this was eventually shown to be true under CH by R. O. Davies [2] in 1971, and then later Kunen [37] did it for $k > 2$ in 1987.

This paper [4] also has the following interesting ZFC result on the borderline between geometry and measure theory: If $H \subseteq \mathbb{R}$ is a Hamel basis, let H_k denote the set of all linear combinations from H using k or fewer elements of H , so $\mathbb{R} = \bigcup_k H_k$. It was already well-known (by an easy difference set argument) that each H_k is either Lebesgue null or non-measurable, so that some H_k must be non-measurable. But Erdős showed, by a sophisticated transfinite recursion, that for each k there is a Hamel basis H such that H_k is null but H_{k+1} is not measurable.

A result that combines geometry, measure theory, and linear structure is in the 1981 paper [19] of Erdős, Kunen, and Mauldin: Assuming CH , there is an $X \in [\mathbb{R}]^{\aleph_1}$ such that X is concentrated on \mathbb{Q} and $N + X$ is a null set for all null N . Here, “concentrated” means that $X \setminus U$ is countable for all open $U \supset \mathbb{Q}$; so $\text{MA}(\aleph_1)$ (or just $\mathfrak{b} > \aleph_1$) implies that there are no uncountable concentrated sets.

His paper [7](1978) is an interesting survey and opens up some new questions. He restates his results from [17] and [4]. He notes that his refutation under $\neg\text{CH}$ of the above partition obtained *distinct* x, y, z, t , and he posed the natural question of whether, in ZFC, one could get $\mathbb{R}^k = \bigcup_n S_n$, where each S_n does not contain an isosceles triangle. This was eventually solved affirmatively by Schmerl [44] in 1996.

This is a good example of Erdős inspiring work by others, since he knew exactly the right questions to ask, even when he couldn’t answer them himself.

The paper [18] (Erdős and Komjáth, 1990) contains some more results in the spirit of the early papers [37, 44] of Kunen and Schmerl mentioned above. [18] shows that assuming CH , one can get $\mathbb{R}^2 = \bigcup_n S_n$, where each S_n does not contain a right triangle. They point out that under $\neg\text{CH}$, this was already known to be false, as remarked earlier in Erdős [7] (1978), but without proof.

We next consider the three papers [6, 15, 16] (1955, 1994, 1997). Actually, Erdős died in 1996, so he never lived to see [16] published.

[15] (Erdős, Jackson, and Mauldin, 1994): Sierpiński, in his papers and his book [46], described many equivalents of CH , many of which have a geometric flavor. For example (see [46], p. 214): In \mathbb{R}^n , let \mathcal{L}_i be the set of lines parallel to the i^{th} axis for $i = 1, \dots, n$. Then CH is equivalent to the statement that \mathbb{R}^3 can be partitioned into 3 sets, S_1, S_2, S_3 , such that

for each i and $L \in \mathcal{L}_i$, $L \cap S_i$ is finite. Partitioning \mathbb{R}^4 , replacing 3 by 4 throughout, one gets an equivalent of $2^{\aleph_0} \leq \aleph_2$.

Erdős in [6] (1955) asked whether one could get similar results by varying the definition of the \mathcal{L}_i . In [15] there are some answers. Let $\mathcal{L}^n (= RP^{n-1})$ be the set of all lines in \mathbb{R}^n . Then CH is equivalent to the statement: Whenever \mathcal{L}^3 is partitioned into three subsets, $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$, one can write \mathbb{R}^3 as a disjoint union of sets S_1, S_2, S_3 such that $L \cap S_i$ is finite for $i = 1, 2, 3$ and $L \in \mathcal{L}_i$.

Actually, [15] has a much more general result, a special case of which is the following. Fix s with $1 \leq s < \omega$. Then the following are equivalent:

1. $2^{\aleph_0} \leq \aleph_s$.

2. For all $n, p \geq 2$ and all disjoint $\mathcal{L}_1, \dots, \mathcal{L}_p \subseteq \mathcal{L}^n$, one can write \mathbb{R}^n as a disjoint union of sets S_1, \dots, S_p such that

(*) $\forall i \forall L \in \mathcal{L}_i \ [|S_i \cap L| < \aleph_{\max(0, s+2-p)}]$.

3. For some $n \geq 2$ and p with $2 \leq p \leq s + 2$, and some non-parallel lines ℓ_1, \dots, ℓ_p , (*) holds if we let \mathcal{L}_i be the set of all lines parallel to ℓ_i .

Note that (2) \rightarrow (3) is trivial, and in (2), we would get an equivalent statement if we required the \mathcal{L}_i to form a partition of \mathcal{L}^n . Also, (3), in the special case that the ℓ_i are the coordinate axes and $p = n$, is close in spirit to the work of Sierpiński.

[16] (Erdős, Jackson, Mauldin, 1997) considers countably infinite partitions. Here, the size of \mathfrak{c} is not as important as whether MA holds. For example, assume MA and, for $n \geq 2$, assume that we have partitioned \mathcal{L}^n into $\bigcup_{i \in \omega} \mathcal{L}_i$. Then we can partition \mathbb{R}^n into S_i for $i \in \omega$ such that $|L \cap S_i| \leq 3$ for each $i \in \omega$ and each $L \in \mathcal{L}_i$. The following are two more general results about \mathbb{R}^n (for $n \geq 2$), assuming MA.

Theorem. Let $\mathcal{L}^n = \bigcup_{i \in \omega} \mathcal{L}_i$. Then we can partition \mathbb{R}^n as $\mathbb{R}^n = \bigcup_{i \in \omega} S_i$ such that $|L \cap S_i| \leq 3$ for each $i \in \omega$ and each $L \in \mathcal{L}_i$.

The next theorem is related somewhat to two-point sets (i.e., sets that intersect each line in exactly two points). They were first proved to exist by Mazurkiewicz (1914).

Theorem. Fix $S \subseteq \mathbb{R}^n$ such that $|L \cap S| < \aleph_0$ for all $L \in \mathcal{L}^n$. Then we can partition S as $S = \bigcup_{i \in \omega} S_i$ such that $|L \cap S_i| \leq 3$ for all $L \in \mathcal{L}^n$ and $i < \omega$.

The paper also generalizes these theorems to results that replace lines with higher dimensional hyperplanes.

8. FREE SETS

The Free Set Lemma in its current form was proved by Hajnal [25] in 1960: If $\kappa < \lambda$ are infinite cardinals and $g : \lambda \rightarrow [\lambda]^{<\kappa}$, then there is a free set $F \subseteq \lambda$ of size λ . Here, “free” means that $\alpha \notin g(\beta)$ whenever $\alpha, \beta \in F$ and $\alpha \neq \beta$. But Hajnal’s proof was done after partial results were proved by a number of other people, including Erdős.

Lázár proved it in 1936 for regular λ . This proof is by now an easy exercise in using the Pressing Down Lemma. Erdős [4] proved it in 1950 for singular λ , assuming GCH. Finally, Hajnal [25] got a proof in ZFC. Paper [4] was already mentioned above in the section on geometry, and there is really no relation between the geometry results and the one on free sets; so, the title of [4], “Some remarks on set theory”, is appropriate.

We discuss next his [5] (1954), “Some remarks on set theory III”. In fact, eleven of his papers were similarly titled, including six [3, 4, 5, 6, 13, 20] mentioned in this survey.

Paper [5] addresses the following question. Suppose that $g : \mathbb{R} \rightarrow \mathcal{P}(\mathbb{R})$ and each $g(x)$ is “small” in some sense. What can one say about the existence of “large” free sets? Here, there is actually some relation between the free sets and the geometry.

There are a number of examples where there need not even be a free set of size two: One is when “small” means “ $|g(x)| < \mathfrak{c}$ ”; this is easily done by well-ordering \mathbb{R} in type \mathfrak{c} . Another is when “small” means “ $g(x)$ is not dense in \mathbb{R} ”; for example, let $g(x) = (-\infty, x)$. Somewhat more complicated is the fact that if “small” means “ $|g(x)| < \mathfrak{c}$ and $g(x)$ is not dense in \mathbb{R} ”, then there must be a free set of size two, but there is an example where there is no free set of size three.

He also shows (Theorem 6) that if “small” means “ $g(x)$ is nowhere dense”, then there must be a free set of size \aleph_0 . Erdős says that it is not clear if one can improve this result, even under CH. Actually, Bagemihl [1] (1973) did improve this by showing that one can always get an everywhere dense free set.

Theorem 6 leaves open the question of the existence of a free set of size \aleph_1 or bigger. But, this is independent.

It is false under CH by the following example: List $[\mathbb{R}]^{\aleph_0}$ as $\{E_\xi : \xi < \omega_1\}$ and list \mathbb{R} as $\{x_\alpha : \alpha < \omega_1\}$. Let $g(x_\alpha)$ be an ω -sequence converging to x_α . Make sure that for all $\xi < \alpha$, if $x_\alpha \in \overline{E}_\xi \setminus E_\xi$, then $E_\xi \cap g(x_\alpha) \neq \emptyset$. Now, suppose that $F \in [\mathbb{R}]^{\aleph_1}$ is free. Fix ξ such that $E_\xi \subseteq F \subseteq \overline{E}_\xi$. If $\alpha > \xi$ and $x_\alpha \in F \setminus E_\xi$, then $g(x_\alpha)$ contains a point from $F \setminus \{x_\alpha\}$, contradicting freeness of F .

On the other hand, it is true if $\neg\text{CH}$ and there is a Luzin set L of size \aleph_2 (e.g., as in the Cohen model). Then each $g(x) \cap L$ is countable, so one can apply the standard Free Set Lemma to get a free subset of L of size \aleph_2 .

The paper [14] (Erdős, Hajnal, and Máté, 1973) relates free sets to large cardinals and Suslin trees.

Say we have $g : \lambda \rightarrow \mathcal{P}(\lambda)$. If $\kappa < \lambda$ and all $|g(\alpha)| < \kappa$, then Hajnal's Free Set Lemma implies that there is a free set of size λ . We cannot simply assume that all $|g(\alpha)| < \lambda$; the trivial counter-example being $g(\alpha) = \alpha$, where there is no free set of size 2. But they show that if we now add some structural restrictions on $\text{ran}(g)$ then one can produce large free sets. So, this is similar in spirit to the results in [5] discussed above, except that here the restrictions do not come from geometry. Typical restrictions on a set $S \subseteq [\lambda]^{<\lambda}$ are:

Condition A: For all $F \subseteq \lambda$, the set $\{s \cap F : s \in S\}$ has no increasing λ -chains under inclusion \subsetneq .

Condition B: Whenever $\tau < \lambda$ and $\lambda = \bigcup_{\alpha < \tau} E_\alpha$, with the E_α pairwise disjoint and of size λ , there is an $\alpha < \tau$ and an $F \in [E_\alpha]^\lambda$ such that the set $\{s \cap F : s \in S\}$ has no increasing λ -chains under inclusion \subsetneq .

It is easy to see that Condition A implies Condition B for regular λ . Note that since the trivial example above is a chain, it is natural to avoid such chains for positive results.

A typical theorem in the paper is their Theorem 3.8: Assume that λ is regular and *Condition B* holds with $S = \text{ran}(g)$. Then

1. There is a free set of size \aleph_0 .
2. If $\mu < \lambda$ and $\nu^{<\mu} < \lambda$ for all $\nu < \lambda$, then there is a free set of size μ .
3. If λ is weakly compact, then there is a free set of size λ .

We note that (3) is similar in spirit to the equivalent of weak compactness discussed under large cardinals (Section 2); this was from [13] (Erdős and Hajnal, 1974). Paper [13] mentions that [14] is forthcoming and will give some further information.

In (3), one cannot simply replace “weakly compact” by “strongly inaccessible”, since they point out that there is a counter-example if there is a λ -Suslin tree. Identifying the tree with the set λ , $g(x)$ is simply the set of nodes below x ; here, $\text{ran}(g)$ even has the stronger Condition A. It is not clear whether one can build a counter-example from a λ -Aronszajn tree. But it is also unknown whether there must be a λ -Suslin tree whenever λ is strongly inaccessible and not weakly compact; of course, this is true in L by a well-known result of Jensen.

9. CONCLUSION

This centennial volume documents the contributions of Paul Erdős to many diverse branches of mathematics. We hope that this brief survey has shown the broad range of his contributions within the particular area of set theory.

REFERENCES

Many of the papers of Erdős are on line at

http://www.renyi.hu/~p_erdos/Erdos.html

For these, we include the paper's label from that web site; e.g., [1943-08].

- [1] F. Bagemihl, The existence of an everywhere dense independent set, *Michigan Math. J.* 20 (1973) 112.
- [2] R. O. Davies, Partitioning the plane into denumerably many sets without repeated distances, *Proc. Cambridge Philos. Soc.* 72 (1972) 179–183.
- [3] P. Erdős, Some remarks on set theory, *Ann. of Math.* (2) 44 (1943) 643–646. [1943-08]
- [4] P. Erdős, Some remarks on set theory, *Proc. Amer. Math. Soc.* 1 (1950) 127–141. [1950-13]
- [5] P. Erdős, Some remarks on set theory III, *Michigan Math. J.* 2 (1954) 51–57. [1954-08]
- [6] P. Erdős, Some remarks on set theory IV, *Michigan Math. J.* 2 (1953-54) 169–173 (1955). [1955-14]
- [7] P. Erdős, Set-theoretic, measure-theoretic, combinatorial, and number-theoretic problems concerning point sets in Euclidean space, *Real Anal. Exchange* 4 (1978/79) 113–138. [1978-40]
- [8] P. Erdős and A. Hajnal, On the structure of set-mappings, *Acta Math. Acad. Sci. Hungar.* 9 (1958) 111–131. [1958-12]
- [9] P. Erdős and A. Hajnal, On a property of families of sets, *Acta Math. Acad. Sci. Hungar.* 12 (1961) 87–123. [1961-11]
- [10] P. Erdős and A. Hajnal, On a classification of denumerable order types and an application to the partition calculus, *Fund. Math.* 51 (1962/1963) 117–129. [1962-06]
- [11] P. Erdős and A. Hajnal, Some remarks concerning our paper “On the structure of set mappings”. Non-existence of a two-valued σ -measure for the first uncountable inaccessible cardinal, *Acta Math. Acad. Sci. Hungar.* 13 (1962) 223–226. [1962-20]
- [12] P. Erdős and A. Hajnal, On a problem of B. Jónsson, *Bull. Acad. Polon. Sci. Sr. Sci. Math. Astr. Phys.* 14 (1966) 19–23. [1966-05]
- [13] P. Erdős and A. Hajnal, Some remarks on set theory XI, *Fund. Math.* 81 (1974), 261–265. [1974-33]
- [14] P. Erdős, A. Hajnal, and A. Máté, Chain conditions on set mappings and free sets, *Acta Sci. Math. (Szeged)* 34 (1973) 69–79. [1973-03]

- [15] P. Erdős, S. Jackson, and R. D. Mauldin, On partitions of lines and space, *Fund. Math.* 145 (1994) 101–119.
- [16] P. Erdős, S. Jackson, and R. D. Mauldin, On infinite partitions of lines and space, *Fund. Math.* 152 (1997) 75–95.
- [17] P. Erdős and S. Kakutani, On non-denumerable graphs, *Bull. Amer. Math. Soc.* 49 (1943) 457–461. [1943–05]
- [18] P. Erdős and P. Komjáth, Countable decompositions of R^2 and R^3 . *Discrete Comput. Geom.* 5 (1990) 325–331.
- [19] P. Erdős, K. Kunen, and R. D. Mauldin, Some additive properties of sets of real numbers, *Fund. Math.* 113 (1981) 187–199. [1981–28]
- [20] P. Erdős and M. Makkai, Some remarks on set theory X, *Studia Sci. Math. Hungar.* 1 (1966) 157–159. [1966–19]
- [21] P. Erdős and R. Rado, A partition calculus in set theory, *Bull. Amer. Math. Soc.* 62 (1956) 427–489. [1956–02]
- [22] P. Erdős and M. E. Rudin, A non-normal box product, in *Infinite and Finite Sets* (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II, Colloq. Math. Soc. János Bolyai, Vol. 10, North-Holland, 1975, pp. 629–631. [1975–01]
- [23] P. Erdős and A. Tarski, On families of mutually exclusive sets, *Ann. of Math.* (2) 44 (1943) 315–329. [1943–04]
- [24] P. Erdős and A. Tarski, On some problems involving inaccessible cardinals, in *Essays on the Foundations of Mathematics* Magnes Press, Hebrew Univ., Jerusalem, 1961, pp. 50–82. [1961–14]
- [25] A. Hajnal, Proof of a conjecture of S. Ruziewicz, *Fund. Math.* 50 (1961/1962) 123–128.
- [26] A. Hajnal and P. Hamburger, *Set theory*, Cambridge University Press, 1999.
- [27] A. Hajnal, I. Juhász, and S. Shelah, Strongly almost disjoint families, revisited, *Fund. Math.* 163 (2000) 13–23.
- [28] F. Hausdorff, Grundzüge einer Theorie der Geordnete Mengen, *Math. Ann.* 65 (1908) 435–505.
- [29] T. Jech, *Set Theory*, The third millennium edition, Springer-Verlag, 2003.
- [30] A. Kanamori, *The Higher Infinite. Large cardinals in set theory from their beginnings*, Second edition, Springer-Verlag, 2003.
- [31] H. J. Keisler, Some applications of the theory of models to set theory, in *Logic, Methodology and Philosophy of Science* (Proc. 1960 Internat. Congr.), Stanford Univ. Press, 1962, pp. 80–86.
- [32] H. J. Keisler, Six classes of theories, *J. Austral. Math. Soc. Ser. A* 21 (1976) 257–266.
- [33] H. J. Keisler and A. Tarski, From accessible to inaccessible cardinals, *Fund. Math.* 53 (1963/1964) 225–308.
- [34] K. Kunen, Indescribability and the continuum, in *Axiomatic Set Theory* (Proc. Sympos. Pure Math., Vol. XIII, Part I), pp. 199–203, Amer. Math. Soc., 1971.
- [35] K. Kunen, Elementary embeddings and infinitary combinatorics, *J. Symbolic Logic* 36 (1971) 407–413.

- [36] K. Kunen, Some comment on box products, in *Infinite and Finite Sets* (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II, Colloq. Math. Soc. János Bolyai, Vol. 10, North-Holland, 1975, pp. 1011–1016.
- [37] K. Kunen, Partitioning Euclidean space, *Math. Proc. Cambridge Philos. Soc.* 102 (1987) 379–383. 05A17 (04A30 51M20)
- [38] K. Kunen, *Set Theory*, College Publications, 2011.
- [39] K. Kunen and F. D. Tall, Between Martin’s axiom and Souslin’s hypothesis. *Fund. Math.* 102 (1979) 173–181.
- [40] J. Łoś, Linear equations and pure subgroups, *Bull. Acad. Polon. Sci. Sr. Sci. Math. Astr. Phys.* 7 (1959) 13–18.
- [41] E. W. Miller, On a property of families of sets, *Comptes Rendus Varsovie* 30 (1937) 31–38.
- [42] W. Mitchell, Aronszajn trees and the independence of the transfer property, *Ann. Math. Logic* 5 (1972/73) 21–46.
- [43] M. E. Rudin, Countable box products of ordinals, *Trans. Amer. Math. Soc.* 192 (1974) 121–128.
- [44] J. H. Schmerl, Countable partitions of Euclidean space, *Math. Proc. Cambridge Philos. Soc.* 120 (1996) 7–12.
- [45] D. Scott, Measurable cardinals and constructible sets, *Bull. Acad. Polon. Sci. Sr. Sci. Math. Astronom. Phys.* 9 (1961) 521–524.
- [46] W. Sierpiński, *Hypothèse du Continu*, Second Edition, Chelsea Publishing Company, 1956 (first printed in 1934).
- [47] S. Shelah, Stability, the f.c.p., and superstability; model theoretic properties of formulas in first order theory, *Ann. Math. Logic* 3 (1971) 271–362.
- [48] S. Shelah, A combinatorial problem; stability and order for models and theories in infinitary languages, *Pacific J. Math.* 41 (1972) 247–261.
- [49] R. M. Solovay, A nonconstructible Δ^1_3 set of integers, *Trans. Amer. Math. Soc.* 127 (1967) 50–75.
- [50] M. Souslin, Problème 3, *Fundamenta Mathematicae* 1 (1920) p. 223.
- [51] F. D. Tall, Large cardinals for topologists, in *Surveys in General Topology*, Academic Press, 1980, pp. 445–477.
- [52] A. Tarski, Some problems and results relevant to the foundations of set theory, in *Logic, Methodology and Philosophy of Science* (Proc. 1960 Internat. Congr.), Stanford Univ. Press, 1962, pp. 125–135.
- [53] S. Ulam, Zur Masstheorie in der allgemeinen Mengenlehre, *Fund Math.* 16 (1930) 140–150.

Kenneth Kunen

*University of Wisconsin,
Madison,
WI 53706,
U.S.A.*

e-mail: kunen@math.wisc.edu

SOME PROBLEMS AND IDEAS OF ERDŐS IN ANALYSIS AND GEOMETRY

R. DANIEL MAULDIN

I review a meager few of the many problems and ideas Erdős proposed over the years involving a mixture of measure theory, geometry, and set theory.

1. INTRODUCTION

I have selected a few topics from Erdős' many problems and ideas in this area. Some were selected just for the sake of promoting them and others because they have led to several developments and connections. Three sources for some additional problems of Erdős in these areas may be found in [16, 17, 19].

2. SIMILAR COPIES OF SEQUENCES

Even in his article of 1978 in [16], Erdős says he had made the following conjecture for a long time:

Conjecture 2.1. *Let $\{x_n\}$ be a sequence of positive numbers decreasing to 0. Is there a Lebesgue measurable set E with positive measure which does not contain any affine copy of the sequence?*

In his lecture at the Scottish Book conference in 1979, Erdős said that the problem has been open for so long that he should offer \$100 for its solution. He also said at that time that he didn't think the problem was difficult. However, this well known and much studied problem remains open. This problem is discussed in some detail in [6] and more recently in the survey article [35].

3. ADDITIVE NUMBER THEORY AND EFFECTIVE DIMENSION

Erdős conjectured that to each infinite set of positive integers A , there corresponds a complementary set B , an infinite set of positive integers B with density 0 such that the sum set $A + B$ contains every sufficiently large integer. Lorentz proved the conjecture in [27]. In fact, letting $A(n)$ be the number of elements of A not exceeding n , Lorentz proved the following

Theorem 3.1. *There is a constant c such that every infinite set $A \subset \mathbb{N}$, there corresponds an infinite set $B \subset \mathbb{N}$ such that $A + B$ contains every sufficiently large integer and for each n :*

$$(1) \quad B(n) \leq c \sum_{k=1}^n \frac{\log A(k)}{A(k)}.$$

Inequality (1) clearly shows B has density 0. Erdős in [13] shows that inequality (1) is the best possible if one only takes into account the rate of increase of $A(n)$ but not its structural properties:

Theorem 3.2. *There is a sequence A of positive integers with positive lower density such that for every complementary set B satisfies $B(n) > C_1(\log n)^2$. This is in agreement with estimate (1).*

Erdős also made an improvement if A is the set of primes. For this set, Lorentz's estimate yields the existence of a complementary set B with $B(n) < C_2(\log n)^3$. Erdős shows there is some B with $B(n) < C_3(\log n)^2$.

By the way, in [13] Erdős posed the following problem.

Problem 3.3. Is there a set B of positive integers with $B(n) < C_4 \frac{n}{\log n}$ such that the sets $B + 2^k$ cover all but finitely many positive integers?

In [33] Ruzsa gave an affirmative answer and later in [34], he even determined the best constant. Lorentz proceeds to prove Theorem 3.1 by first proving a finite version of it:

Theorem 3.4. *There is a constant C such that if m and n are integers, k is a positive integer, and A is a set of integers with $A \subset [m, m+k)$ with $\text{card}(A) \geq l \geq 2$, then there are integers $b_1 < b_2 < \dots < b_K$ in the interval $(n-k, n+k)$ such that the translates $A + b_i$ cover the integers in the interval $(m+n, m+n+k)$ and*

$$(2) \quad K \leq Ck \frac{\log l}{l}.$$

The idea is to select the b_j 's greedily and estimate the number of steps required until the interval is covered. An immediate consequence of Lorentz's estimate is:

Theorem 3.5. *Let n be a positive integer. If a_1, \dots, a_l is a set of incongruent residues modulo n , there is another set of residues b_1, \dots, b_k with*

$$(3) \quad k \leq Cn \frac{\log l}{l}$$

such that each residue modulo n is of the form $a_i + b_j$.

If one would like to somehow measure the structural properties of A , a finite set of, say, positive integers with cardinality at least 2, one could consider what Randall Dougherty calls $den_{cover}(A)$, the 'covering density of A .' This is defined as follows. For each n , let $C(A, n)$ be the minimal number of translated copies of A needed to cover $[1, n] \cap \mathbb{N}$. Then

$$(4) \quad den_{cover}(A) = \lim_{n \rightarrow \infty} \frac{\text{card}(A)C(A, n)}{n}.$$

Clearly, $den_{cover}(\{1, 2, 3\}) = 1$. But, $den_{cover}(\{1, 2, 4\}) = 6/5$.

In another direction, Erdős, Kunen and I in [18] used Lorentz's theorem to prove the following:

Theorem 3.6. *Let P be a nonempty perfect subset of \mathbb{R} . Then there is a perfect set M with Lebesgue measure zero such that $P + M = \mathbb{R}$.*

One could consider extensions of these theorems and ideas to groups other than \mathbb{R} .

P. Elias in [10] has obtained a stronger form of Theorem 3.6. Using Kronecker's approximation theorem, he has shown that the set M of the theorem may be taken to be Dirichlet set. A set M is said to be a Dirichlet set if there exists an increasing sequence of positive integers $\{n_k\}$ such that the sequence of functions $\{\sin n_k x\}$ converges uniformly to 0 on M .

Also, Lorentz's theorem has a direct application in *effective geometric measure theory*. The *Kolmogorov complexity* of a string σ , denoted $K(\sigma)$, is the length (in this paper we will measure length in ternary units) of the shortest program (under a fixed universal machine) which outputs σ [26]. For a real number x , $x \upharpoonright n$ denotes the first n digits in a ternary expansion of x . Martin-Löf random reals have high initial segment complexity [8]; indeed every Martin-Löf random real r satisfies $\lim_n K(r \upharpoonright n)/n = 1$. This fact conforms with our intuition that the M-L random objects do not compress much.

Recall some classical dimension notions. Let $E \subseteq \mathbb{R}^n$. The *diameter* of E , denoted $|E|$, is the supremum of the distances between any two points in E . A *cover* \mathcal{G} for a set E is a collection of sets whose union contains E , and \mathcal{G} is a δ -*mesh cover* if the diameter of each member G is at most δ . For a number $\beta \geq 0$, the β -*dimensional Hausdorff measure* of E , written $\mathcal{H}^\beta(E)$, is given by $\lim_{\delta \rightarrow 0} \mathcal{H}_\delta^\beta(E)$ where

$$(5) \quad \mathcal{H}_\delta^\beta(E) = \inf \left\{ \sum_{G \in \mathcal{G}} |G|^\beta : \mathcal{G} \text{ is a countable } \delta\text{-mesh cover of } E \right\}.$$

The *Hausdorff dimension* of a set E , denoted $\dim_H(E)$, is the unique number α where the α -dimensional Hausdorff measure of E transitions from being negligible to being infinitely large; if $\beta < \alpha$, then $\mathcal{H}^\beta(E) = \infty$ and if $\beta > \alpha$, then $\mathcal{H}^\beta(E) = 0$ [22].

The *effective* (or *constructive*) β -*dimensional Hausdorff measure* of a set E , $c\mathcal{H}^\beta(E)$, is defined exactly in the same way as Hausdorff measure with the restriction that the covers be uniformly c.e. (= computably enumerable) open sets [8, Definition 13.3.3]. This yields the corresponding notion of the *effective* (or *constructive*) *Hausdorff dimension* of a set E , $\text{cdim}_H E$.

Lutz [28] showed that constructive dimension of a set is determined by the constructive dimension of its points:

$$(6) \quad \text{cdim}_H E = \sup\{\text{cdim}_H\{x\} : x \in E\},$$

and from work of Mayordomo [32](\geq) and Levin [25](\leq) (also see [8]) we have for any real number x ,

$$(7) \quad \text{cdim}_H\{x\} = \liminf_{n \rightarrow \infty} \frac{K(x \upharpoonright n)}{n}.$$

We define the *constructive dimension* of a point x to be the effective Hausdorff dimension of the singleton $\{x\}$. In [7], Lorentz's theorem plays a central role in the proof of the following.

Theorem 3.7. *Let C be the standard middle-third Cantor set. For any α satisfying $1 - \dim_H(C) \leq \alpha \leq 1$, and for any Martin-Löf random $r \in [0, 1]$, we have*

$$\dim_H((C + r) \cap E_{=\alpha}) = \dim_H((C + r) \cap E_{\leq\alpha}) = \alpha - 1 + \dim_H(C),$$

where $E_{=\alpha}$ consists of all real numbers with constructible dimension α and $E_{\leq\alpha}$ is the set of reals of dimension at most α .

The constructive dimension of $(C + r) \cap E_{=\alpha}$ is α whereas the Hausdorff dimension of this set is $\alpha - 1 + \dim_{\mathbb{H}}(C)$. This means that for a given M-L random real r there are many points x in the Cantor set which cancels the randomness of r , i.e., $x + r$ has lower constructive dimension; the initial strings of $r + x$ have a factor less Kolmogorov complexity than the corresponding initial strings of r .

It seems that we have just begun to delve into the possibilities in this direction. For example, one could investigate analogues of Theorem 3.7 for other totally disconnected self similar or self conformal sets in \mathbb{R} or \mathbb{R}^n .

4. DIMENSION OF SUBGROUPS AND RINGS

Erdős and Volkmann in [15] proved the following theorem.

Theorem 4.1. *For each α with $0 < \alpha < 1$, there is an additive Borel subgroup of the reals with Hausdorff dimension α .*

Several proofs of this fact have now been given. They all involve some set of numbers which are well approximated by rationals. For example, (see [22]), fix $0 < \alpha < 1$ and let n_k be a sequence of positive integers which increases sufficiently rapidly. Let

$$(8) \quad x \in G \iff \exists M \forall k \exists \text{ integer } p : \left| x - \frac{p}{n_k} \right| < \frac{M}{n_k^\alpha}.$$

Clearly, G is an additive subgroup of \mathbb{R} and it can be shown that $\dim_{\mathbb{H}}(G) = \alpha$. However, if one asks about subrings of \mathbb{R} , Edgar and Miller [9] showed the answer is quite different.

Theorem 4.2. *If the Borel set F is a subring of \mathbb{R} , then either $\dim_{\mathbb{H}}(F) = 0$ or $\dim_{\mathbb{H}}F = 1$.*

In fact, Edgar and Miller show that

Theorem 4.3. *If the Borel set F is a subring of \mathbb{C} , then either $\dim_{\mathbb{H}}(F) = 0$ or $F = \mathbb{R}$ or $F = \mathbb{C}$.*

Independently, Bourgain [1] also proved Theorem 4.2 by more delicate quantitative methods. This leads to the following problem.

Problem 4.4. For which α other than 0, 1 or 2 are there subrings of \mathbb{R} or \mathbb{C} with Hausdorff dimension α ?

This is really a question about transfinite constructions. Things are not so clear for other rings. Consider the example of D. Goldstein.

Example 4.5. Let the Borel set G be an additive subgroup of \mathbb{R} with $\dim_{\mathbb{H}}(G) = \alpha$. Let F consist of all 2×2 matrices M of the form

$$M = \begin{bmatrix} m & x \\ 0 & n \end{bmatrix},$$

where $x \in G$ and $m, n \in \mathbb{Z}$.

Then for any matrix norm, we have for the Borel subring F , $\dim_{\mathbb{H}}(F) = \alpha$.

Thus, for every α with $0 \leq \alpha \leq 1$, there are Borel subrings of the space of 2×2 matrices with dimension α . But we don't know the answer for larger α .

Problem 4.6. For which $\alpha > 1$ does the space of 2×2 real valued matrices have a (Borel) subring with Hausdorff dimension α ? Of course, one can consider this problem in a more general context.

Buhler, Butler, de Launey and Graham in [5] investigated 'Origami rings' in \mathbb{C} generated as follows. Let $L_{\alpha}(p)$ be the line in the complex plane through p with angle α . Given a collection U of angles, let $R(U)$ be the points that can be obtained by starting with 0 and 1, and then recursively adding intersection points of the form $L_{\alpha}(p) \cap L_{\beta}(q)$, where p, q have been already been generated, and α, β are in U and the lines are distinct. For each n , let U_n be the group of the n equally spaced angles $k\pi/n$, $0 \leq k < n$. They characterize the subrings of \mathbb{C} generated by the finite subgroups U where $3 \leq \text{card}(U)$ as follows.

Theorem 4.7. *Let $n \geq 3$. If n is prime, the $R(U_n) = \mathbb{Z}[\zeta_n]$, the cyclotomic integer ring. If n is not a prime, then $R(U_n) = \mathbb{Z}[1/n, \zeta_n]$, the cyclotomic integer ring localized at the primes dividing n . Moreover, if $n > 3$, then $R(U_n)$ is dense in the plane.*

This led Goldstein and I to construct uncountable subgroups G of the circle group which are the union of countably many compact sets each with box counting dimension 0. (Actually, such subgroups had been constructed much earlier by Laczkovich and Ruzsa in [29].) It follows from this that the subring of \mathbb{C} generated by G still has Hausdorff dimension 0. This leads to the following problem.

Problem 4.8. Is there a subgroup G of the circle group with $\dim_{\mathbb{H}}(G) = 0$ such that G is not the union of countably many sets with lower box counting dimension 0 and yet the ring generated by G still has dimension 0?

5. SETS CONTAINING THE VERTICES OF A TRIANGLE OF AREA 1

Many years ago Erdős noted that if E is a Lebesgue measurable subset of the plane with infinite measure, then for every $c > 0$, E contains the vertices of a triangle of area c . As several people have noted, this remains true if E has positive measure and is unbounded.

In [16] and again in [17, 19], Erdős poses what he said was an interesting and perhaps difficult problem, even though, as far as I know, he never did offer any money for its solution.

Problem 5.1. Is there a finite constant C such that if a Lebesgue measurable set E has measure greater than C , then E contains the vertices of a triangle of area 1? Moreover, is it true that the best constant is $c_0 = 4\pi/3\sqrt{3}$, the area of the disk such that the area of the inscribed equilateral triangle is 1?

Chris Freiling and I have studied this problem. Using some standard approximations in measure theory, Erdős' problem is equivalent to the following problem.

Problem 5.2. Is there a finite constant c such that for every $n \in \mathbb{N}$ if E is the union of the interiors of no more than n compact convex sets and E has measure greater than c , then E contains the vertices of a triangle of area 1. Moreover, is c_0 the best possible constant?

We showed in [30] that the constant c_0 is the best possible if n is 1. I reiterate the argument here. Suppose one has then a compact convex set K of positive area which is "small" meaning K does not contain the vertices of a triangle of area greater than 1. If one takes a line l , then the Steiner symmetrical of K about l has the same area as K and also does not contain the vertices of a triangle of area greater than 1. There is a sequence K_m , each of which is obtained by iterating the process of taking Steiner symmetrizations of K about a finite number of lines through the origin which converges to the closed disk centered at the origin with the same area as K , (see [36]). From this, it follows that the area of K is no more than c_0 . So, Erdős' conjecture is true if $n = 1$.

Let E be the union of the interiors of the compact convex sets K_1, \dots, K_n and suppose E does not contain the vertices of a triangle of area 1. Then the area of any triangle whose vertices belong to two of the sets K_i must be no more than 1. If i, j, k are different, then either the area of every triangle with one vertex from each of K_i, K_j, K_k is at most 1, or the areas of all such triangles is at least 1.

For $n = 2$, c_0 is still the best constant. If we have two compact convex bodies K_1 and K_2 such that their union does not contain the vertices of a triangle of area greater than 1, then their compact convex hull doesn't either, (see [30]).

Even for $n = 3$, Freiling and I argue c_0 is the best constant. Suppose $E = E_1 \cup E_2 \cup E_3$, where each set E_i is the interior of a compact convex set K_i . If E_1, E_2 , and E_3 form a "small" triple, i.e., the area of every triangle with its vertices in different sets E_i has area less than 1, then since their closed convex hull would have no triangle with area greater than 1 (see [30]), we are reduced to the case $n = 1$. On the other hand, if E_1, E_2 , and E_3 form a "large" triple, we use the following redistribution of mass argument. Let us suppose K_1 has the smallest area of the three bodies. There must be a line L which supports both K_2 and K_3 such that K_2 and K_3 lie in one half plane determined by L and K_1 lies in the interior of the other half plane. Let $A \in L \cap K_2$ and $B \in L \cap K_3$. Let $C \in K_1$. The triangle with vertices A, B and C must have area at least 1. Let us take lines l parallel to L and cutting the interior of both K_2 and K_3 . The line l intersect K_2 in points A_1 and A_2 and meets K_3 in points B_1 and B_2 , where A_2 and B_1 are closer together than A_1 and B_2 . Since the area of triangle A_1A_2C is no more than one and the area of triangle A_2B_1C is at least 1, $\|A_1 - A_2\| \leq \|A_2 - B_1\|$. Similarly, $\|B_1 - B_2\| \leq \|A_2 - B_1\|$. This is so for lines l until we reach a line that is a support line to either K_2 or to K_3 . In either case, this implies the area we have swept out between K_2 and K_3 is at least the area of the smaller of the areas of K_2 and K_3 and therefore the area is at least as large as the area of K_1 . So, if we replace the three bodies K_1, K_2 , and K_3 with the single body formed by K_2, K_3 and the area between them, we are back to the case $n = 1$.

The case $n = 4$ is still open.

6. PARTITIONS OF LINES AND PLANES

In generalizing a result of Sierpinski, Erdős in [14] proved the following.

Theorem 6.1. *The following two statements are equivalent:*

- (1) *CH, the continuum hypothesis holds: $2^\omega = \omega_1$.*
- (2) *If the lines in \mathbb{R}^2 (\mathbb{R}^3) are colored with 2 colors, then there exists a coloring of \mathbb{R}^2 (\mathbb{R}^3) with the same colors such that each line contains only countably many points with its color.*

Erdős, Jackson and I in [20] answered one of Erdős' question in [14] by proving the following.

Theorem 6.2. *The following two statements are equivalent:*

- (1) *CH, the continuum hypothesis holds: $2^\omega = \omega_1$.*
- (2) *If the lines in \mathbb{R}^2 (\mathbb{R}^3) are colored with three colors, then there exists a coloring of \mathbb{R}^2 (\mathbb{R}^3) with the same colors such that each line contains only finitely many points with its color.*

These results and several others involving flats in \mathbb{R}^n , $n \geq 2$ are discussed in [30]. Recently, Humke and Laczkovich used Erdős's original result to show that assuming CH holds there are subsets of the plane with some very unusual linear density properties [23]. One can imagine that there are several other types of strange examples using other partition results.

7. EXACT DIMENSION OF CONTINUED FRACTIONS USING ONLY THE PRIMES

In [31], Urbanski and I studied S_I , the set of continued fractions of the form

$$\frac{1}{b_1 + \frac{1}{b_2 + \frac{1}{b_3 + \frac{1}{\ddots}}}}$$

where I is a fixed subset of \mathbb{N} and each $b_n \in I$. We developed a *pressure* function which allowed us to determine the Hausdorff dimension $\alpha = \alpha_I$ of S_I . We showed that for those sets I for which the pressure function has a zero, there is a natural *conformal* probability measure supported on S_I and a corresponding Gauss measure, a measure supported on S_I equivalent to the conformal measure and which is invariant under the shift map on S_I . Using further properties such as the generalized density of I we found some conditions to determine whether $\mathcal{H}^\alpha(S_I)$ is 0, positive and finite, or ∞ . We also found some conditions such as some properties of the gaps in I which help to determine whether the α -dimensional packing measure $\mathcal{P}^\alpha(S_I)$ is 0, positive and finite, or ∞ . For example, if $p \geq 2$ and $I = \{n^p : n \in \mathbb{N}\}$, then $0 < \mathcal{H}^\alpha(S_I) < \infty$ and $\mathcal{P}^\alpha(S_I) = \infty$. On the other hand, if I has bounded gaps, then $\mathcal{P}^\alpha(S_I) < \infty$. If I is the set of primes, using Erdős' theorem that there are arbitrarily large two sided gaps in the sequence of primes [11], we showed that there is a conformal measure and a corresponding Gauss measure for this system, and yet $0 = \mathcal{H}^\alpha(S_I)$ and $\mathcal{P}^\alpha(S_I) = \infty$. A natural question which we posed in [31] is:

Problem 7.1. Let S be the set of all standard continued fractions of the form

$$\frac{1}{b_1 + \frac{1}{b_2 + \frac{1}{b_3 + \frac{1}{\ddots}}}}$$

where each b_i is a prime. Is there a Hausdorff gauge function g of the form $g(t) = t^\alpha L(t)$, where $L(t)$ is slowly varying such that $0 < \mathcal{H}^g(S) < \infty$?

If I is a finite subset of \mathbb{N} or if $I = \mathbb{N}$, then both $\mathcal{H}^\alpha(S_I)$ and $\mathcal{P}^\alpha(S_I)$ are positive and finite. We also don't know the answer to the following problem:

Problem 7.2. Is there a proper infinite subset I of \mathbb{N} such that both $\mathcal{H}^\alpha(S_I)$ and $\mathcal{P}^\alpha(S_I)$ are positive and finite?

I wish to thank Miklos Laczkovich for his help in preparing this paper.

REFERENCES

- [1] J. Bourgain, On the Erdős–Volkmann and Katz–Tao Ring Conjectures, *Geom. Funct. Anal.*, 19 (2003), 334–365.
- [2] Z. Buczolich and R. D. Mauldin, On the convergence of series of translates for measurable functions, *Mathematika*, 46 (1999), 337–341.
- [3] Z. Buczolich, J.-P. Kahane and R. D. Mauldin, On series of translates of positive functions, *Acta Math. Hungarica*, 98 (2001), 171–188.
- [4] Z. Buczolich and R. D. Mauldin, On series of translates of positive functions II, *Indagationes Math.*, 12 (2001), 317–327.
- [5] J. P. Buhler, S. Butler, W. de Launey, R. Graham, Rings Arising in Oragami, *J. Australian Math. Soc.*, to appear.
- [6] H. T. Croft, K. J. Falconer, Richard K. Guy, *Unsolved problems in geometry*, Springer-Verlag, Berlin, 1994.
- [7] R. Dougherty, J. H. Lutz, R. D. Mauldin, J. Teutsch, Translating the Cantor set by a random real, *Trans. Amer. Math. Soc.*, to appear.
- [8] R. G. Downey and D. R. Hirschfeldt, *Algorithmic randomness and complexity, Theory and Applications of Computability*, Springer, New York, 2010.
- [9] G. A. Edgar and C. Miller, Borel subrings of the reals, *Proc. Amer. Math. Soc.*, 131, 1121–1129, 2002.
- [10] P. Elias, Dirichlet sets, Erdős–Kunen–Mauldin theorem and analytic subgroups of the reals. *Proc. Amer. Math. Soc.*, 139, (2010), 2093–2104.

- [11] P. Erdős, On the difference of consecutive primes, *Quart. J. Oxford*, 6 (1935), 124–128.
- [12] P. Erdős, On the strong law of large numbers, *Trans. Amer. Math. Soc.*, 67 (1949), 51–56.
- [13] P. Erdős, Some results on additive number theory, *Proc. Amer. Math. Soc.*, 5 (1954), 847–853.
- [14] P. Erdős, Some remarks on set theory. IV, *Mich Math. J.*, 2 (1953–54), 169–173 (1955).
- [15] P. Erdős and B. Volkmann, Additive Gruppen mit vorgegebener Hausdorffscher dimension. *J. Reine Angew. Math.*, 221 (1966), 203–208.
- [16] P. Erdős, Set-theoretic, measure-theoretic, combinatorial and number-theoretic problems concerning point sets in Euclidean space, *Real Anal. Exchange*, 4, no. 2, (1978/79), 113–138.
- [17] P. Erdős, My Scottish Book Problems, in: *The Scottish Book, Mathematics from the Scottish Café*. Edited by R. Daniel Mauldin, *Birkhäuser*, Boston, Mass., 1981.
- [18] P. Erdős, K. Kunen, R. D. Mauldin, Some additive properties of sets of real numbers, *Fund. Math.*, 113 (1981), 187–199.
- [19] P. Erdős, Some combinatorial, geometric and set theoretic problems in measure theory, in *Measure Theory*, Oberwolfach 1983, *Lecture Notes in Mathematics* 1089, Springer-Verlag (1984).
- [20] P. Erdős, S. Jackson, R. D. Mauldin, On partitions of lines and planes, *Fund. Math.*, 145 (1994), 101–119.
- [21] K. J. Falconer, *The geometry of fractal sets*, Cambridge Tracts in Mathematics, vol. 85, Cambridge university press, Cambridge, 1986.
- [22] K. Falconer, *Fractal Geometry*, John Wiley & Sons Inc., Hoboken, NJ, second edition, 2003.
- [23] P. D. Humke, M. Laczkovich, Transference of Density, preprint, 2012.
- [24] D. Khoshnevisan, *Probability*, Graduate Studies in Mathematics, AMS, 2007.
- [25] L. A. Levin, The concept of a random sequence, *Dokl. Akad. Nauk SSSR*, 212 (1973), 548–550.
- [26] Ming Li and Paul Vitanyi, *An introduction to Kolmogorov complexity and its applications*, third ed., Texts in Computer Science, Springer, New York, 2008.
- [27] G. G. Lorentz, On a problem of additive number theory, *Proc. Amer. Math. Soc.*, 5 (1954), 838–841.
- [28] J. H. Lutz, *The dimensions of individual strings and sequences*, *Inform. and Comput.*, 187 (2003), no. 1, 49–79.
- [29] M. Laczkovich and I. Z. Rusza, Measure of sumsets and ejective sets I., *Real Analysis Exchange*, 22 (1996-1997), 153–166.
- [30] R. D. Mauldin, Some problems in set theory, analysis and geometry, in *Paul Erdős and his Mathematics I*, 493-505, Springer, 2002.
- [31] R. D. Mauldin and M. Urbanski, Conformal iterated function systems with applications to the geometry of continued fractions, *Trans. Amer. Math. Soc.*, 351 (1999), 4995–5025.

- [32] Elvira Mayordomo, A Kolmogorov complexity characterization of constructive Hausdorff dimension, *Inform. Process. Lett.*, 84 (2002), no. 1, 1–3.
- [33] I. Z. Ruzsa, On a problem of P. Erdős, *Canad. Math. Bull.*, 15 (1972), 309–310.
- [34] I. Z. Ruzsa, Additive completion of lacunary sequences, *Combinatorics*, 21 (2001), 279–291.
- [35] R. E. Svetic, The Erdős similarity problem, *Real Analysis Exchange*, 26(2) (2000), 525–540.
- [36] Roger Webster, *Convexity*, Oxford University Press, 1994.

R. Daniel Mauldin

*Department of Mathematics,
University of North Texas,
Denton,
TX 76203*

e-mail: `mauldin@unt.edu`

L^2 MAJORANT PRINCIPLES

HUGH L. MONTGOMERY

Dedicated to Erdős Pál on his centenary

In this short historical note, we discuss an important majorant principle introduced by Erdős & Fuchs [1].

1. INTRODUCTION

Let $R(x)$ be defined by the relation

$$(1.1) \quad \sum_{\substack{a,b \in \mathbb{Z} \\ a^2 + b^2 \leq x}} 1 = \pi x + R(x).$$

It is classical (see Hardy [4]) that

$$(1.2) \quad R(x) = \Omega((x \log x)^{1/4}).$$

This is shown by using what amounts to a Fourier expansion of $R(x)$. Erdős & Turán [2] considered the more general problem of counting sums of two members of a given sequence of non-negative integers. Suppose that $a_1 \leq a_2 \leq \dots$ is a sequence of non-negative integers, and let $R_A(x)$ be defined by the relation

$$(1.3) \quad \sum_{\substack{j,k \\ a_j + a_k \leq x}} 1 = cx + R_A(x)$$

where c is a suitable positive constant. Erdős & Turán conjectured that there is no sequence $\{a_j\}$ for which $R_A(x) = O(1)$. This was emphatically confirmed by Erdős & Fuchs [1], who showed that

$$(1.4) \quad R_A(x) = \Omega(x^{1/4}/(\log x)^{1/2}).$$

This is amazingly close to (1.2), considering that $\{a_j\}$ is an arbitrary sequence. Subsequently, Jurkat (unpublished), Hayashi (unpublished, but see [5]) and Montgomery & Vaughan [7] by three different methods have shown that

$$(1.5) \quad R_A(x) = \Omega(x^{1/4}).$$

Our focus on this occasion is not the Erdős–Fuchs theorem itself, but rather an important lemma that they introduced in their work. Suppose that $f \in L^1(\mathbb{T})$ with Fourier coefficients

$$\widehat{f}(n) = \int_0^1 f(x)e(-nx) dx$$

where $e(\theta) = e^{2\pi i\theta}$ is the complex exponential with period 1. Suppose also that $\widehat{f}(n) \geq 0$ for all n , and that $\sum_{-\infty}^{\infty} \widehat{f}(n) < \infty$. The Erdős–Fuchs lemma asserts that

$$(1.6) \quad \int_{-\theta}^{\theta} |f(x)|^2 dx \gg \theta \int_0^1 |f(x)|^2 dx$$

for any $\theta \in (0, 1/2]$. This lemma, and others like it, proved using the same ideas, are very useful. At the same time, and independently, Wiener & Wintner [9], [8, pages 758–764] showed that if $F(s)$ is the Laplace transform of a nonnegative function, convergent for $\Re s > 1$, then for any $\sigma > 1$, $a > 0$, $\varepsilon > 0$ we have

$$(1.7) \quad \int_{-a}^a |F(\sigma + it)|^2 dt \leq (8[a/\varepsilon] + 1) \int_{-\varepsilon}^{\varepsilon} |F(\sigma + it)|^2 dt.$$

This paper is rather poorly written, which makes the editorial notes of Bateman and Diamond [8, pages 788–790] especially valuable. In the same vein, Halász [3] showed that if $|b_n| \leq a_n$ for all n , and if $\sum_{n=1}^{\infty} a_n/n^\sigma < \infty$, then

$$(1.8) \quad \int_{T_0-1}^{T_0+1} \left| \sum_{n=1}^{\infty} \frac{b_n}{n^{\sigma+it}} \right|^2 dt \ll \int_{-1}^1 \left| \sum_{n=1}^{\infty} \frac{a_n}{n^{\sigma+it}} \right|^2 dt$$

uniformly in T_0 . The following polished form (due to Wirsing) of this principle implies all these results.

Theorem. *Suppose that $|b_n| \leq a_n$ for all n , that $\sum_{n=1}^\infty a_n < \infty$, and that $\lambda_1, \lambda_2, \dots$ are real numbers. Then*

$$(1.9) \quad \int_{-T}^T \left| \sum_{n=1}^\infty b_n e(\lambda_n t) \right|^2 dt \leq 3 \int_{-T}^T \left| \sum_{n=1}^\infty a_n e(\lambda_n t) \right|^2 dt.$$

Logan [6] has shown that the constant 3 is best-possible.

By applying the above with b_n replaced by $b_n e(\lambda_n T_0)$, we see that the above implies that

$$(1.10) \quad \int_{T_0-T}^{T_0+T} \left| \sum_{n=1}^\infty b_n e(\lambda_n t) \right|^2 dt \leq 3 \int_{-T}^T \left| \sum_{n=1}^\infty a_n e(\lambda_n t) \right|^2 dt$$

for all real T_0 . The special case of this with $b_n = a_n$ is noteworthy:

$$(1.11) \quad \int_{T_0-T}^{T_0+T} \left| \sum_{n=1}^\infty a_n e(\lambda_n t) \right|^2 dt \leq 3 \int_{-T}^T \left| \sum_{n=1}^\infty a_n e(\lambda_n t) \right|^2 dt.$$

This clearly implies the Erdős–Fuchs lemma (1.6). While our Theorem is a little more flexible, its proof involves only ideas already found in the Erdős–Fuchs proof.

2. PROOF OF THE THEOREM

Let $K(t) = \max(0, 1 - |t|/T)$. Then $K \in L^1(\mathbb{R})$, and

$$\widehat{K}(u) = T \left(\frac{\sin \pi T u}{\pi T u} \right)^2 \geq 0.$$

Thus

$$(2.1) \quad \begin{aligned} \int_{-T}^T K(t) \left| \sum_{n=1}^\infty b_n e(\lambda_n t) \right|^2 dt &= \sum_{m=1}^\infty \sum_{n=1}^\infty b_m \overline{b_n} \widehat{K}(\lambda_n - \lambda_m) \\ &\leq \sum_{m=1}^\infty \sum_{n=1}^\infty a_m a_n \widehat{K}(\lambda_n - \lambda_m) \\ &= \int_{-T}^T K(t) \left| \sum_{n=1}^\infty a_n e(\lambda_n t) \right|^2 dt. \end{aligned}$$

By replacing b_n by $b_n e(\lambda_n T_0)$, it follows that

$$(2.2) \quad \int_{T_0-T}^{T_0+T} K(t-T_0) \left| \sum_{n=1}^{\infty} b_n e(\lambda_n t) \right|^2 dt \leq \int_{-T}^T K(t) \left| \sum_{n=1}^{\infty} a_n e(\lambda_n t) \right|^2 dt$$

for any real T_0 . But

$$K(t+T) + K(t) + K(t-T) = \begin{cases} 0 & (t \leq -2T), \\ t+2T & (-2T \leq t \leq -T), \\ 1 & (-T \leq t \leq T), \\ 2T-t & (T \leq t \leq 2T), \\ 0 & (t \geq 2T). \end{cases}$$

Since this majorizes the characteristic function of the interval $[-T, T]$, it follows that the left hand side of (1.9) is

$$\leq \int_{-2T}^{2T} (K(t+T) + K(t) + K(t-T)) \left| \sum_{n=1}^{\infty} b_n e(\lambda_n t) \right|^2 dt.$$

By three applications of (2.1) it follows that the above is

$$\leq 3 \int_{-T}^T K(t) \left| \sum_{n=1}^{\infty} a_n e(\lambda_n t) \right|^2 dt \leq 3 \int_{-T}^T \left| \sum_{n=1}^{\infty} a_n e(\lambda_n t) \right|^2 dt.$$

This completes the proof.

REFERENCES

- [1] P. Erdős & W. H. J. Fuchs, *On a problem of additive number theory*, J. London Math. Soc. 31 (1956), 67–73.
- [2] P. Erdős & P. Turán, *On a problem of Sidon in additive number theory, and some related problems*, J. London Math. Soc. 16 (1941), 212–215; *addendum*, *ibid.* 19 (1944), 208.
- [3] G. Halász, *Über die Mittelwerte multiplicativer zahlentheoretischer Funktionen*, Acta Math. Acad. Sci. Hungar. 19 (1968), 365–403.
- [4] G. H. Hardy, *On the expression of a number as a sum of two squares*, Quart. J. Math. 46 (1915), 263–283.

- [5] E. K. Hayashi, *Omega theorems for the iterated additive convolution of a non-negative arithmetic function*, Ph.D. Thesis, University of Illinois at Urbana–Champaign, 1973.
- [6] B. F. Logan, *An interference problem for exponentials*, Michigan Math. J. 35 (1988), 369–393.
- [7] H. L. Montgomery & R. C. Vaughan, *On the Erdős–Fuchs theorems*, A Tribute to Paul Erdős. Cambridge University Press, 1990.
- [8] Norbert Wiener, *Collected Works with Commentaries*, Vol. II, MIT Press, 1979.
- [9] N. Wiener & A. Winter, *On a local L^2 -variant of Ikehara’s theorem*, Rev. Math. Cuyana 2 (1956), 53–59.

Hugh L. Montgomery

Department of Mathematics,
University of Michigan,
Ann Arbor,
MI 48109–1043,
USA

e-mail: hlm@umich.edu

A COMBINATORIAL CLASSIC – SPARSE GRAPHS WITH HIGH CHROMATIC NUMBER

JAROSLAV NEŠETŘIL*

Remembering dědeček Paul Erdős

1. INTRODUCTION

It seems that combinatorics, and graph theory in particular, reached mathematical maturity relatively recently. Perhaps as a result of this there are not too many essential stories which have determined the course of the subject over a long period, enduring stories which appear again and again as a source of inspiration and motivate and challenge research.

In this article we attempt to demonstrate one example of such a story which we believe motivated some of the key parts of modern combinatorics. (Of course there are other stories, see for example [59].) Moreover the main result is related to the central theme of this book – the work and mathematical legacy of Paul Erdős.

Let $G = (V, E)$ be an (undirected) graph. We need to recall only a few facts and definitions. The *chromatic number* $\chi(G)$ of G is the minimal number of classes (“colors”) of a partition of V into independent sets. A set $A \subseteq V$ is called *independent* if it doesn’t contain any edge. The maximal size of an independent set is denoted by $\alpha(G)$. It is obvious that

$$(1) \quad \alpha(G) \cdot \chi(G) \geq |V|$$

holds for every graph G .

This leads to the lower bound

$$(2) \quad \chi(G) \geq \frac{|V|}{\alpha(G)},$$

*Partially supported by the Project LL1201 ERCCZ CORES and by CE-ITI P202/12/G061 of GACR.

which is one of the very few lower bounds available for the chromatic number.

It is a classical (and folklore) result that a graph has chromatic number ≤ 2 iff it doesn't contain a cycle of odd length. An even easier statement is that a forest (i.e. a graph without any cycle) has chromatic number ≤ 2 .

The minimal length of a cycle in G is called the girth of G and denoted here by $girth(G)$. The central result of this paper has the following innocent form:

Theorem 1 (Erdős [23]). *For every choice of positive integers k and l there exists a graph $G_{k,l} = G$ with the following properties:*

- 1) $\chi(G) \geq k$;
- 2) $girth(G) > l$.

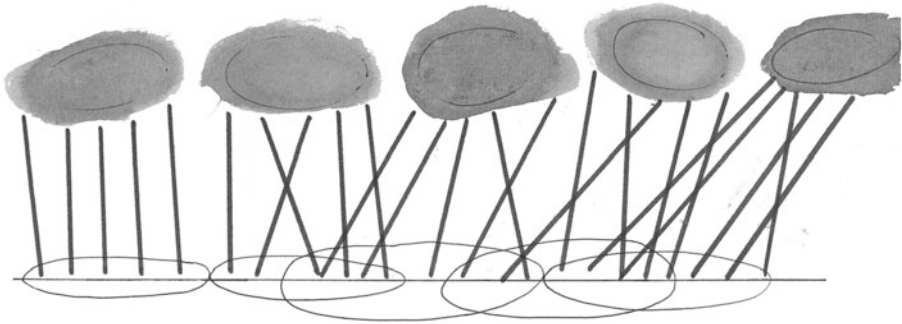
Thus the absence of a short cycle (of length $\leq l$) cannot guarantee bounded chromatic number. By interpreting the chromatic number as a dimension or as a measure of complexity we see that Theorem 1 claims that there exists high dimensional (or highly complex) graphs which are locally as trivial as forests (i.e. graphs without any cycles) can be. An old saying existing in several languages is very fitting here: We do not see the wood for the trees! Yes, these are paradoxical objects.

Theorem 1 is both a culmination of long development and the start of important consequent research and methods. The literature is large and we find Theorem 1 in most books dealing with graphs. With various proofs one can find it in many combinatorial graph theory books and particularly in books relating to probabilistic methods in combinatorics, see e.g. [11], [5], [39], [13], [21], [63], [10], [57]. In this survey we concentrate on various structural extensions and theoretical implications of Theorem 1 (and we indicate various proof methods).

2. EARLY CONSTRUCTIONS

Theorem 1 was proved in 1958 by Erdős in his seminal paper [23]. But already at that time this result was firmly based in advanced combinatorics and it also had an interesting history. Let us review it briefly for completeness from a contemporary perspective.

The first nontrivial instance of Theorem 1 is the case $l = 3$. In this form claims the existence of a triangle free graph $G_{k,3}$ with $\chi(G_{k,3}) > k$. This was proved independently by W. Tutte (alias Blanche Descartes) [20] and A. Zykov [92]. The proofs are constructive and can be visualised as follows:



Here is a more formal sketch: We proceed by induction on k . Given $G = G_{k,3}$ with n vertices we consider a set X with $k(n - 1) + 1$ vertices and for every subset $Y \subseteq X$, $|Y| = n$ we take an isomorphic copy G_Y of G every vertex of which is joined by a matching E_Y to Y . Denote by $G_{k+1,3}$ the graph consisting of all edges in all graphs G_Y , $Y \subseteq X$, $|Y| = n$, and all matchings E_Y . It is easy to see that $G_{k+1,3}$ has no triangles and, assuming $\chi(G_{k,2}) \geq k$, we get $\chi(G_{k+1,3}) \geq k + 1$. ■

Tutte’s construction is a prototype of many subsequent proofs and variants, as we shall see in Sections 4 and 5. Let us give some further constructions of triangle-free graphs (i.e. $l = 3$), most of which are regarded as classical.

Note that already in [43] it was observed that the above inductive construction does not even create cycles of length ≤ 5 . However this remained the best result (with respect to girth l) until [23].

Another early construction for $l = 3$ was provided by [66]. The construction proceeds again by induction on k : In each step we create a sibling x' for every vertex x and join x' to a vertex y if and only if x and y are joined. Then we add a (universal) vertex joined to all the siblings vertices produced. Call the resulting graph $M(G)$ (*Mycielskian* of G). $M(G)$ has no triangle and $\chi(M(G)) = \chi(G) + 1$. (Thus from K_2 we obtain C_5 and from C_5 the Grötzsch graph.)

An interesting variation of this construction of graphs $G_{k,3}$ is to iterate siblings. By this we mean that every vertex has siblings x_1, \dots, x_t and sibling x_{i+1} is joined to those siblings y for which $\{x, y\} \in E$. A universal vertex is then joined to all siblings x_t . These graph (and their variants) were studied in [32], [87], [8].

One of the simplest constructions is provided by the *shift graphs* S_n : the vertices of S_n are all pairs (a, b) of integers $1 \leq a < b \leq n$ with edges formed by pairs $(a, b)(b, c)$. Clearly S_n has no triangle (but contains large complete bipartite graphs) and $\chi(S_n) = \lceil \log n \rceil$. These remarkable graphs

are important in the infinite case as well and they can be traced to Erdős-Specker graphs [28].

Other early constructions of triangle-free graphs with high chromatic number are geometrical (distance graphs, see already [27]). A particularly elegant combinatorial geometric construction [19] was discovered in the context of computational complexity:

We consider the set of all *flags* (i.e. all incidence pairs (p, L)) in a projective plane of order k with an arbitrary linear ordering $<$. These are the vertices of our graph G . Vertices (p, L) and (p', L') will form an edge of G if $(p, L) < (p', L')$, all p, L, p', L' are distinct, and if $p \in L'$. This graph has no K_3 , and it can be shown that $\alpha(G) \leq k + 1$. Thus $\chi(G) \geq k^2 + k + 1$ as G has $(k^2 + k + 1)(k + 1)$ vertices.

Another by now classical example is provided by Kneser graphs. The *Kneser graph* $K\binom{n}{p}$ has as vertices all p -element subsets of $[n] = \{1, 2, \dots, n\}$. Edges of $K\binom{n}{p}$ are formed by pairs of disjoint sets. In (another) landmark paper [56] Lovász proved that $\chi(K\binom{n}{p}) = n - 2p + 2$. This (lower bound) was achieved by relating the coloring problem to algebraic topology. This powerful tool found many applications (see Matoušek's book [61] devoted to this subject). This is clearly an "advanced" construction (with which we deal in the next section) but it is related to girth 4 only. It follows that the Kneser graphs $K\binom{2m+k-2}{m}$ (for any m) provide another nice example in playing the role of $G_{k,3}$.

All these constructions have been thoroughly studied. Any new construction (such as [19] or [49]) is welcome with high hopes and then investigated thoroughly (see e.g. [8], [46] [32], [87]). But all these old-new constructions, which we have not listed exhaustively, are related to small girth. Indeed very small: $l \leq 6$ and mostly even $l = 3$. One should stress that the odd girth condition (i.e. the absence of short odd cycles) is in the context of chromatic number a much easier condition than the girth. Rectangles present a problem and there are structural reasons for it (see more on that in the last section).

3. ERDŐS THEME

Theorem 1 for $l = 3$ (i.e. the existence of triangle-free graphs with large chromatic number) provided only one part of the motivation for Erdős’ proof for general l . The other motivation (and certainly at that time for Erdős more important motivation) was the setting which relies on the inequality (2) and relates to Ramsey theory [30]. It is interesting to follow [23]:

Denote by $r(k, 3)$ the minimal number of vertices n such that every triangle-free graph G_n with n vertices contains an independent set of size k . Formally,

$$r(k, 3) = \min\{n; \text{ either } K_3 \subseteq G_n \text{ or } \alpha(G_n) \geq k\}.$$

Erdős proved in [23] that $r(k, 3) > k^{1+1/6}$, which using (2) implies that there are graphs $G_{k,3}$. The asymptotic behavior of Ramsey numbers $r(k, 3)$ was determined in a sequence of important papers [44], [1],

$$c_1 \frac{k^2}{\log k} \leq r(k, 3) \leq c_2 \frac{k^2}{\log k},$$

and gave rise to Rödl’s nibble, or semi-random methods [85], [5], [41]. The numbers $r(k, 3)$ were the first asymptotically known Ramsey numbers (since then there have been others, see [3], [4]).

Returning to the history of Theorem 1, Erdős made a stronger statement:

Theorem 2 [23]. *Let l be fixed, let $0 < \eta < \frac{1}{2l}$. For every sufficiently large n there exists a graph $G = (V, E)$ with n vertices and the following properties:*

- 1) $girth(G) > l$;
- 2) $\alpha(G) < n^{1-\eta}$.

(Note that this not only implies Theorem 1 but proves further that $\chi(G) > n^\epsilon$.)

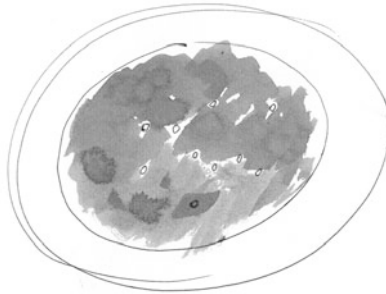
The proof of Theorem 2 is probabilistic and in fact it may be viewed as the cradle of the probabilistic method [88]. Nowadays it is found in every good graph theory book. Here is a very brief sketch:

We consider a random graph G with n vertices and $n^{1+\epsilon}$ edges, $\epsilon = 2\eta$, and prove that almost all such graphs satisfy 2) and that they contain $o(n)$ edges in cycles of length $\leq l$. They can then be deleted while 2) still holds.



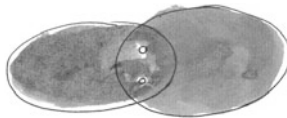
Putting it more poetically: A thin soup of $n^{1+\varepsilon}$ edges on n points contains (on average) few cycles but still shares some properties of the complete graph, namely small independence number.

How to depict this proof? Like a soup.



The proof of Theorem 2 has many variants and many more problems were solved by this method. Continuing in our line, for example, Erdős and Hajnal [25] generalized the result to hypergraphs as follows:

A p -uniform hypergraph is a pair (X, \mathcal{M}) , where $\mathcal{M} \subseteq \binom{X}{p} = \{M; M \subseteq X, |M| = p\}$. Elements of \mathcal{M} are still called edges. A cycle in (X, \mathcal{M}) and its length and girth are defined analogously as for graphs; a cycle of length 2 is formed by any pair of edges which intersect in (at least) 2 points.



Hypergraphs without 2-cycles are called simple (or linear).

The chromatic number $\chi(X, \mathcal{M})$ is defined analogously as for graphs: it is the minimal number of colors needed in a coloring of vertices so that no edge is monochromatic (this seems to be the most common definition of the chromatic number for hypergraphs; of course there are other possibilities).

Theorem 3 [25]. *Let $p \geq 2$, k, l be positive integers. Then there exists a p -uniform hypergraph $G_{p,k,l} = (X, \mathcal{M})$ such that*

- 1) $\chi(X, \mathcal{M}) \geq k$;
- 2) $\text{girth}(X, \mathcal{M}) > l$.

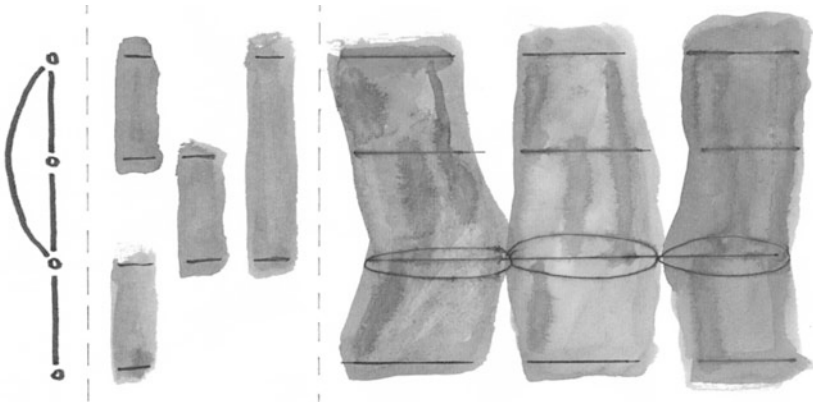
The probabilistic proofs of Theorems 2 and 3 are similar. Yet the connection between these two statements provide some challenging open problems (as we shall stress at several places in this article).

4. ADVANCED CONSTRUCTIONS

Already in Erdős’ paper [23] the question about a constructive proof of Theorem 1 is raised. The progress has been very slow here. The $l = 6$ barrier was broken only a decade later [67] and there were speculations about the untractability (in some sense) of the problem. Even from today’s point of view there is no easy (and elementarily justified) construction of a graph, say, $G_{4,27}$.

The first constructive proof of Theorem 1 was obtained by Lovász [55]. His striking proof is based on proving (the more general) Theorem 3, i.e. the existence of the hypergraph $G_{p,k,l}$. The proof proceeds by double induction on p and l (for a fixed k) and is too complicated to be explained here.

Another construction was provided in [77]. This is an outgrowth of structural Ramsey theory. It is called *partite construction* or *amalgamation construction* [80],[74], [75], [73] and in the structural Ramsey theory it seems to be one of the basic methods for obtaining structural results. The partite construction when applied to coloring of vertices is indicated by the following:



Here is a very rough sketch: Put $a = (p - 1)(k - 1) + 1$. We start with a system (V, \mathcal{M}) of p -tuples which are organised on the set $V = \bigcup_{i=1}^a V_i$, (the V_i are disjoint sets called *parts*), in such a way that for any p -tuple of parts V_{i_1}, \dots, V_{i_p} there exists an edge $M \in \mathcal{M}$ with $M \subseteq \bigcup_{j=1}^p V_{i_j}$. We call this the *partite system* P_0 .

In the inductive step we assume that we are given a partite system $P_{i-1} = (V, \mathcal{M})$ with $V = V_1 \cup \dots \cup V_a$. Put $|V_i| = P$ and apply induction to

get system (Y, \mathcal{N}) with properties $(P, k, l - 1)$. Now extend (like in Tutte's construction) every $N \in \mathcal{M}$ to a copy of P_{i-1} while keeping the distribution to parts. One then proves (see [38] for more details) that P_a has properties of $G_{p,k,l}$.

In a way the partite construction is a multipartite generalization of Tutte's construction. (The situation is not so straightforward for Ramsey theory and the amalgamation is more complicated.)

However for both Lovász's construction and as well as for the partite construction the size of the constructed (hyper)graph is not bounded by a tower function of bounded height. Even the (somewhat less precise) question whether one can prove Theorem 1 without referring to Theorem 3 was asked (and answered positively in [49]). Further variants of constructions of graphs $G_{k,l}$ are given in [50] and more recent [89].

One should stress that the size of the graphs $G_{k,l}$ is not merely a combinatorial question. The graphs $G_{k,l}$ are closely linked to special graphs used in the theory of algorithms and complexity theory. In particular, *expander graphs* (see for example extensive) [38] form a cornerstone of the modern theory of computing (see for example Ajtai-Komlos-Szemerédi [2]). One can see easily that large d -regular expander graphs with girth l may be used to construct graphs $G_{k,l}$.

A polynomial size construction of expanders, and thereby of graphs $G_{k,l}$, came as a real surprise from a different corner of mathematics as a combination of mainly harmonic analysis, number theory and algebraic graph theory. The resulting graphs, often called *Ramanujan graphs* defined by Margulis [60] and Lubotzky, Phillips and Sarnak [58], are fascinating in their own right.

There is a large literature (an interested reader may consult a survey article [38] and references given there) and several books. For completeness we state the main consequence for the context of this paper:

Let p, q be primes with Legendre symbol $(\frac{p}{q}) = 1$, q sufficiently larger than p . Then there exists a graph $X^{p,q} = (V, E)$ (we preserve the standard notation of these graphs) with the following properties:

- 1) $|V| = n = q(q^2 - 1)/2$. (The vertex set of $X^{p,q}$ is the set of points of the projective linear group $PSL_2(q)$);
- 2) $X^{p,q}$ is $(p + 1)$ -regular;
- 3) $girth(X^{p,q}) \geq 2 \log_p q$;
- 4) $\alpha(X^{p,q}) \leq \frac{2\sqrt{p}}{p+1}n$;
- 5) $\chi(X^{p,q}) \geq \frac{p+1}{2\sqrt{p}}$.

Hence $X^{p,q}$ can be chosen as an example of graph $G_{k,l}$ with at most k^{3l} vertices.

This whole area is a source of many applications (and beautiful mathematics) which exceeds the scope of this paper. But in passing let us stress that no full analogy of graphs $X^{p,q}$ is known for hypergraphs (see for example recent [54]). In particular, no small explicit construction of hypergraphs is known. The best result here is the work of G. Kun [51], where he constructs hypergraphs $G_{p,k,l}$ with number of vertices bounded by a primitive recursive function of p, k, l . This construction uses a “twisted” product to reduce the number of short cycles in a constructed hypergraph. One proves that this is a polynomial process, yet randomized at each step.

A fully deterministic small (or even bounded by a tower function of bounded height) construction of hypergraphs $G_{p,k,l}$ is still an open problem. Admittedly, however the derandomization techniques and advances of theoretical computer science make the “constructive questions” less clear (and probably less important too) than they were in the 1960s.

5. RANDOM PLACEMENT CONSTRUCTION

Here we present perhaps the simplest probabilistic proof of both Theorem 1 and 3 (however not Theorem 2). Surprisingly, this proof seems to be a little known. We need the following lemma [72].

Lemma 4. *Fix $p \geq 3, l \geq 3$ positive integers. For any $\varepsilon > 0$ there exist $n_0(p, l, \varepsilon)$ such that for every $n \geq n_0(p, l, \varepsilon)$ there exists a p -uniform hypergraph $H(n, p, l, \varepsilon) = (X, \mathcal{M})$ with the following properties:*

- 1) $|X| = n$;
- 2) $|\mathcal{M}| \geq n^{1+\varepsilon}$;
- 3) $girth(X, \mathcal{M}) \geq l$.

This (for graphs) is the easier part of Erdős proof of Theorem 2: one considers (for large n) random k -uniform hypergraph $(X, \mathcal{M}), X = \{1, 2, \dots, n\}$ with $m = \lceil 2n^{1+\frac{1}{\varepsilon}} \rceil$ edges. It is easy to prove that for these values of m , the average number of edges in a cycle of length $< l$ is $o(n)$. Taking a witness (X, \mathcal{M}) of this inequality and by deleting the corresponding edges in short cycles we get the desired hypergraph $H(n, p, l, \varepsilon)$.

However simple this lemma has many consequences.

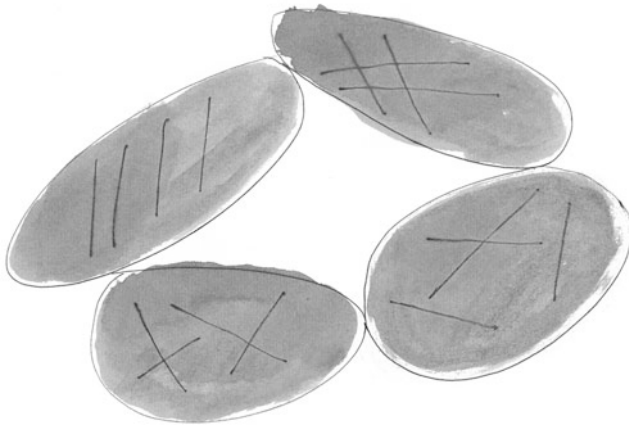
Proof of Theorem 3 ([72]). Let p, k, l be fixed. Put $P = (p - 1)(k - 1) + 1$ and consider $H(n, P, l, \varepsilon) = (X, \mathcal{M})$ as in Lemma 4. Let \mathcal{H} be the class of

all p -uniform hypergraphs (X, \mathcal{N}) where every $M \in \mathcal{M}$ contains exactly one edge $N \in \mathcal{N}$ (i.e. we assume $|\mathcal{N} \cap \binom{M}{p}| = 1$ for every $M \in \mathcal{M}$). Every (X, \mathcal{N}) obviously has girth $\geq l$. \mathcal{H} is a large set: $|\mathcal{H}| = a^{n^{1+\varepsilon}}$ where $a = \binom{P}{p}$.

However, given a partition π of X by $(k - 1)$ colors we have only at most $(a - 1)^{n^{1+\varepsilon}}$ hypergraphs in \mathcal{H} for which π is a coloring of (X, \mathcal{N}) (with no monochromatic edge). Thus there are at most $(k - 1)^n (a - 1)^{n^{1+\varepsilon}} < a^{n^{1+\varepsilon}} = |\mathcal{H}|$ hypergraphs in \mathcal{H} with chromatic number $< k$. Thus there exists a witness for Theorem 3. ■

In this way the desired high chromatic large girth hypergraph is constructed by randomly replacing edges of \mathcal{M} (i.e. P -tuples) with copies of a fixed hypergraph H_0 . In this proof H_0 is a hypergraph with P vertices containing a single edge (p -tuple).

The above *random placement construction* is very flexible. Satisfaction of the difficult condition on the girth is inherited from $\mathcal{H}(n, p, l, \varepsilon)$ and the chromatic number follows by the above easy counting argument. We have tried to illustrate it by following figure:



Yet another application is given in [72]. This is related to recent work on ergodic properties of topological subgroups of S_ω . The combinatorial part of this development is motivated by the following definition which originated in structural Ramsey theory. We formulate it for graphs (for hypergraphs and, more generally, relational structures the definition and subsequent statements hold with little change).

An ordered graph \vec{G} is a graph $G = (V, E)$ together with a linear ordering \leq of V . We say that a graph $G' = (V', E')$ has the *ordering property* for

\vec{G} if for any ordering \preccurlyeq of V' (i.e. for any ordered graph \vec{G}') there exists an embedding $\varphi : G \rightarrow G'$ which is monotone with respect to \leq and \preccurlyeq .

For example, it is a classical result of graph theory that $\chi(G) \geq k$ if and only if G has the ordering property for the monotone path with k vertices (known as the Gallai-Hasse-Vitaver-Roy theorem).

We have the following:

Theorem 5. *For every graph $G = (V, E)$ there exists a graph G' with the following properties:*

- 1) G' has the ordering property for any ordered graph \vec{G} ;
- 2) if $\text{girth}(G) \geq l$ then $\text{girth}(G') \geq l$.

Proof. Given \vec{G} with p vertices, consider $H(n, p, l, \varepsilon) = (X, \mathcal{M})$ and consider all random placements H of G on edges of \mathcal{M} . Put $\frac{p!}{|\text{Aut}(G)|} = a$ (this is the number of distinct placements of G on a p -element set). The number of all graphs H is thus $a^{|\mathcal{M}|} = a^{n^{1+\varepsilon}}$. However only at most $(a - 1)^{n^{1+\varepsilon}} \cdot n!$ do not have the ordering property for a \vec{G} . Thus there is a witness G' which has the ordering property for all \vec{G} . ■

In particular, for the cycle C_l , $l > 3$, we obtain an undirected graph of girth l which fails to be a cover graph of any partial order.

The existence of high girth non cover graph of posets (proved in [72]) led to the proof that the following recognition problem is NP-complete [78], [14]:

Input: A graph G .

Question: Is G a cover graph of a finite poset?

The question was refined in [86] to lattices and this paper also contains a polynomial algorithm (using Ramanujan graphs) which constructs for given k and l a graph $G_{k,l}$ with $\text{girth}(G_{k,l}) = l$ and $\chi(G_{k,l}) = k$ (see also [22]).

It is clear that every ordering of G' constructed in this way contains many copies of \vec{G} . Recently Angel, Kechris and Lyons [6] isolated in the interesting context of topological dynamics (characterizing structures with unique ergodic measure) the following property of a random placement graph G' : For graphs G, G' we denote by $\text{emb}(G, G')$ the number of all embeddings of G into G' . Similarly $\text{emb}(\vec{G}, \vec{G}')$ denotes the number of monotone embeddings (with respect to orderings of \vec{G}, \vec{G}') of G into G'

Proposition 6. *Let $\varepsilon > 0$ be given. For every 2-connected graph G there exists a graph G' such that for every pair of ordered graphs \vec{G} and \vec{G}'*

$$\left| \frac{\text{emb}(\vec{G}, \vec{G}')}{\text{emb}(G, G')} - \frac{1}{n!} \right| < \varepsilon.$$

The proof follows again by letting G' be the random placement of copies of G and applying Chernoff's inequality.

A generalization of Proposition 6 has been proven recently in [79], which further exploits the random placement construction to ordering property. Let us review it briefly.

In this setting it is convenient to view orderings as permutations. Let $\sigma : [n] \rightarrow [n]$ be a permutation of $[n] = \{1, 2, \dots, n\}$. For $X \subseteq [n]$ let σ_X be the subpermutation of σ induced by the set X (i.e. if $X = \{i_1 < i_2 < \dots < i_k\}$ then $\sigma_X(a) < \sigma_X(b)$ iff $\sigma(i_a) < \sigma(i_b)$).

Let $k \leq n$ (and typically k is much smaller than n) and let $\pi_1, \dots, \pi_{k!}$ be a fixed enumeration of all permutations of $[k]$. The k -statistics of σ is a sequence $s_1^\sigma, \dots, s_{k!}^\sigma$ where $s_i^\sigma = |\{X \in \binom{[n]}{k}; \sigma_X = \pi_i\}| / \binom{n}{k}$.

An ordered graph \vec{G} on $[n]$ may be coded as (G, σ) for a permutation σ of $[n]$. We still call (G, σ) an ordered graph (by permutation σ).

Let (G', σ') be an ordered graph on $[N]$. An embedding (G, σ) into (G', σ') is a monotone injection $f : [n] \rightarrow [N]$ which is embedding of G into G' and which satisfies

$$\sigma(i) < \sigma(j) \text{ if and only if } \sigma'(f(i)) < \sigma'(f(j)).$$

Theorem 7 [79]. *Let G be a 2-connected graph with k vertices. Let $\vec{a} = (a_1, \dots, a_{k!})$ be a stochastic vector. Then for any $\varepsilon > 0$ there exists a graph H with n vertices with the following properties*

- 1) $\text{girth}(G) = \text{girth}(H)$;
- 2) if σ is a permutation of $[n]$ with k -statistics $(s_1, s_2, \dots, s_{k!})$ then

$$\left| \frac{\text{emb}((G, \pi_l), (H, \sigma))}{\text{emb}(G, H)} - b_l \right| < \varepsilon$$

where

$$b_l = \sum \{a_i s_j^\sigma; \pi_i \circ \pi_j = \pi_l\}.$$

It is easy to see that for the uniform probability $\vec{a} = (1/k!, 1/k!, \dots, 1/k!)$ we get Proposition 6.

We also obtain the following “sparsification lemma”, which is perhaps of independent interest

Lemma 8. *For every $l, k \geq 2, \varepsilon > 0$, there exists n and $\mathcal{M} \subseteq \binom{[n]}{k}$ such that*

- 1) $([n], \mathcal{M})$ has no cycles of length $\leq l$;
- 2) for every permutation σ of $[n]$ it holds that $|s_i^\sigma - s_i^\sigma(\mathcal{M})| < \varepsilon$, where $s_i^\sigma(\mathcal{M}) = |\{M \in \mathcal{M}; \sigma_M = \pi_i\}|/|\mathcal{M}|$.

Thus the k -statistics of every permutation σ on $[n]$ are approximated by k -statistics on edges of \mathcal{M} (and yet \mathcal{M} has no short cycles).

The random placement construction has further applications to most coloring problems studied. For example it readily implies one of the main results of [91]. Other applications of random placement construction to coloring of graphs and hypergraphs are contained in [48], [45].

6. OTHER VOICES, OTHER ROOMS

Coloring problems are among the most frequently studied combinatorial problems. One general approach is based on the notion of a homomorphism: Given graphs $G = (V, E)$ and $G' = (V', E')$, a *homomorphism* is any mapping $f : V \rightarrow V'$ which satisfies $\{x, y\} \in E \Rightarrow \{f(x), f(y)\} \in E'$.

It is easy to see that G has a homomorphism to a complete graph K_k if and only if $\chi(G) \leq k$. Motivated by this, a homomorphism $G \rightarrow H$ is also called an H -coloring. Of course, if $G \not\rightarrow K_k$ then also $G \not\rightarrow H$ for every H with $\chi(H) \leq k$. But as homomorphisms compose (i.e. form a category) we can prove stronger and (arguably more elegant) statements. These also indicate that sparsity is not a strong restriction in many coloring problems.

The main results proved in [83] may be formulated as follows. Because of its connection to rigid graphs and homomorphism order it is called sparse incomparability lemma (see e.g. [36]) and it holds for relational systems generally.

Theorem 9. *For every graph H and for all positive integers k and l there exists a graph G with the following properties:*

- 1) $\text{girth}(G) > l$;
- 2) for every graph F with at most k vertices, there exists a homomorphism $g : G \rightarrow F$ if and only if there exists a homomorphism $f : H \rightarrow F$.

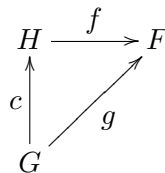
For the statement of the second result of [83] we need the following notion:

A graph F is said to be *pointed for* a graph H (or shortly H -pointed) if any two distinct homomorphisms $H \rightarrow F$ differ in at most 2 vertices. In other words this means that if two homomorphisms $g, g' : H \rightarrow F$ satisfy $g(x) = g'(x)$ for all $x \neq x'$ (for some fixed vertex $x_0 \in V(H)$) then it also holds that $g(x_0) = g'(x_0)$. A graph H is called a *core* if any homomorphism $H \rightarrow H$ is an automorphism. Note that any core graph H is H -pointed and it follows that most graphs H on a large set are H -pointed.

Theorem 10. *For every graph H and for every choice of positive integers k and l there exists a graph G together with a surjective homomorphism $c : G \rightarrow H$ with the following properties:*

- 1) $girth(G) > l$;
- 2) for every graph F with at most k vertices, there exists a homomorphism $g : G \rightarrow F$ if and only if there exists a homomorphism $f : H \rightarrow F$;
- 3) for every H -pointed graph F with at most k vertices and for every homomorphism $g : G \rightarrow F$ there exists a unique homomorphism $f : H \rightarrow F$ such that $g = f \circ c$.

Conditions 2) and 3) may be expressed by the following diagram:



Theorem 10 may look like a technical extension of Theorem 9. However, it has several interesting corollaries from which we obtain structural extension of Erdős' Theorem 1.

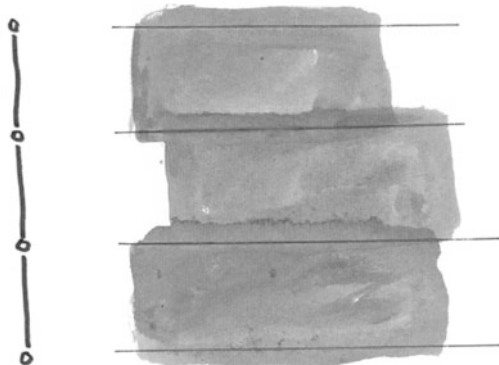
A graph G is *uniquely H -colorable* if there is a surjective homomorphism c from G onto H , and any other homomorphism from G to H is the composition $\sigma \circ c$ of c with an automorphism σ of H . (Note that this implies that H is a core graph.)

The problem of the existence of uniquely k -colorable graphs with large girth has an interesting history: the triangle-free case (i.e. $l = 3$) was settled in [68] and this was improved in [31] to graphs not containing short odd cycles. The general case was solved by Vladimír Müller [64, 65]. Müller's proof is constructive and uses a constructive proof of Theorem 1. A non-constructive proof has been published in [12] and the particular case $H = H'$

of our Theorem 10 (i.e. the existence of uniquely H -colorable graph G with girth $> l$) is proved in [91]. The above Theorem 10 then implies that there is a graph G which is *strongly uniquely H -colorable* in the sense that any homomorphism $G \rightarrow F$ to any small H -pointed graph F is induced by a homomorphism $H \rightarrow F$.

A probabilistic proof of Theorem 10 ([83]) is yet another variant of Erdős method and follows by a now standard pattern [51], [62]. Suppose H has vertices $\{1, 2, \dots, a\}$ and let H have q edges. Let V_1, \dots, V_a be disjoint sets each of (large) size n . Let G_0 be the graph with vertex set $V = V_1 \cup \dots \cup V_a$ and let $\{x, y\} \in E(G_0)$ if and only if $x \in V_i, y \in V_j$ and $\{i, j\} \in E(H)$. Let \mathbb{G} be a random subgraph of G_0 with $qn^{1+\varepsilon}$ edges where $0 < \varepsilon < 1/4l$. This may be viewed as we are replacing each vertex of H by a large cloud (with n vertices) and then taking a sparse random subgraph between clouds corresponding to edges of H . Theorem 9 then follows: If we have a coloring $c : V \rightarrow \{1, \dots, k'\}, k' \leq k$, then for each $i = 1, \dots, a$ let $V'_i \subseteq V$ be the largest monochromatic subset and call the corresponding color $c(i)$. One then observes that if $\{i, h\} \in E(H)$ then $c(i) \neq c(j)$ because between V'_i and V'_j there have to be some edges. Thus if c is a homomorphism $\mathbb{G} \rightarrow H'$, $V(H) = \{1, \dots, a'\}$ then with high probability c induces a homomorphism $H \rightarrow H'$. The proof of part 3) of Theorem 10 needs more care as we have to treat small subsets, see [83].

In the style of this paper we add a schematic figure:



Note that this proof can be derandomized and an explicit (polynomial size) construction can be given (using Ramanujan graphs) [62]. Kun [51] gives a polynomial algorithm to construct set system (and more generally relational systems) satisfying 1), 2) of Theorem 10. This allowed to close the hierarchy of descriptive complexity of classes defined and asked in [29]. In technical terms this amounts to $MMSNP = CSP$ [51].

There is here more than meets the eye. One can prove also a strong extension of Müller's Extension Theorem [64, 65]. To do so we need another (this time categorical) notion:

A *t-projective* graph is a graph G with the property that for every homomorphism (in this setting usually called polymorphism) $f : \underbrace{G \times \cdots \times G}_t \rightarrow G$

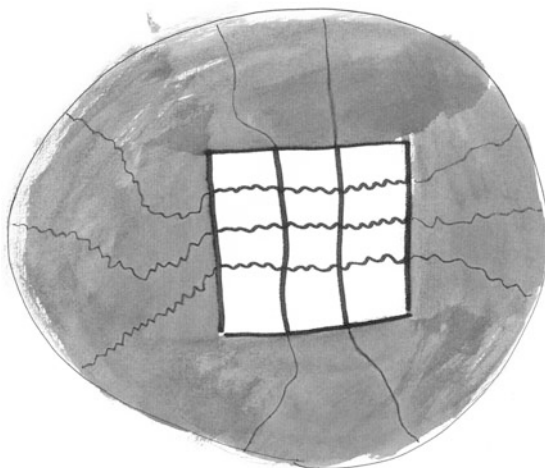
which satisfies $f(x, \dots, x) = x$, there exists i_0 such that $f(x_1, x_2, \dots, x_t) = x_{i_0}$ (i.e. every idempotent homomorphism is a projection). It was proved in [53] that a graph is 2-projective if and only if it is t -projective for every t (this is not true in general for relational structures).

The following result takes us from colorings to arbitrary H -colorings (and holds for general finite relational structures as well):

Corollary 11. *Let H be projective graph with k vertices, and l a positive integer. Let A be a finite set and let f_1, f_2, \dots, f_t be distinct mappings $A \rightarrow V(H)$. Then there exists a graph $G = (V, E)$ such that the followings hold:*

- 1) A is a subset of V ;
- 2) for every $i = 1, 2, \dots, t$ there exists unique homomorphism $g_i : G \rightarrow H$ such that g_i restricted to the set A coincides with the mapping f_i ;
- 3) for every homomorphism $f : G \rightarrow H$ there exists $i, 1 \leq i \leq t$ and an automorphism h of H such that $h \circ f_i = f$;
- 4) G has girth $> l$.

Müller's Extension Theorem corresponds to k -colorings (i.e. $H = K_k$) which uniquely extend a given set partition and is depicted in the following figure:



This is not just a generalization, this is as far as we can go:

Corollary 12. *For a core graph H , the following statements are equivalent:*

- I. *Corollary 11 holds.*
- II. *The graph H is projective.*

These results and corresponding notions are not only interesting as an ultimate strengthening of Erdős' Theorem 1 in a more structural setting. In fact the above results hold for relational structures (or finite relational models). This is important for the complexity of algorithms, particularly in the context of Constraint Satisfaction Problems – CSP:

For a given a relational structure \mathbf{H} , $CSP(\mathbf{H})$ denotes the following decision problem:

Input: A structure \mathbf{A}

Question: Does there exists a homomorphism $\mathbf{A} \rightarrow \mathbf{H}$.

It is conjectured [29] that this problem falls into just two classes:

NP-Complete problems and polynomially solvable problems. This Dichotomy Conjecture [29] was investigated in the context of (universal) algebra [40], [15], [7], combinatorics and graph theory [36], threshold phenomena and random walks [52], for survey of this development see e.g. [36]. The dichotomy conjecture has then a refined form which conjectures an actual form of the dichotomy, see [37] Theorem 3.4. One of these formulations (using term “block projectivity”) was isolated in [82]. Let us remark that recently the analogous question of dichotomy for counting homomorphisms (and CSP) was solved in the full generality in a major paper [16].

Theorem 11 plays an important role in this reduction (via so called fibre construction). It further follows from this result that the conjectured Dichotomy is very robust: it does not change if we restrict to objects with girth $\geq l$ and to structures with degree of its vertices bounded by $D(\mathbf{H})$. For this we need an effective version of Corollary 11 which is provided by [62, 51]. As a particular case the following problem is NP-complete for any non-bipartite graph H (and any fixed l):

Input: Graph G with girth $\geq l$.

Question: Does there exists an H -coloring of G ?

So after all the work the large girth restriction is not much of an obstacle.

7. LIMITATIONS, PERSPECTIVES AND PROBLEMS

We can interpret negatively many of the above results: Despite the apparent difficulty high girth graphs with large chromatic number exist and their complexity and special properties seem not to be influenced by this (severe) restriction.

Here we add a few structural results which indicate opposite. We start with two infinite limitations.

7.1. No universal C_4 -free graph

Theorem 13 [33]. *There is no countable graph of girth > 4 which contains every countable graph of girth > 4 as a subgraph.*

In other words, there is no countable universal graph for the class of all graphs of girth > 4 . The same holds for the class of graphs with girth $> l \geq 4$. On the other hand, such a universal graph exists for the class of all triangle-free graphs. (The same is true, more generally, for the class of all graphs not containing short odd cycles [18].) In fact one can “forbid” homomorphisms from any finite set of graphs, see [17]. But this does not surprise an interested reader: odd cycles are easier in the whole paper.

7.2. No (transfinite) unbounded χ

Theorem 14 [25]. *Every graph G with chromatic number $> \omega$ (i.e. of uncountable chromatic number) contains a complete bipartite subgraph $K_{\omega,n}$ for arbitrary finite n and thus in particular the quadrangle $K_{2,2}$.*

Thus the graphs $G_{k,l}$ are strictly finite objects which do not have, in full generality, an analogy in the infinite.

7.3. Erdős-Hajnal

Conjecture 7.1 [24]. *For every k, l there exists $f(k, l)$ such that any graph G with $\chi(G) \geq f(k, l)$ contains a subgraph G' of girth $> l$ and $\chi(G') \geq k$.*

This beautiful and (at least at first glance) plausible conjecture is in fact a very hard problem. The only known non-trivial case is that $f(k, 4)$ exists (Rödl [84]). (Note that Theorem 1 is equivalent to saying that Conjecture 7.1 holds for complete graphs.)

This conjecture appeared recently in the context of (homomorphism) restricted dualities, cf. [71, Chapter 11].

7.4. Victor Neumann-Lara Conjecture

Conjecture 7.2. *There exists a function $g : \mathbb{N} \rightarrow \mathbb{N}$ with the following property: Let $G = (V, E)$ be an undirected graph with $\chi(G) \geq g(k)$. Then there exists an orientation $\vec{G} = (V, \vec{E})$ of G such that in every k -coloring of V \vec{G} contains a monochromatic directed cycle.*

There are many (perhaps too many) variants of chromatic number.

The minimal number of colors which suffice for coloring of vertices of digraphs so that no directed cycle is monochromatic is called (*directed*) *acyclic chromatic number* (thus the defining property is that the color classes induce acyclic subgraphs). Directed acyclic chromatic number was investigated and the analogy of Theorem 1 for directed acyclic coloring immediately follows from the random placement method (for a different proof see [9]). The existence of uniquely acyclic colorable was proved recently in [34].

7.5. The Pentagon problem

Conjecture 7.3. *There exists an integer l with the following property: If G is a subcubic graph (i.e. every vertex has degree ≤ 3) with girth $\geq l$ then $G \rightarrow C_5$.*

Brook’s theorem implies that every subcubic graph not containing K_4 satisfies $G \rightarrow C_3$. On the other hand this problem has a negative solution for C_{11} [47], C_9 [90], and finally C_7 [35].

It is not even known whether high girth subcubic graphs have circular chromatic number < 3 .

7.6. No Ramsey classes with girth > 3

A class \mathcal{C} of graphs (or, more generally, structures) is said to be *Ramsey* if the following holds ([80], [81], [70]): For every $\mathbf{A}, \mathbf{B} \in \mathcal{C}$ and positive integer k there exists $\mathbf{C} \in \mathcal{C}$ such that $\mathbf{C} \rightarrow (\mathbf{B})_k^{\mathbf{A}}$ where (Erdős-Rado) partition arrow has the following meaning:

For every partition $\binom{\mathbf{C}}{\mathbf{A}} = \mathcal{A}_1 \cup \dots \cup \mathcal{A}_k$ there exists $\mathbf{B}' \in \binom{\mathbf{C}}{\mathbf{B}}$ such that $\binom{\mathbf{B}'}{\mathbf{A}} \subseteq \mathcal{A}_i$ for some $i \in \{1, \dots, k\}$. Here $\binom{\mathbf{C}}{\mathbf{A}}$ denotes the set of all substructures of \mathbf{C} which are isomorphic to \mathbf{A} .

Ramsey’s theorem claims that the class of complete graphs is a Ramsey class.

Theorem 15 [69]. *The class \mathcal{C}_l of all ordered graphs with girth $\geq l > 4$ fails to be a Ramsey class.*

This follows from [69] where the connection between Ramsey classes and ultrahomogeneous structures is isolated and as a result of this all Ramsey classes of graphs are determined (or better: it is shown that we know them all). This line of research was studied later intensively, see e.g. the important paper [42], and connections to extreme amenability of subgroups of S_ω were established.

7.7. Edge Ramsey for large girth

Problem 16. Does the class \mathcal{C}_l have the edge-Ramsey property?

Explicitly: Is it true that for every $G \in \mathcal{C}_l$ there exists $H = (V, E) \in \mathcal{C}_l$ such that for every partition $E_1 \cup E_2$ of E there exists a subgraph G' of H , G' isomorphic to G such that $E(G')$ is a subset of either E_1 or E_2 (i.e. $H \rightarrow (G)_2^{K_2}$ in the above notation). This is known to be true for $l \leq 6$ [65]. This problem (together with Pisier type problems [26]) is one of the few that remained open in structural Ramsey theory.

7.8. Persistence of Old Motivations

The problems addressed in this paper are active problems attacked (and sometimes) solved by many. As an example of recent striking result let us mention the work of T. Bohman and P. Keevash and, independently of G. F. Pontiveros, S. Griffiths and R. Morris on asymptotics of Ramsey numbers $r(k, 3)$. But other problems remain. Particularly, the basic challenge in this area of complex large girth graphs is to find new constructions. The old questions remain. The recent advances of theoretical computer science put these problems in a new context and make these questions very actual.

Acknowledgment: I thank to Martin Bálek and Andrew Goodall for the help when preparing this article.

REFERENCES

- [1] M. Ajtai, J. Komlós, E. Szemerédi, *A note on Ramsey numbers*, J. Combi. Th. Series A, 29 (1980), 354–360.
- [2] M. Ajtai, J. Komlós, E. Szemerédi, *An $O(n \log n)$ sorting network*, Proceedings of the 15th Annual ACM Symposium on Theory of Computing, 1983, 1–9.

- [3] N. Alon, *Discrete Mathematics: methods and challenges*, Proc. of the International Congress of Mathematicians (ICM), Beijing 2002, China, Higher Education Press (2003), 119–135.
- [4] N. Alon, V. Rödl, *Sharp bounds for some multicolor Ramsey numbers*, *Combinatorica* 25 (2005), 125–141.
- [5] N. Alon, J. H. Spencer, *The probabilistic method* Second ed. Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, 2000.
- [6] O. Angel, A. Kechris, R. Lyons, *Random Orderings and Unique Ergodicity of Automorphism Groups*, arXiv: 1208.2389 (2012).
- [7] L. Barto, M. Kozik, *New conditions for Taylor varieties and CSP*, In: Proceedings of LICS'10, IEEE, 2010, 100–109.
- [8] S. Baum, M. Stiebitz, *Coloring Of Graphs Without Short Odd Paths Between Vertices Of The Same Color Class*, (preprint, TU Ilmenau).
- [9] D. Bokal, G. Fijavž, M. Juvan, P. M. Kayll, B. Mohar, *The circular chromatic number of a digraph*, *J. Graph Theory* 46(2004), no. 3, 227–240.
- [10] B. Bollobás, *Random Graphs*, Academic Press, 1985.
- [11] B. Bollobás, *Modern Graph Theory*, Springer-Verlag, 1998.
- [12] B. Bollobás and N. Sauer, *Uniquely colorable graphs with large girth*, *Can. J. Math.* 28(1976), 1340–1344.
- [13] J. A. Bondy, U. S. R. Murty, *Graph theory*, Graduate Texts in Mathematics, 244, Springer, 2008.
- [14] G. Brightwell, *On the complexity of diagram testing*, *Order* 10,4 (1983), 297–303.
- [15] A. A. Bulatov, P. Jeavons, A. A. Krokhin, *Classifying the Complexity of Constraints Using Finite Algebras*, *SIAM J. Comput.* 34(3), 2005, 720–742.
- [16] J.-Y. Cai, X. Chen, P. Lu, *Graph Homomorphisms with Complex Values: A Dichotomy Theorem*, arXiv: 0903.4728v2.
- [17] G. Cherlin, S. Shelah, N. Shi, *Universal graphs with forbidden subgraphs and algebraic closure*, *Adv. in Applied Math.* 22 (1999), 454–491.
- [18] G. Cherlin, N. Shi, *Graphs omitting a finite set of cycles*, *J. of Graph Th.* 21 (1997), 351–355.
- [19] B. Codenotti, P. Pudlák, J. Resta, *Some structural properties of low rank matrices related to computational complexity*, *Theoretical Computer Sci.* 235 (2000), 89–107.
- [20] B. Descartes, *A three colour problem*, *Eureka* 21, 1947.
- [21] R. Diestel, *Graph theory*, Graduate Texts in Mathematics, 173, Springer, Heidelberg, 2010.
- [22] T. Emden-Weinert, S. Hongardy, B. Kreutzer, *Uniquely colorable graphs and hardness of colouring of graphs of large girth*, *Comb. Prob. Comp.* 7,4 (1998), 375–386.
- [23] P. Erdős, *Graph theory and probability*, *Canad. J. Math.* 11 (1959), 34–38.
- [24] P. Erdős, *Problems and results in combinatorial analysis and graph theory*, In: Proof Techniques in Graph Theory, Academic Press, 1969, 27–35.
- [25] P. Erdős, A. Hajnal, *On chromatic number of graphs and set-systems*, *Acta Math. Acad. Sci. Hungar.* 17 (1966), 61–99.

- [26] P. Erdős, J. Nešetřil, V. Rödl *On Pisier type problems and results (combinatorial applications to number theory)*, In: Mathematics of Ramsey Theory, Springer 1990, 214–231.
- [27] P. Erdős, C. A. Rogers, *The construction of certain graphs*, Canad. J. Math. 14 (1962), 702–707.
- [28] P. Erdős, E. Specker, *On a theorem in the theory of relations and a solution of a problem of Knaster*, Colloq. Math. 8 (1961), 19–21.
- [29] T. Feder, M. Y. Vardi, *The computational structure of monotone monadic SNP and constraint satisfaction: A study through Datalog and group theory*, SIAM J. Comput. 28, 1 (1999), 57–104.
- [30] R. Graham and B. Rothschild and J. Spencer, *Ramsey Theory*, Wiley, New York, 1990.
- [31] D. Greenwell and L. Lovász, *Applications of product coloring*, Acta Math. Acad. Sci. Hungar. 25(1974), 335–340.
- [32] A. Gyárfás, T. Jensen, M. Stiebitz, *On graphs with strongly independent colour-classes*, JGT 46 (2004), 1–14.
- [33] A. Hajnal, J. Pach, *Monochromatic paths in infinite graphs*, In: Finite and Infinite Sets, Coll. Math. Soc. J. Bolyai, Eger, 1981, 359–369.
- [34] A. Harutyunyan, P. M. Kayll, B. Mohar, L. Rafferty, *Uniquely D -colourable Digraphs with Large Girth*, Canad. J. Math. Vol. 64 (6), 2012, 1310–1328.
- [35] H. Hatami, *Random cubic graphs are not homomorphic to the cycle of size 7*, J. Comb. Th. B 93 (2005), 319–325.
- [36] P. Hell, J. Nešetřil, *Graphs and homomorphisms*, Oxford University Press, 2004.
- [37] P. Hell, J. Nešetřil, *Colouring, Constraint Satisfaction, and Complexity*, Comp. Sci. Review 2,3 (2008) 134–164.
- [38] S. Hoory, N. Linial, A. Wigderson, *Expander graphs and their applications*, Bulletin (New series) of the American Mathematical Society 43 (4), 2006, 439–561.
- [39] S. Janson, T. Luczak, A. Rucinski, *Random Graphs*, Wiley, 2000.
- [40] P. Jeavons, *On the algebraic structure of combinatorial problems*, Theor. Comp. Sci 200 (1998), 185–204.
- [41] J. Kahn, *Recent results on some not-so-recent hypergraph matching and covering problems*, Proc. 1st Int'l Conference on Extremal Problems for Finite Sets, Visegrad, 1991.
- [42] A. Kechris, Pestov, S. Todorcevic, *Fraïssé limits, Ramsey theory and topological dynamics of automorphism groups*, Geometrical and Functional Analysis 15(1), 2005, 106–189.
- [43] J. Kelly and L. Kelly, *Path and circuits in critical graphs*, Amer. J. Math. 76:786–792, 1954.
- [44] J. H. Kim, *The Ramsey Number $R(3, t)$ has order of magnitude $t^2/\log t$* , Random Structures and Algorithms 7 (1995), 173–207.
- [45] A. V. Kostochka, D. Mubayi, J. Verstraete, *Hypergraph Ramsey Numbers: Triangles versus Cliques*, submitted.

- [46] A. V. Kostochka J. Nešetřil, *Properties of Descartes' construction of triangle-free graphs with high chromatic number*, *Combin. Probab. Comput.* 8(1999), no. 5, 467–472.
- [47] A. Kostochka, J. Nešetřil, P. Smolřková, *Coloring and homomorphism of degenerate and bounded degree graphs*, *Discrete Math.* 233, 1–3 (2001), 257–276.
- [48] A. V. Kostochka, V. Rödl, *Constructions of Sparse Uniform Hypergraphs with High Chromatic Number*, *Random tructures and Algorithms*, 2009, 46–56.
- [49] I. Kříž, *A hypergraph free construction of highly chromatic graphs without short cycles*, *Combinatorica* 9 (1989), 227–229.
- [50] I. Kříž, J. Nešetřil, *Chromatic number of the Hasse diagrams, eyebrows and dimension*, *Order* 8 /1991/, 41–48.
- [51] G. Kun, *Constraints, MMSN and expander relational structures*, arXiv: 0706.1701 (2007).
- [52] G. Kun, M. Szegedy, *A new line of attack on the dichotomy conjecture*, *STOC 2009*, 725–734. See also *Electronic Coll. on Comp. Compl. (ECCC)* 16:59 (2009), 44p.
- [53] B. Larose, C. Tardif, *Graph coloring problems*, Wiley, 1995.
- [54] J. Lenz, D. Mubayi, *The poset of hypergraph quasirandomness*, *The Poset of Hypergraph Quasirandomness*, arXiv:1208.5978 [math.CO] 29 Aug 2012.
- [55] L. Lovász, *On chromatic number of finite set-systems*, *Acta Math. Acad. Sci. Hungar.* 19(1968), 59–67.
- [56] L. Lovász, *Kneser's conjecture, chromatic number, and homotopy*, *J. Combin. Theory. Ser. A* 25 (1978), 319–324.
- [57] L. Lovász, *Combinatorial Problems and Exercises*, Akad. Kiad, Budapest, 1979.
- [58] A. Lubotzky, R. Phillips, P. Sarnak, *Ramanujan graphs*, *Combinatorica*, 8(3), 1988, 261–277.
- [59] M. Mareš, *The saga of minimum spanning trees*, *Comp. Sci. Review* 2 (2008), 165–221.
- [60] G. A. Margulis, *Explicit constructions of graphs without short cycles and low density codes*, *Problemy Pereači Informacii*, 9(4), 1973, 71–80.
- [61] J. Matoušek, *Using the Borsuk-Ulam theorem*, *Lectures on topological methods in combinatorics and geometry*, Springer, 2003.
- [62] J. Matoušek, J. Nešetřil, *Constructions of sparse graphs with given homomorphisms*, (manuscript).
- [63] M. Molloy, B. Reed, *Graph colouring and the probabilistic method*, *Algorithms and Combinatorics*, 23, Springer, 2002.
- [64] V. Müller, *On colorable critical and uniquely colorable critical graphs*, in: *Recent Advances in Graph Theory* (ed. M. Hiedler), Academia, Prague, 1975.
- [65] V. Müller, *On coloring of graphs without short cycles*, *Discrete Math.*, 26(1979), 165–179.
- [66] J. Mycielski, *Sur le coloriage des graphes*, *Colloq. Math.* 3, 161–162, 1955.

- [67] J. Nešetřil, *K-chromatic graphs without cycles of length ≤ 7* , (in Russian), Comment. Math. Univ. Carol., 7, 3(1966), 373–376.
- [68] J. Nešetřil, *On uniquely colorable graphs without short cycles*, Časopis Pěst. Mat. 98(1973), 122–125.
- [69] J. Nešetřil, *For graphs there are only four types of hereditary Ramsey classes*, J. Combin. Theory Ser. B 46, (1989), no. 2, 127–132.
- [70] J. Nešetřil, *Ramsey Theory*, In: Handbook of Combinatorics (ed. R. L. Graham, M. Grötschel, L. Lovász), North-Holland, 1995, 1331–1403.
- [71] J. Nešetřil, P. Ossona de Mendez, *Sparsity – Graph, Structures, and Algorithms*, Springer, 2012.
- [72] J. Nešetřil, V. Rödl, *On a probabilistic graph-theoretic Method*, Proc. Amer. Math. Soc. 72 (1978), 417–421.
- [73] J. Nešetřil, V. Rödl, *A short proof of the existence of restricted Ramsey graphs by means of a partite construction*, Combinatorica 1,2 (1982), 199–202.
- [74] J. Nešetřil, V. Rödl, *Sparse Ramsey graphs*, Combinatorica 4, 1 (1984), 71–78.
- [75] J. Nešetřil, V. Rödl, *Combinatorial partitions of finite posets and lattices – Ramsey lattices*, Algebra Univ. 19 (1984), 106–119.
- [76] J. Nešetřil, V. Rödl, *Complexity of diagrams*, Order 3 (1987), 321–330.
- [77] J. Nešetřil, V. Rödl, *Chromatically optimal rigid graphs*, J. Combin. Th.(B), 46(1989), 133–141.
- [78] J. Nešetřil, V. Rödl, *More on complexity of diagrams*, Comm. Math. Univ. Carol. 36,2 (1995), 269–278.
- [79] J. Nešetřil, V. Rödl, *Statistics of Orderings*, (to appear).
- [80] J. Nešetřil, V. Rödl, *Partitions of Finite Relational and Set Systems*, J. Comb. Th. A 22,3 (1977), 289–312.
- [81] J. Nešetřil, V. Rödl, *Partition (Ramsey) Theory – a survey*, In: Coll. Math. Soc. János Bolyai, 18. Combinatorics, Keszthely 1976, North Holland, 1978, 759–792.
- [82] J. Nešetřil, M. Siggers, L. Zadori, *A Combinatorial Constraint Satisfaction Problem Dichotomy Classification Conjecture*, European J. Comb. 31 (1), 2010, 280–296.
- [83] J. Nešetřil, X. Zhu, *On sparse graphs with given colorings and homomorphisms*, J. Combin. Theory Ser. B 90(2004), no. 1, 161–172.
- [84] V. Rödl, *On the chromatic number of subgraphs of a given graph*, Proc. Amer. Math. Soc. 64 (1977), 370–371.
- [85] V. Rödl, *On a packing and covering problem*, Europ. J. Combinatorics 5 (1985), 69–78.
- [86] V. Rödl, L. Thoma, *The complexity of cover graph recognition for some vertices of finite lattices*, Order 12,4 (1995), 351–374.
- [87] G. Simonyi and G. Tardos, *Local chromatic number, Ky Fan’s theorem, and circular colorings*, Combinatorica 26 (2006), 587–620.
- [88] J. Spencer, *Ten Lectures on the Probabilistic Method*, Society for Industrial and Applied Mathematics, 1987.

- [89] W. T. Trotter, R. Wang, *Incidence Posets and Cover Graphs*, to appear in *Order*.
- [90] I. M. Wanless, N. C. Wormald, *Regular graphs with no homomorphisms onto cycles*, *J. Comb. Th. B* 82 (2001), 155–160.
- [91] X. Zhu, *Uniquely H -colorable graphs with large girth*, *J. Graph Theory*, **23** (1996), 33–41.
- [92] A. A. Zykov, *On some properties of linear complexes*, *Mat. Sbornik* 24:313–319, 1949. (In Russian).

Jaroslav Nešetřil

Computer Science Institute of Charles

University,

and

Institute for Theoretical Computer Science

(ITI),

Charles University,

Malostranské nám.25,

11800 Praha 1,

Czech Republic

e-mail: `nesetril@iuuk.mff.cuni.cz`

SMALL BALL PROBABILITY, INVERSE THEOREMS, AND APPLICATIONS

HOI H. NGUYEN* and VAN H. VU†

Let ξ be a real random variable with mean zero and variance one and $A = \{a_1, \dots, a_n\}$ be a multi-set in \mathbf{R}^d . The random sum

$$S_A := a_1\xi_1 + \dots + a_n\xi_n$$

where ξ_i are iid copies of ξ is of fundamental importance in probability and its applications.

We discuss the *small ball* problem, the aim of which is to estimate the maximum probability that S_A belongs to a ball with given small radius, following the discovery made by Littlewood–Offord and Erdős almost 70 years ago. We will mainly focus on recent developments that characterize the structure of those sets A where the small ball probability is relatively large. Applications of these results include full solutions or significant progresses of many open problems in different areas.

1. LITTLEWOOD–OFFORD AND ERDŐS ESTIMATES

Let ξ be a real random variable with mean zero and variance one and $A = \{a_1, \dots, a_n\}$ be a multi-set in \mathbf{R} (here $n \rightarrow \infty$). The random sum

$$S_A := a_1\xi_1 + \dots + a_n\xi_n$$

where ξ_i are iid copies of ξ plays an essential role in probability. The Central Limit Theorem, arguably the most important theorem in the field, asserts that if the a_i 's are the same, then

$$\frac{S_A}{\sqrt{\sum_{i=1}^n |a_i|^2}} \longrightarrow \mathbf{N}(0, 1).$$

*The first author is supported by research grant DMS-1256802.

†The second author is supported by research grants DMS-0901216 and AFOSAR-FA-9550-12-1-0083.

Furthermore, Berry-Esséen theorem shows that if ξ has bounded third moment, then the rate of convergence is $O(n^{-1/2})$. This, in particular, implies that for any small open interval I

$$\mathbf{P}(S_A \in I) = O(|I|/n^{1/2}).$$

The assumption that the a_i 's are the same are, of course, not essential. Typically, it suffices to assume that none of the a_i 's is dominating; see [13] for more discussion.

The probability $\mathbf{P}(S_A \in I)$ (and its high dimensional generalization) will be referred to as *small ball* probability throughout the paper. In 1943, Littlewood and Offord, in connection with their studies of random polynomials [33], raised the problem of estimating the small ball probability for *arbitrary* coefficients a_i . Notice that when we do not assume anything about the a_i 's, even the Central Limit Theorem may fail, so Berry-Esséen type bounds no longer apply. Quite remarkably, Littlewood and Offord managed to show

Theorem 1.1. *If ξ is Bernoulli (taking values ± 1 with probability $1/2$) and a_i have absolute value at least 1, then for any open interval I of length 2,*

$$\mathbf{P}(S_A \in I) = O\left(\frac{\log n}{n^{1/2}}\right).$$

Shortly after Littlewood–Offord result, Erdős [10] gave a beautiful combinatorial proof of the following refinement, which turned out to be sharp.

Theorem 1.2. *Under the assumption of Theorem 1.1*

$$(1) \quad \mathbf{P}(S_A \in I) \leq \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n} = O\left(\frac{1}{n^{1/2}}\right).$$

Proof (of Theorem 1.2). Erdős' proof made an ingenious use of Sperner's lemma, which asserts that if \mathcal{F} is an anti-chain on a set of n elements, then \mathcal{F} has at most $\binom{n}{\lfloor n/2 \rfloor}$ elements (an anti-chain is a family of subsets none of which contains the other). Let x be a fixed number. By reversing the sign of a_i if necessary, one can assume that $a_i \geq 1$ for all i . Now let \mathcal{F} be the set of all subsets X of $[n] := \{1, 2, \dots, n\}$ such that

$$\sum_{i \in X} a_i - \sum_{j \in \bar{X}} a_j \in (x - 1, x + 1).$$

One can easily verify that \mathcal{F} is an anti-chain. Hence, by Sperner's lemma,

$$|\mathcal{F}| \leq \frac{\binom{n}{n/2}}{2^n},$$

completing the proof. ■

The problem was also studied in probability by Kolmogorov, Rogozin, and others; we refer the reader to [30, 31] and [43]. Erdős' result is popular in the combinatorics community and has become the starting point for a whole theory that we now start to discuss.

Notation. We use the asymptotic notation such as O, o, Θ under the assumption that $n \rightarrow \infty$; $O_\alpha(1)$ means the constant in big O depends on α . All logarithms have natural base, if not specified otherwise.

2. HIGH DIMENSIONAL EXTENSIONS

Let ξ be a real random variable and $A = \{a_1, \dots, a_n\}$ a multi-set in \mathbf{R}^d , where d is fixed. For a given radius $R > 0$, we define

$$\rho_{d,R,\xi}(A) := \sup_{x \in \mathbf{R}^d} \mathbf{P}(a_1 \xi_1 + \dots + a_n \xi_n \in \mathbf{B}(x, R)),$$

where ξ_i are iid copies of ξ , and $\mathbf{B}(x, R)$ denotes the open disk of radius R centered at x in \mathbf{R}^d . Furthermore, let

$$p(d, R, \xi, n) := \sup_A \rho_{d,R,\xi}(A)$$

where A runs over all multi-sets of size n in \mathbf{R}^d consisting of vectors with norm at least 1. Erdős' theorem can be reformulated as

$$p(1, 1, \text{Ber}, n) = \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n} = O(n^{-1/2}).$$

In the case $d = 1$, Erdős obtained the optimal bound for any fixed R . In what follows we define $s := \lfloor R \rfloor + 1$.

Theorem 2.1. *Let $S(n, m)$ denote the sum of the largest m binomial coefficients $\binom{n}{i}$, $0 \leq i \leq n$. Then*

$$(2) \quad p(1, R, \text{Ber}, n) = 2^{-n} S(n, s).$$

The case $d \geq 2$ is much more complicated and has been studied by many researchers. In particular, Katona [24] and Kleitman [25] showed that $p(2, 1, Ber, n) = 2^{-n} \binom{n}{\lfloor n/2 \rfloor}$. This result was extended by Kleitman [26] to arbitrary dimension d ,

$$(3) \quad p(d, 1, Ber, n) = \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n}.$$

The estimate for general radius R is much harder. In [27], Kleitman showed that $2^n p(2, R, Ber, n)$ is bounded from above by the sum of the $2 \lfloor R/\sqrt{2} \rfloor$ largest binomial coefficients in n . For general d , Griggs [19] proved that

$$p(d, R, Ber, n) \leq 2^{2^{d-1}-2} \lceil R\sqrt{d} \rceil \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n}.$$

This result was then improved by Sali [48, 49] to

$$p(d, R, Ber, n) \leq 2^d \lceil R\sqrt{d} \rceil \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n}.$$

A major improvement is due to Frankl and Füredi [14], who proved

Theorem 2.2. *For any fixed d and R*

$$(4) \quad p(d, R, Ber, n) = (1 + o(1))2^{-n} S(n, s).$$

This result is asymptotically sharp. In view of (2) and (3), it is natural to ask if the exact estimate

$$(5) \quad p(d, R, Ber, n) = 2^{-n} S(n, s),$$

holds for all fixed dimension d . However, this has turned out to be false. The authors of [26, 14] observed that (5) fails if $s \geq 2$ and

$$(6) \quad R > \sqrt{(s-1)^2 + 1}.$$

Example 2.3. Take $v_1 = \dots = v_{n-1} = \mathbf{e}_1$ and $v_n = \mathbf{e}_2$, where $\mathbf{e}_1, \mathbf{e}_2$ are two orthogonal unit vectors. For this system, let B be the ball of radius R centered at $v = (v_1 + \dots + v_n)/2$. Assume that n has the same parity with s , then by definition we have

$$\mathbf{P}(S_V \in \mathbf{B}(v, R)) = 2 \sum_{(n-s)/2 \leq i \leq (n+s)/2} \binom{n-1}{i} / 2^n > 2^{-n} S(n, s).$$

Frankl and Füredi raised the following problem.

Conjecture 2.4 [14, Conjecture 5.2]. *Let R, d be fixed. If $s - 1 \leq R < \sqrt{(s - 1)^2 + 1}$ and n is sufficiently large, then*

$$p(d, R, Ber, n) = 2^{-n}S(n, s).$$

The conjecture has been confirmed for $s = 1$ by Kleitman (see (3)) and for $s = 2, 3$ by Frankl and Füredi [14] (see [14, Theorem 1.2]). Furthermore, Frankl and Füredi showed that (5) holds under a stronger assumption that $s - 1 \leq R \leq (s - 1) + \frac{1}{10s^2}$. A few years ago, Tao and the second author proved Conjecture 2.4 for $s \geq 3$. This, combined with the above mentioned earlier results, established the conjecture in full generality [66].

Theorem 2.5. *Let R, d be fixed. Then there exists a positive number $n_0 = n_0(R, d)$ such that the following holds for all $n \geq n_0$ and $s - 1 \leq R < \sqrt{(s - 1)^2 + 1}$*

$$p(d, R, Ber, n) = 2^{-n}S(n, s).$$

We will present a short proof of Theorems 2.2 and 2.5 in Section 17.

3. REFINEMENTS BY RESTRICTIONS ON A

A totally different direction of research started with the observation that the upper bound in (1) improves significantly if we make some extra assumption on the additive structure of A . In this section, it is more natural to present the results in discrete form. In the discrete setting, one considers the probability that S_A takes a single value (for instance, $\mathbf{P}(S_A = 0)$).

Erdős’s result in the first section implies

Theorem 3.1. *Let a_i be non-zero real numbers, then*

$$\sup_{x \in \mathbf{R}} \mathbf{P}(S_A = x) \leq \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n} = O(n^{-1/2}).$$

Erdős and Moser [11] showed that under the condition that the a_i are different, the bound improved significantly.

Theorem 3.2. *Let a_i be distinct real numbers, then*

$$\sup_{x \in \mathbf{R}} \mathbf{P}(S_A = x) = O(n^{-3/2} \log n).$$

They conjectured that the $\log n$ term is not necessary. Sárközy and Szemerédi’s [50] confirmed this conjecture

Theorem 3.3. *Let a_i be distinct real numbers, then*

$$\rho_A := \sup_{x \in \mathbf{R}} \mathbf{P}(S_A = x) = O(n^{-3/2}).$$

In [54], Stanley found a different (algebraic) proof for a more precise result, using the hard-Lefschetz theorem from algebraic geometry.

Theorem 3.4 (Stanley's theorem). *Let n be odd and $A_0 := \{ -\frac{n-1}{2}, \dots, \frac{n-1}{2} \}$. Let A be any set of n distinct real numbers, then*

$$\rho(A) := \sup_{x \in \mathbf{R}} \mathbf{P}(S_A = x) \leq \sup_{x \in \mathbf{R}} \mathbf{P}(S_{A_0} = x).$$

A similar result holds for the case n is even, see [54]. Later, Proctor [41] found a simpler proof for Stanley's theorem. His proof is also algebraic, using tools from Lie algebra. It is interesting to see whether algebraic approaches can be used to obtain *continuous* results. (For the continuous version of Theorem 3.3, see Section 6.)

A hierarchy of bounds. We have seen that the Erdős' bound of $O(n^{-1/2})$ is sharp, if we allow the a_i to be the same. If we forbid this, then the next bound is $O(n^{-3/2})$, which can be attained if the a_i form an arithmetic progression. Naturally, one would ask what happen if we forbid the a_i to form an arithmetic progression and so forth. Halász' result, discussed in Section 6, gives a satisfying answer to this question.

Remark 3.5. To conclude this section, let us mention that while discrete theorems such as Theorem 3.4 are formalized for real numbers, it holds for any infinite abelian groups, thanks to a general trick called Freiman isomorphism (see [67] and also Appendix A). In particular, this trick allows us to assume that the a_i 's are integers in the proofs. Freiman isomorphism, however, is not always applicable in continuous settings.

4. LITTLEWOOD–OFFORD TYPE BOUNDS FOR HIGHER DEGREE POLYNOMIALS

For simplicity, we present all results in this section in discrete form. The extension to continuous setting is rather straightforward, and thus omitted.

One can view the sum $S = a_1\xi_1 + \dots + a_n\xi_n$ as a linear function of the random variables ξ_1, \dots, ξ_n . It is natural to study general polynomials of

higher degree k . Let us first consider the case $k = 2$. Following [8], we refer to it as the Quadratic Littlewood–Offord problem.

Let ξ_i be iid Bernoulli random variables, let $A = (a_{ij})$ be an $n \times n$ symmetric matrix of real entries. We define the *quadratic concentration probability* of A by

$$\rho_q(A) := \sup_{a \in \mathbf{R}} \mathbf{P} \left(\sum_{i,j} a_{ij} \xi_i \xi_j = a \right).$$

Similar to the problem considered by Erdős and Littlewood–Offord, we may ask what upper bound one can prove for $\rho_q(A)$ provided that the entries a_{ij} are non-zero? This question was first addressed by Costello, Tao and the second author in [8], motivated by their study of Weiss’ problem concerning the singularity of a random symmetric matrix (see Section 5).

Theorem 4.1. *Suppose that $a_{ij} \neq 0$ for all $1 \leq i, j \leq n$. Then*

$$\rho_q(A) = O(n^{-1/8}).$$

The key to the proof of Theorem 4.1 is a decoupling lemma, which can be proved using Cauchy-Schwarz inequality. The reader may consider this lemma an exercise, or consult [8] for details.

Lemma 4.2 (Decoupling lemma). *Let Y and Z be independent random variables and $E = E(Y, Z)$ be an event depending on Y and Z . Then*

$$\mathbf{P}(E(Y, Z)) \leq \mathbf{P}(E(Y, Z) \wedge E(Y', Z) \wedge E(Y, Z') \wedge E(Y', Z'))^{1/4}$$

where Y' and Z' are independent copies of Y and Z , respectively. Here we use $A \wedge B$ to denote the event that A and B both hold.

Consider the quadratic form $Q(x) := \sum_{i,j} a_{ij} \xi_i \xi_j$, and fix a non-trivial partition $\{1, \dots, n\} = U_1 \cup U_2$ and a non-empty subset S of U_1 . For instance one can take U_1 to be the first half of the indices and U_2 to be the second half. Define $Y := (\xi_i)_{i \in U_1}$ and $Z := (\xi_i)_{i \in U_2}$. We can write $Q(x) = Q(Y, Z)$. Let ξ'_i be an independent copy of ξ_i and set $Y' := (\xi'_i)_{i \in U_1}$ and $Z' := (\xi'_i)_{i \in U_2}$. By Lemma 4.2, for any number x

$$\mathbf{P}(Q(Y, Z) = x) \leq \mathbf{P}(Q(Y, Z) = Q(Y, Z') = Q(Y', Z) = Q(Y', Z') = x)^{1/4}.$$

On the other hand, if $Q(Y, Z) = Q(Y, Z') = Q(Y', Z) = Q(Y', Z') = x$ then regardless the value of x

$$R := Q(Y, Z) - Q(Y', Z) - Q(Y, Z') + Q(Y', Z') = 0.$$

Furthermore, we can write R as

$$R = \sum_{i \in U_1} \sum_{j \in U_2} a_{ij} (\xi_i - \xi'_i) (\xi_j - \xi'_j) = \sum_{i \in U_1} R_i w_i,$$

where w_i is the random variable $w_i := \xi_i - \xi'_i$, and R_i is the random variable $\sum_{j \in U_2} a_{ij} w_j$.

We now can conclude the proof by applying Theorem 3.1 twice. First, combining this theorem with a combinatorial argument, one can show that (with high probability), many R_i are non-zero. Next, one can condition on the non-zero R_i and apply Theorem 3.1 for the linear form $\sum_{i \in U_1} R_i w_i$ to obtain a bound on $\mathbf{P}(R = 0)$.

The upper bound $n^{-1/8}$ in Theorem 4.1 can be easily improved to $n^{-1/4}$. The optimal bound was obtained by Costello [7] using, among others, the inverse theorems from Section 7.

Theorem 4.3 (Quadratic Littlewood–Offord inequality). *Suppose that $a_{ij} \neq 0$, $1 \leq i, j \leq n$. Then*

$$\rho_q(A) \leq n^{-1/2+o(1)}.$$

The exponent $1/2 + o(1)$ is best possible (up to the $o(1)$ term) as demonstrated by the quadratic form $\sum_{i,j} \xi_i \xi_j = \left(\sum_{i=1}^n \xi_i \right)^2$. Both Theorems 4.1 and 4.3 hold in a general setting where the ξ_i are not necessary Bernoulli and only a fraction of the a_{ij} 's are non-zero.

One can extend the argument above to give bounds of the form n^{-c_k} for a general polynomial of degree k . However, due to the repeated use of the decoupling lemma, c_k decreases very fast with k .

Theorem 4.4. *Let f be a multilinear polynomial of real coefficients in n variables ξ_1, \dots, ξ_n with $m \times n^{k-1}$ monomials of maximum degree k . If ξ_i are iid Bernoulli random variables, then for any value x*

$$\mathbf{P}(f = x) = O\left(m^{-\frac{1}{2(k^2+k)/2}}\right).$$

By a more refined analysis, Razborov and Viola [42] recently obtained a better exponent of order roughly $\frac{1}{2k}$ (see Section 16). On the other hand, it might be the case that the bound $n^{-1/2+o(1)}$ holds for all degrees $k \geq 2$, under some reasonable assumption on the coefficients of the polynomial.

Quadratic (and higher degree) Littlewood–Offord bounds play important roles in the study of random symmetric matrices and Boolean circuits. We will discuss these applications in Sections 5 and 16, respectively.

5. APPLICATION: SINGULARITY OF RANDOM BERNOULLI MATRICES

Let M_n be a random matrix of size n whose entries are iid Bernoulli random variables. A notorious open problem in probabilistic combinatorics is to estimate p_n , the probability that M_n is singular (see [23, 57] for more details).

Conjecture 5.1. $p_n = (1/2 + o(1))^n$.

To give the reader a feeling about how the Littlewood–Offord problem can be useful in estimating p_n , let us consider the following process. We expose the rows of M_n one by one from the top. Assume that the first $n - 1$ rows are linearly independent and form a hyperplane with normal vector $\mathbf{v} = (a_1, \dots, a_n)$. Conditioned on these rows, the probability that M_n is singular is

$$\mathbf{P}(X \cdot \mathbf{v} = 0) = \mathbf{P}(a_1\xi_1 + \dots + a_n\xi_n = 0),$$

where $X = (\xi_1, \dots, \xi_n)$ is the last row.

As an illustration, let us give a short proof for the classical bound $p_n = o(1)$ (first showed by Komlós in [28] using a different argument).

Theorem 5.2. $p_n = o(1)$.

We with a simple observation [23].

Fact 5.3. *Let H be a subspace of dimension $1 \leq d \leq n$. Then H contains at most 2^d Bernoulli vectors.*

To see this, notice that in a subspace of dimension d , there is a set of d coordinates which determine the others. This fact implies

$$p_n \leq \sum_{i=1}^{n-1} \mathbf{P}(\mathbf{x}_{i+1} \in H_i) \leq \sum_{i=1}^{n-1} 2^{i-n} \leq 1 - \frac{2}{2^n},$$

where H_i is the subspace generated by the the first i rows $\mathbf{x}_1, \dots, \mathbf{x}_i$ of M_n .

This bound is quite the opposite of what we want to prove. However, we notice that the loss comes at the end. Thus, to obtain the desired upper bound $o(1)$, it suffices to show that the sum of the last (say) $\log \log n$ terms is at most (say) $\frac{1}{\log^{1/3} n}$. To do this, we will exploit the fact that the H_i are spanned by random vectors. The following lemma (which is a more effective version of the above fact) implies the theorem via the union bound.

Lemma 5.4. *Let H be the subspace spanned by d random vectors, where $d \geq n - \log \log n$. Then with probability at least $1 - \frac{1}{n}$, H contains at most $\frac{2^n}{\log^{1/3} n}$ Bernoulli vectors.*

We say that a set S of d vectors is k -universal if for any set of k different indices $1 \leq i_1, \dots, i_k \leq n$ and any set of signs $\varepsilon_1, \dots, \varepsilon_n$ ($\varepsilon_i = \pm 1$), there is a vector V in S such that the sign of the i_j -th coordinate of V matches ε_j , for all $1 \leq j \leq k$.

Fact 5.5. *If $d \geq n/2$, then with probability at least $1 - \frac{1}{n}$, a set of d random vectors is k -universal, for $k = \log n/10$.*

To prove this, notice that the failure probability is, by the union bound, at most

$$\binom{n}{k} \left(1 - \frac{1}{2^k}\right)^d \leq n^k \left(1 - \frac{1}{2^k}\right)^{n/2} \leq n^{-1}.$$

If S is k -universal, then any non-zero vector \mathbf{v} in the orthogonal complement of the subspace spanned by S should have more than k non-zero components (otherwise, there would be a vector in S having positive inner product with \mathbf{v}). If we fix such \mathbf{v} , and let \mathbf{x} be a random Bernoulli vector, then by Theorem 3.1

$$\mathbf{P}(\mathbf{x} \in \text{span}(S)) \leq \mathbf{P}(\mathbf{x} \cdot \mathbf{v} = 0) = O\left(\frac{1}{k^{1/2}}\right) = o\left(\frac{1}{\log^{1/3} n}\right),$$

proving Lemma 5.4 and Theorem 5.2.

The symmetric version of Theorem 5.2 is much harder and has been open for quite sometime (the problem was raised by Weiss the 1980s). Let p_n^{sym} be the singular probability of a random symmetric matrix whose upper diagonal entries are iid Bernoulli variables. Weiss conjectured that $p_n^{sym} = o(1)$. This was proved by Costello, Tao, and the second author [8]. Somewhat interestingly, this proof made use of the argument of Komlós in [28] which he applied for non-symmetric matrices. Instead of exposing the matrix row by row, one needs to expose the principal minors one by one, starting with the top left entry. At step i , one has a symmetric matrix M_i of size i and the next matrix M_{i+1} is obtained by adding a row and its transpose. Following Komlós, one defines X_i as the co-rank of the matrix at step i and shows that the sequence X_i behaves as a bias random walk with a positive drift. Carrying out the calculation carefully, one obtains that $X_n = 0$ with high probability.

The key technical step of this argument is to show that if M_i has full rank then so does M_{i+1} , with very high probability. Here the quadratic Littlewood–Offord bound is essential. Notice that if we condition on M_i , then $\det(M_{i+1})$ is a quadratic form of the entries in the additional $((i + 1)$ -th) row, with coefficients being the co-factors of M_i . By looking at these co-factors closely and using Theorem 4.1 (to be more precise, a variant of it where only a fraction of coefficients are required to be non-zero), one can establish Weiss’ conjecture.

Theorem 5.6.

$$p_n^{sym} = o(1).$$

Getting strong quantitative bounds for p_n and p_n^{sym} is more challenging, and we will continue this topic in Section 13 and 14, after the introduction of inverse theorems.

6. HALÁSZ’ RESULTS

In [21] (see also in [67]), Halász proved the following very general theorem.

Theorem 6.1. *Suppose that there exists a constant $\delta > 0$ such that the following holds*

- (General position) for any unit vector \mathbf{e} in \mathbf{R}^d one can select at least δn vectors a_k with $|\langle a_k, \mathbf{e} \rangle| \geq 1$;
- (Separation) among the 2^d vectors b of the form $\pm a_{k_1} \pm \dots \pm a_{k_d}$ one can select at least $\delta 2^d$ with pairwise distance at least 1.

Then

$$\rho_{d,1,Ber}(A) = O_{\delta,d}(n^{-3d/2}).$$

Halász’ method is Fourier analytic, which uses the following powerful Esséen-type concentration inequality as the starting point (see [21], [12]).

Lemma 6.2. *There exists an absolute positive constant $C = C(d)$ such that for any random variable X and any unit ball $\mathbf{B} \subset \mathbf{R}^d$*

$$(7) \quad \mathbf{P}(X \in \mathbf{B}) \leq C \int_{\|t\|_2 \leq 1} |\mathbf{E}(\exp(i\langle t, X \rangle))| dt.$$

Proof (of Lemma 6.2). With the function $k(t)$ to be defined later, let $K(x)$ be its Fourier’s transform

$$K(x) = \int_{\mathbf{R}^d} \exp(i\langle x, t \rangle) k(t) dt.$$

Let $H(x)$ be the distribution function and $h(x)$ be the characteristic function of X respectively. By Parseval’s identity we have

$$(8) \quad \int_{\mathbf{R}^d} K(x)dH(x) = \int_{\mathbf{R}^d} k(t)h(t)dt.$$

If we choose $k(t)$ so that

$$\begin{cases} k(t) = 0 & \text{for } \|t\|_2 \geq 1, \\ |k(t)| \leq c_1 & \text{for } \|t\|_2 \leq 1, \end{cases}$$

then the RHS of (8) is bounded by that of (7) modulo a constant factor.

Also, if

$$\begin{cases} K(x) \geq 1, \|x\|_2 \leq c_2, & \text{for some constant } c_2, \\ K(x) \geq 0 & \text{for } \|x\|_2 \geq c_2, \end{cases}$$

then the LHS of (8) is at least $\int_{\|x\|_2 \leq c_2} dH(x)$.

Similarly, by translating $K(x)$ (i.e. by multiplying $k(x)$ with a phase of $\exp(it_0, x)$), we obtain the same upper bound for $\int_{\|x-t_0\|_2 \leq c_2} dH(x)$. Thus, by covering the unit ball \mathbf{B} with balls of radius c_2 , we arrive at (7) for some constant C depending on d .

To construct $k(t)$ with the properties above, one may take it to have the convolution form

$$k(x) := \int_{x \in \mathbf{R}^d} k_1(x)k_1(t - x) dx,$$

where $k_1(x) = 1$ if $\|x\|_2 \leq 1/2$ and $k_1(x) = 0$ otherwise. ■

To illustrate Halász’ method, let us give a quick proof of Erdős’ bound $O(n^{-1/2})$ for the small ball probability $\rho_{1,1,Ber}(A)$ with A being a multi-set of n real numbers of absolute value at least 1. In view of Lemma 6.2, it suffices to show that

$$\int_{|t| \leq 1} \left| \mathbf{E} \left(\exp \left(it \sum_{j=1}^n a_j \xi_j \right) \right) \right| dt = O(1/\sqrt{n}).$$

By the independence of the ξ_j , we have

$$\left| \mathbf{E} \left(\exp \left(it \sum_{j=1}^n a_j \xi_j \right) \right) \right| = \prod_{j=1}^n |\mathbf{E}(\exp(it a_j \xi_j))| = \left| \prod_{j=1}^n \cos(t a_j) \right|.$$

By Hölder’s inequality

$$\int_{|t| \leq 1} \left| \mathbf{E} \left(\exp \left(it \sum_{j=1}^n a_j \xi_j \right) \right) \right| dt \leq \prod_{j=1}^n \left(\int_{|t| \leq 1} |\cos(t a_j)|^n dt \right)^{1/n}.$$

But since each a_j has magnitude at least 1, it is easy to check that $\int_{|t| \leq 1} |\cos(t a_j)|^n dt = O(1/\sqrt{n})$, and the claim follows.

Using Halász technique, it is possible to deduce

Corollary 6.3 [67, Corollary 7.16]. *Let A be a multi-set in \mathbf{R} . Let l be a fixed integer and R_l be the number of solutions of the equation $a_{i_1} + \dots + a_{i_l} = a_{j_1} + \dots + a_{j_l}$. Then*

$$\rho_A := \sup_x \mathbf{P}(S_A = x) = O\left(n^{-2l - \frac{1}{2}} R_l\right).$$

This result provides the hierarchy of bounds mentioned in the previous section, given that we forbid more and more additive structures on A . Let us consider the first few steps of the hierarchy.

- If the a_i ’s are distinct, then we can set $l = 1$ and $R_1 = n$ (the only solutions are the trivial ones $a_i = a_i$). Thus, we obtain Sárközy-Szemerédi’s bound $O(n^{-3/2})$.
- If we forbid the a_i ’s to satisfy equations $a_i + a_j = a_l + a_k$, for any $\{i, j\} \neq \{k, l\}$ (in particular this prohibits A to be an arithmetic progression), then one can fix $l = 2$ and $R_2 = n^2$ and obtain $\rho_A = O(n^{-5/2})$.
- If we continue to forbid equations of the form $a_h + a_i + a_j = a_k + a_l + a_m$, $\{h, i, j\} \neq \{k, l, m\}$, then one obtains $\rho_A = O(n^{-7/2})$ and so on.

Halász’ method is very powerful and has a strong influence on the recent developments discussed in the coming sections.

7. INVERSE THEOREMS: DISCRETE CASE

A few years ago, Tao and the second author [60] brought a new view to the small ball problem. Instead of working out a hierarchy of bounds by imposing new assumptions as done in Corollary 6.3, they tried to find the underlying reason as to why the small ball probability is large (say, polynomial in n).

It is easier and more natural to work with the discrete problem first. Let A be a multi-set of integers and ξ be the Bernoulli random variable.

Question 7.1 (Inverse problem, [60]). *Let $n \rightarrow \infty$. Assume that for some constant C*

$$\rho_A = \sup_x \mathbf{P}(S_A = x) \geq n^{-C}.$$

What can we say about the elements a_1, \dots, a_n of A ?

Denote by M the sum of all elements of A and rewrite $\sum_i a_i \xi_i$ as $M - 2 \sum_{i; \xi_i = -1} a_i$. As A has 2^n subsets, the bound $\rho_A \geq n^{-C}$ implies that at least $2^n/n^C$ among the subset sums are exactly $(M - x)/2$. This overwhelming collision suggests that A must have some strong additive structure. Tao and the second author proposed

Inverse Principle:

(9)

A set with large small ball probability must have strong additive structure.

The issue is, of course, to quantify the statement. Before attacking this question, let us recall the famous Freiman's inverse theorem from Additive Combinatorics. As the readers will see, this theorem strongly motivates our study.

In the 1970s, Freiman considered the collection of pairwise sums $A + A := \{a + a' \mid a, a' \in A\}$ [15]. Normally, one expects this collection to have $\Theta(|A|^2)$ elements. Freiman proved a deep and powerful theorem showing that if $A + A$ has only $O(|A|)$ elements (i.e., a huge number of collision occurs) then A must look like an arithmetic progression. (Notice that if A is an arithmetic progression then $|A + A| \approx 2|A|$.)

To make Freiman's statement more precise, we need the definition of *generalized arithmetic progressions* (GAPs).

Definition 7.2. A set Q of an abelian group Z is a *GAP of rank r* if it can be expressed in the form

$Q = \{g_0 + m_1 g_1 + \dots + m_r g_r \mid M_i \leq m_i \leq M'_i, m_i \in \mathbf{Z} \text{ for all } 1 \leq i \leq r\}$
for some $g_0, \dots, g_r \in Z$ and some real numbers $M_1, \dots, M_r, M'_1, \dots, M'_r$.

It is convenient to think of Q as the image of an integer box $B := \{(m_1, \dots, m_r) \in \mathbf{Z}^r \mid M_i \leq m_i \leq M'_i\}$ under the linear map

$$\Phi : (m_1, \dots, m_r) \mapsto g_0 + m_1 g_1 + \dots + m_r g_r.$$

The numbers g_i are the *generators* of Q , the numbers M'_i, M_i are the *dimensions* of Q , and $\text{Vol}(Q) := |B|$ is the *volume* of B . We say that Q is *proper* if this map is one to one, or equivalently if $|Q| = \text{Vol}(Q)$. For non-proper GAPs, we of course have $|Q| < \text{Vol}(Q)$. If $-M_i = M'_i$ for all $i \geq 1$ and $g_0 = 0$, we say that Q is *symmetric*.

If Q is symmetric and $t > 0$, the dilate tQ is the set

$$\{m_1 g_1 + \dots + m_r g_r \mid -tM'_i \leq m_i \leq tM'_i \text{ for all } 1 \leq i \leq r\}.$$

It is easy to see that if Q is a proper map of rank r , then $|Q + Q| \leq 2^r |Q|$. This implies that if A is a subset of density δ in a proper GAP Q of rank r , then as far as $\delta = \Theta(1)$,

$$|A + A| \leq |Q + Q| \leq 2^r |Q| \leq \frac{2^r}{\delta} |A| = O(|A|).$$

Thus, dense subsets of a proper GAP of constant rank satisfies the assumption $|A + A| = O(|A|)$. Freiman’s remarkable inverse theorem showed that this example is the only one.

Theorem 7.3 (Freiman’s inverse theorem in \mathbf{Z}). *Let γ be a given positive number. Let X be a set in \mathbf{Z} such that $|X + X| \leq \gamma |X|$. Then there exists a proper GAP of rank $O_\gamma(1)$ and cardinality $O_\gamma(|X|)$ that contains X .*

For further discussions, including a beautiful proof by Ruzsa, see [67, Chapter 5]; see also [5] for recent and deep developments concerning non-commutative settings (when A is a subset of a non-abelian group).

In our case, we want to find examples for A such that

$$\rho(A) := \sup_x \mathbf{P}(S_A = x)$$

is large. Again, dense subsets of a proper GAP come in as natural candidates.

Example 7.4. Let Q be a proper symmetric GAP of rank r and volume N . Let a_1, \dots, a_n be (not necessarily distinct) elements of Q . By the Central Limit Theorem, with probability at least $2/3$, the random sum

$S_A = \sum_{i=1}^n a_i x_i$ takes value in the dilate $10n^{1/2}Q$. Since $|tQ| \leq t^r N$, by the pigeon hole principle, we can conclude that there is a point x where

$$\mathbf{P}(S_A = x) = \Omega\left(\frac{1}{n^{r/2}N}\right).$$

Thus if $|Q| = N = O(n^{C-r/2})$ for some constant $C \geq r/2$, then

$$\rho(A) \geq \mathbf{P}(S_A = x) = \Omega\left(\frac{1}{n^C}\right).$$

This example shows that if the elements of A are elements of a symmetric proper GAP with a small rank and small cardinality, then $\rho(A)$ is large. Inspired by Freiman's theorem, Tao and the second author [62, 60] showed that the converse is also true.

Theorem 7.5. *For any constant C, ε there are constants r, B such that the following holds. Let A be a multi-set of n real numbers such that $\rho(A) \geq n^{-C}$, then there is a GAP Q of rank r and volume n^B such that all but n^ε elements of A belong to Q .*

The dependence of B on C, ε is not explicit in [60]. In [62], Tao and the second author obtained an almost sharp dependence. The best dependence, which mirrors Example 7.4 was proved in a more recent paper [39] of the current authors. This proof is different from those in earlier proofs and made a direct use of Freiman's theorem (see Appendix A).

Theorem 7.6 (Optimal inverse Littlewood–Offord theorem, discrete case) [39]. *Let $\varepsilon < 1$ and C be positive constants. Assume that*

$$\rho(A) \geq n^{-C}.$$

Then for any $n^\varepsilon \leq n' \leq n$, there exists a proper symmetric GAP Q of rank $r = O_{C,\varepsilon}(1)$ which contains all but at most n' elements of A (counting multiplicities), where

$$|Q| = O_{C,\varepsilon}(\rho(A)^{-1} n'^{-\frac{r}{2}}).$$

In particular, there exists a proper symmetric GAP of rank $O_{C,\varepsilon}(1)$ and cardinality $O_{C,\varepsilon}(\rho(A)^{-1} n^{-\frac{r}{2}})$ which contains all but at most εn elements of A (counting multiplicities).

The existence of the exceptional set cannot be avoided completely. For more discussions, see [60, 39]. There is also a trade-off between the size of the exceptional set and the bound on $|Q|$. In fact, the main result of [62] has a better bound on the exceptional set with a loss of a small polynomial factor in the volume bound.

Let us also point out that the above inverse theorems hold in a very general setting where the random variables ξ_i are not necessarily Bernoulli and independent (see [60, 62, 39, 38] for more details).

8. APPLICATION: FROM INVERSE TO FORWARD

One can use the “inverse” Theorem 7.6 to quickly prove several “forward” theorems presented in earlier sections. As an example, let us derive Theorems 3.1 and 3.3.

Proof (of Theorem 3.1). Assume, for contradiction, that there is a set A of n non-zero numbers such that $\rho(A) \geq c_1 n^{-1/2}$ for some large constant c_1 to be chosen. Set $\varepsilon = .1, C = 1/2$. By Theorem 7.6, there is a GAP Q of rank r and size $O\left(\frac{1}{c_1} n^{C-\frac{r}{2}}\right)$ that contains at least $.9n$ elements from A . However, by setting c_1 to be sufficiently large (compared to the constant in big O) and using the fact that $C = 1/2$ and $r \geq 1$, we can force $O\left(\frac{1}{c_1} n^{C-\frac{r}{2}}\right) < 1$. Thus, Q has to be empty, a contradiction. ■

Proof (of Theorem 3.3). Similarly, assume that there is a set A of n distinct numbers such that $\rho(A) \geq c_1 n^{-3/2}$ for some large constant c_1 to be chosen. Set $\varepsilon = .1, C = 3/2$. By Theorem 7.6, there is a GAP Q of rank r and size $O\left(\frac{1}{c_1} n^{C-\frac{r}{2}}\right)$ that contains at least $.9n$ elements from A . This implies $|Q| \geq .9n$. By setting c_1 to be sufficiently large and using the fact that $C = 3/2$ and $r \geq 1$, we can guarantee that $|Q| \leq .8n$, a contradiction. ■

The readers are invited to work out the proof of Corollary 6.3.

Let us now consider another application of Theorem 7.6, which enables us to make very precise counting arguments. Assume that we would like to count the number of multi-sets A of integers with $\max |a_i| \leq M = n^{O(1)}$ such that $\rho(A) \geq n^{-C}$.

Fix $d \geq 1$, fix¹ a GAP Q with rank r and volume $|Q| \leq c\rho(A)^{-1}n^{-\frac{r}{2}}$ for some constant c depending on C and ε . The dominating term in the calculation will be the number of multi-sets which intersect with Q in subsets of size at least $(1 - \varepsilon)n$. This number is bounded by

$$(10) \quad \sum_{k \leq \varepsilon n} |Q|^{n-k} (2M)^k \leq \sum_{k \leq \varepsilon n} \left(c\rho(A)^{-1}n^{-\frac{r}{2}} \right)^{n-k} (2M)^k$$

$$\leq (O_{C,\varepsilon}(1))^n n^{O_\varepsilon(1)n} \rho(A)^{-n} n^{-\frac{n}{2}}.$$

We thus obtain the following useful result.

Theorem 8.1 (Counting theorem: Discrete case). *The number N of multi-sets A of integers with $\max |a_i| \leq n^{C_1}$ and $\rho(A) \geq n^{-C_2}$ is bounded by*

$$N = (O_{C_1,C_2,\varepsilon}(1))^n n^{O_\varepsilon(1)n} (\rho(A)^{-1}n^{-1/2})^n,$$

where ε is an arbitrary constant between 0 and 1.

Due to their asymptotic nature, our inverse theorems do not directly imply Stanley’s precise result (Theorem 3.4). However, by refining the proofs, one can actually get very close and with some bonus, namely, additional strong *rigidity* information. For instance, in [37] the first author showed that if the elements of A are distinct, then

$$\mathbf{P}(S_A = x) \leq \left(\sqrt{\frac{24}{\pi}} + o(1) \right) n^{-3/2},$$

where the constant on the RHS is obtained when A is the symmetric arithmetic progression A_0 from Theorem 3.4. It was showed that if $\rho(A)$ is close to this value, then A needs to be very close to a symmetric arithmetic progression.

Theorem 8.2 [37]. *There exists a positive constant ε_0 such that for any $0 < \varepsilon \leq \varepsilon_0$, there exists a positive number $\varepsilon' = \varepsilon'(\varepsilon)$ such that $\varepsilon' \rightarrow 0$ as $\varepsilon \rightarrow 0$ and the following holds: if A is a set of n distinct integers and*

$$\rho(A) \geq \left(\sqrt{\frac{24}{\pi}} - \varepsilon \right) n^{-\frac{3}{2}},$$

¹A more detailed version of Theorem 7.6 tells us that there are not too many ways to choose the generators of Q . In particular, if $|a_i| \leq M = n^{O(1)}$, the number of ways to fix these is negligible compared to the main term.

then there exists an integer l which divides all $a \in A$ and

$$\sum_{a \in A} \left(\frac{a}{l}\right)^2 \leq (1 + \varepsilon') \sum_{a \in A_0} a^2 = (1 + \varepsilon' + o(1)) \frac{n^3}{12}.$$

We remark that a slightly weaker stability can be shown even when we have a much weaker assumption $\rho(A) \geq \varepsilon n^{-3/2}$.

As the reader will see, in many applications in the following sections, we do not use the inverse theorems directly, but rather their counting corollaries, such as Theorem 8.1. Such counting results can be used to bound the probability of a bad event through the union bound (they count the number of terms in the union). This method was first used in studies of random matrices [57, 60, 45], but it is simpler to illustrate the idea by the following more recent result of Conlon, Fox, and Sudakov [6].

A Hilbert cube is a set of the form $x_0 + \Sigma(\{x_1, \dots, x_d\})$ where $\Sigma(X) = \{\sum_{x \in Y} x | Y \subset X\}$, and $0 \leq x_0, 0 < x_1 < \dots < x_d$ are integers. Following the literature, we refer to the index d as the dimension. One of the earliest results in Ramsey theory is a theorem of Hilbert [22] stating that for any fixed r and d and n sufficiently large, any coloring of the set $[n] := \{1, \dots, n\}$ with r colors must contain a monochromatic Hilbert cube of dimension d . Let $h(d, r)$ be the smallest such n . The best known upper bound for this function is [22, 20]

$$h(d, r) \leq (2r)^{2^{d-1}}.$$

The density version of [55] states that for any natural number d and $\delta > 0$ there exists an n_0 such that if $n \geq n_0$ then any subset of n of density δ contains a Hilbert cube of dimension d . One can show that

$$d \geq c \log \log n$$

where c is a positive constant depending only on δ .

On the other hand, Hegyvári shows an upper bound of the form $O(\sqrt{\log n \log \log n})$ by considering a random subset of density δ . Using the discrete inverse theorems (Section 7), Conlon, Fox, and Sudakov [6] removed the $\log \log n$ term, obtaining $O(\sqrt{\log n})$, which is sharp up to the constant in big O , thanks to another result of Hegyvári.

Conlon et al. started with the following corollary of Theorem 7.5.

Lemma 8.3. *For every $C > 0$, $1 > \varepsilon > 0$ there exist positive constants r and C' such that if X is a multiset with d elements and $|\Sigma(X)| \leq d^C$, then there is a GAP Q of dimension r and volume at most $d^{C'}$ such that all but at most $d^{1-\varepsilon}$ elements of X are contained in Q .*

From this, one can easily prove the following counting lemma.

Lemma 8.4. *For $s \leq \log d$, the number of d -sets $X \subset [n]$ with $\Sigma(X) \leq 2^s d^2$ is at most $n^{O(s)} d^{O(d)}$.*

Let A be a random set of $[n]$ obtained by choosing each number with probability δ independently. Let E be the event that A contains a Hilbert cube of dimension $c\sqrt{\log n}$. We aim to show that

$$(11) \quad \mathbf{P}(E) = o(1),$$

given c sufficiently large.

Trivially $\mathbf{P}(E) \leq n \sum_{X \subset [n]} \delta^{|\Sigma(X)|}$, where the factor n corresponds to the number of ways to choose x_0 . Let m_t be the number of X such that $|\Sigma(X)| = t$. The RHS can be bounded from above by $n \sum_t m_t \delta^t$.

If t is large, say $t \geq d^3$, we just crudely bound $\sum_{t \geq d^3} m_t$ by n^d (which is the total number of ways to choose x_1, \dots, x_d). The contribution in probability in this case is at most $n \times n^d \times \delta^{d^3} = o(1)$, if c is sufficiently large. In the case $t < d^3$, we make use of the counting lemma above to bound m_t and a routine calculation finishes the job.

9. INVERSE THEOREMS: CONTINUOUS CASE I.

In this section and the next, we consider sets with large small ball probability.

We say that a vector $v \in \mathbf{R}^d$ is δ -close to a set $Q \subset \mathbf{R}^d$ if there exists a vector $q \in Q$ such that $\|v - q\|_2 \leq \delta$. A set X is δ -close to a set Q if every element of X is δ -close to Q . The continuous analogue of Example 7.4 is the following.

Example 9.1. Let Q be a proper symmetric GAP of rank r and volume N in \mathbf{R}^d . Let a_1, \dots, a_n be (not necessarily distinct) vectors which are $\frac{1}{100}\beta n^{-1/2}$ -close to Q . Again by the Central Limit Theorem, with probability at least $2/3$, S_A is β -close to $10n^{1/2}Q$. Thus, by the pigeon hole principle, there is a point x in $100n^{1/2}Q$ such that

$$\mathbf{P}(S_A \in B(x, \beta)) \geq |10n^{1/2}Q|^{-1} \geq \Omega(n^{-r/2}|Q|^{-1}).$$

It follows that if Q has cardinality $n^{C-\frac{r}{2}}$ for some constant $C \geq r/2$, then

$$(12) \quad \rho_{d,\beta,Ber}(A) = \Omega\left(\frac{1}{n^C}\right).$$

Thus, in view of the Inverse Principle (9) and Theorem 7.6, we would expect that if $\rho_{d,\beta,Ber}(A)$ is large, then most of the a_i is close to a GAP with small volume. This statement turned out to hold for very general random variable ξ (not only for Bernoulli). In practice, we can consider any real random variable ξ , which satisfies the following condition: there are positive constants C_1, C_2, C_3 such that

$$(13) \quad \mathbf{P}(C_1 \leq |\xi_1 - \xi_2| \leq C_2) \geq C_3,$$

where ξ_1, ξ_2 are iid copies of ξ .

Theorem 9.2 [39]. *Let ξ be a real random variable satisfying (13). Let $0 < \varepsilon < 1; 0 < C$ be constants and $\beta > 0$ be a parameter that may depend on n . Suppose that $A = \{a_1, \dots, a_n\}$ is a (multi-)subset of \mathbf{R}^d such that $\sum_{i=1}^n \|a_i\|_2^2 = 1$ and that A has large small ball probability*

$$\rho := \rho_{d,\beta,\xi}(A) \geq n^{-C}.$$

Then there exists a symmetric proper GAP Q of rank $r \geq d$ and of size $|Q| = O(\rho^{-1}n^{-(r+d)/2})$ such that all but εn elements of A are $O\left(\frac{\beta \log n}{n^{1/2}}\right)$ -close to Q .

In applications, one often chooses β to be at least $\exp(-n^\varepsilon)$ for some small constant ε . Our next result gives more information about Q , but with a weaker approximation.

Theorem 9.3. *Under the assumption of the above theorem, the following holds. For any number n' between n^ε and n , there exists a proper symmetric GAP $Q = \{\sum_{i=1}^r x_i g_i : |x_i| \leq L_i\}$ such that*

- *At least $n - n'$ elements of A are β -close to Q .*
- *Q has small rank, $r = O(1)$, and small cardinality*

$$|Q| \leq \max \left(O \left(\frac{\rho^{-1}}{\sqrt{n'}} \right), 1 \right).$$

- *There is a non-zero integer $p = O(\sqrt{n'})$ such that all steps g_i of Q have the form $g_i = (g_{i1}, \dots, g_{id})$, where $g_{ij} = \beta \frac{p_{ij}}{p}$ with $p_{ij} \in \mathbf{Z}$ and $p_{ij} = O(\beta^{-1}\sqrt{n'})$.*

Theorem 9.3 immediately implies the following result which can be seen as a continuous analogue of Theorem 8.1. This result was first proved by Tao and the second author for the purpose of verifying the Circular Law in random matrix theory [58, 60] using a more complicated argument.

Let n be a positive integer and β, ρ be positive numbers that may depend on n . Let $\mathcal{S}_{n,\beta,\rho}$ be the collection of all multisets $A = \{a_1, \dots, a_n\}, a_i \in \mathbf{R}^2$ such that $\sum_{i=1}^n \|a_i\|_2^2 = 1$ and $\rho_{2,\beta,Ber}(A) \geq \rho$.

Theorem 9.4 (Counting theorem, continuous case) [58, 60]. *Let $0 < \varepsilon \leq 1/3$ and $C > 0$ be constants. Then, for all sufficiently large n and $\beta \geq \exp(-n^\varepsilon)$ and $\rho \geq n^{-C}$ there is a set $\mathcal{S} \subset (\mathbf{R}^2)^n$ of size at most*

$$\rho^{-n} n^{-n(\frac{1}{2}-\varepsilon)} + \exp(o(n))$$

such that for any $A = \{a_1, \dots, a_n\} \in \mathcal{S}_{n,\beta,\rho}$ there is some $A' = (a'_1, \dots, a'_n) \in \mathcal{S}$ such that $\|a_i - a'_i\|_2 \leq \beta$ for all i .

Proof (of Theorem 9.4). Set $n' := n^{1-\frac{3\varepsilon}{2}}$ (which is $\gg n^\varepsilon$ as $\varepsilon \leq 1/3$). Let \mathcal{S}' be the collection of all subsets of size at least $n - n'$ of GAPs whose parameters satisfy the conclusion of Theorem 9.3.

Since each GAP is determined by its generators and dimensions, the number of such GAPs is bounded by

$$((\beta^{-1}\sqrt{n'})\sqrt{n'})^{O(1)} \left(\frac{\rho^{-1}}{\sqrt{n'}}\right)^{O(1)} = \exp(o(n)).$$

(The term $\left(\frac{\rho^{-1}}{\sqrt{n'}}\right)^{O(1)}$ bounds the number of choices of the dimensions M_i .)
Thus

$$|\mathcal{S}'| = \left(O\left(\left(\frac{\rho^{-1}}{\sqrt{n'}}\right)^n\right) + 1\right) \exp(o(n)).$$

We approximate each of the exceptional elements by a lattice point in $\beta \cdot (\mathbf{Z}/d)^d$. Thus if we let \mathcal{S}'' to be the set of these approximated tuples then $|\mathcal{S}''| \leq \sum_{i \leq n'} (O(\beta^{-1}))^i = \exp(o(n))$ (here we used the assumption $\beta \geq \exp(-n^\varepsilon)$).

Set $\mathcal{S} := \mathcal{S}' \times \mathcal{S}''$. It is easy to see that $|\mathcal{S}| \leq O(n^{-1/2+\varepsilon}\rho^{-1})^n + \exp(o(n))$. Furthermore, if $\rho(A) \geq n^{-O(1)}$ then A is β -close to an element of \mathcal{S} , concluding the proof. ■

10. INVERSE THEOREMS: CONTINUOUS CASE II.

Another realization of the Inverse Principle (9) was given by Rudelson and Vershynin in [45, 47] (see also Friedland and Sodin [16]). Let a_1, \dots, a_n be real numbers. Rudelson and Vershynin defined the essential *least common*

denominator (**LCD**) of $\mathbf{a} = (a_1, \dots, a_n)$ as follows. Fix parameters α and γ , where $\gamma \in (0, 1)$, and define

$$\mathbf{LCD}_{\alpha, \gamma}(\mathbf{a}) := \inf \left\{ \theta > 0 : \text{dist}(\theta \mathbf{a}, \mathbf{Z}^n) < \min(\gamma \|\theta \mathbf{a}\|_2, \alpha) \right\},$$

where the distance from a vector \mathbf{v} to a set S is defined by $\text{dist}(\mathbf{v}, S) := \inf_{\mathbf{s} \in S} \|\mathbf{v} - \mathbf{s}\|_2$.

The requirement that the distance is smaller than $\gamma \|\theta \mathbf{a}\|_2$ forces us to consider only non-trivial integer points as approximations of $\theta \mathbf{a}$. One typically assume γ to be a small constant, and $\alpha = c\sqrt{n}$ with a small constant $c > 0$. The inequality $\text{dist}(\theta \mathbf{a}, \mathbf{Z}^n) < \alpha$ then yields that most coordinates of $\theta \mathbf{a}$ are within a small distance from non-zero integers.

Theorem 10.1 (Diophantine approximation [45, 46]). *Consider a sequence $A = \{a_1, \dots, a_n\}$ of real numbers which satisfies $\sum_{i=1}^n a_i^2 \geq 1$. Let ξ be a random variable such that $\sup_a \mathbf{P}(\xi \in B(a, 1)) \leq 1 - b$ for some $b > 0$, and x_1, \dots, x_n be iid copies of ξ . Then, for every $\alpha > 0$ and $\gamma \in (0, 1)$, and for*

$$\beta \geq \frac{1}{\mathbf{LCD}_{\alpha, \gamma}(\mathbf{a})},$$

we have

$$\rho_{1, \beta, \xi}(A) \leq \frac{C\beta}{\gamma\sqrt{b}} + Ce^{-2b\alpha^2}.$$

One can use Theorem 10.1 to prove a special case of the forward result of Erdős and Littlewood–Offord when most of the a_i have the same order of magnitude (see [45, p. 6]).² Indeed, assume that $K_1 \leq |a_i| \leq K_2$ for all i , where $K_2 = cK_1$ with $c = O(1)$. Set $a'_i := a_i / \sqrt{\sum_j a_j^2}$ and $\mathbf{a}' := (a'_1, \dots, a'_n)$. Choose $\gamma = c_1, \alpha = c_2\sqrt{n}$ with sufficiently small positive constants c_1, c_2 (depending on c), the condition $\text{dist}(\theta \mathbf{a}', \mathbf{Z}^n) < \min(\gamma \|\theta \mathbf{a}'\|_2, \alpha)$ implies that $|\theta a'_i - n_i| \leq 1/3$ with $n_i \in \mathbf{Z}, n_i \neq 0$ for at least $c_3 n$ indices i , where c_3 is a positive constant depending on c_1, c_2 . It then follows that for these indices, $\theta^2 a_i'^2 \geq 4n_i^2/9$. Summing over i , we obtain $\theta^2 = \Omega(n)$ and so $\mathbf{LCD}_{\alpha, \gamma}(\mathbf{a}') = \Omega(\sqrt{n})$. Applying Theorem 10.1 to the vector \mathbf{a}' with $\beta = 1/\mathbf{LCD}_{\alpha, \gamma}(\mathbf{a}')$, we obtain the desired upper bound $O(1/\sqrt{n})$ for the concentration probability.

Theorems 10.1 is not exactly *inverse* in the Freiman sense. On the other hand, it is convenient to use and in most applications provides a sufficient

²One can also handle this case by conditioning on the abnormal a_i and use Berry-Esseen for the remaining sum.

amount of structural information that allows one derive a counting theorem. An extra advantage here is that this theorem enables one to consider sets A with small ball probability as small as $(1 - \varepsilon)^n$, rather than just n^{-C} as in Theorem 9.2.

The definition of the essential least common denominator above can be extended naturally to higher dimensions. To this end, we define the product of such multi-vector \mathbf{a} and a vector $\theta \in \mathbf{R}^d$ as

$$\theta \cdot \mathbf{a} = (\langle \theta, a_1 \rangle, \dots, \langle \theta, a_n \rangle) \in \mathbf{R}^n.$$

Then we define, for $\alpha > 0$ and $\gamma \in (0, 1)$,

$$\mathbf{LCD}_{\alpha, \gamma}(\mathbf{a}) := \inf \left\{ \|\theta\|_2 : \theta \in \mathbf{R}^d, \text{dist}(\theta \cdot \mathbf{a}, \mathbf{Z}^n) < \min(\gamma \|\theta \cdot \mathbf{a}\|_2, \alpha) \right\}.$$

The following generalization of Theorem 10.1 gives a bound on the small ball probability for the random sum $\sum_{i=1}^n a_i x_i$ in terms of the additive structure of the coefficient sequence \mathbf{a} .

Theorem 10.2 (Diophantine approximation, multi-dimensional case) [46, 16]. Consider a sequence $A = \{a_1, \dots, a_n\}$ of vectors $a_i \in \mathbf{R}^d$, which satisfies

$$(14) \quad \sum_{i=1}^n \langle a_i, x \rangle^2 \geq \|x\|_2^2 \quad \text{for every } x \in \mathbf{R}^d.$$

Let ξ be a random variable such that $\sup_a \mathbf{P}(\xi \in B(a, 1)) \leq 1 - b$ for some $b > 0$ and x_1, \dots, x_n be iid copies of ξ .

Then, for every $\alpha > 0$ and $\gamma \in (0, 1)$, and for

$$\beta \geq \frac{\sqrt{d}}{\mathbf{LCD}_{\alpha, \gamma}(\mathbf{a})},$$

we have

$$\rho_{d, \beta \sqrt{d}, \xi}(A) \leq \left(\frac{C\beta}{\gamma \sqrt{b}} \right)^d + C^d e^{-2b\alpha^2}.$$

We will sketch the proof of Theorem 10.1 in Appendix B.

11. INVERSE QUADRATIC LITTLEWOOD–OFFORD

In this section, we revisit the quadratic Littlewood–Offord bound in Section 4 and consider its inverse. We first consider a few examples of A where (the quadratic small ball probability) $\rho_q(A)$ is large.

Example 11.1 (Additive structure implies large small ball probability). Let Q be a proper symmetric GAP of rank $r = O(1)$ and of size $n^{O(1)}$. Assume that $a_{ij} \in Q$, then for any $\xi_i \in \{\pm 1\}$

$$\sum_{i,j} a_{ij} \xi_i \xi_j \in n^2 Q.$$

Thus, by the pigeon-hole principle,

$$\rho_q(A) \geq n^{-2r} |Q|^{-1} = n^{-O(1)}.$$

But unlike the linear case, additive structure is not the only source for large small ball probability. Our next example shows that algebra also plays a role.

Example 11.2 (Algebraic structure implies large small ball probability). Assume that

$$a_{ij} = k_i b_j + k_j b_i$$

where $k_i \in \mathbf{Z}, |k_i| = n^{O(1)}$ and such that $\mathbf{P}(\sum_i k_i \xi_i = 0) = n^{-O(1)}$.

Then we have

$$\mathbf{P}\left(\sum_{i,j} a_{ij} \xi_i \xi_j = 0\right) = \mathbf{P}\left(\sum_i k_i \xi_i \sum_j b_j \xi_j = 0\right) = n^{-O(1)}.$$

Combining the above two examples, we have the following general one.

Example 11.3 (Structure implies large small ball probability). Assume that $a_{ij} = a'_{ij} + a''_{ij}$, where $a'_{ij} \in Q$, a proper symmetric GAP of rank $O(1)$ and size $n^{O(1)}$, and

$$a''_{ij} = k_{i1} b_{1j} + k_{j1} b_{1i} + \dots + k_{ir} b_{rj} + k_{jr} b_{ri},$$

where b_{1i}, \dots, b_{ri} are arbitrary and k_{i1}, \dots, k_{ir} are integers bounded by $n^{O(1)}$, and $r = O(1)$ such that

$$\mathbf{P}\left(\sum_i k_{i1} \xi_i = 0, \dots, \sum_i k_{ir} \xi_i = 0\right) = n^{-O(1)}.$$

Then we have

$$\sum_{i,j} a_{ij} \xi_i \xi_j = \sum_{i,j} a'_{ij} \xi_i \xi_j + 2 \left(\sum_i k_{i1} \xi_i \right) \left(\sum_j b_{1j} \xi_j \right) + \cdots + 2 \left(\sum_i k_{ir} \xi_i \right) \left(\sum_j b_{rj} \xi_j \right).$$

Thus,

$$\mathbf{P} \left(\sum_{i,j} a_{ij} \xi_i \xi_j \in n^2 Q \right) = n^{-O(1)}.$$

It then follows, by the pigeon-hole principle, that $\rho_q(A) = n^{-O(1)}$.

We have demonstrated the fact that as long as most of the a_{ij} can be decomposed as $a_{ij} = a'_{ij} + a''_{ij}$, where a'_{ij} belongs to a GAP of rank $O(1)$ and size $n^{O(1)}$ and the symmetric matrix (a''_{ij}) has rank $O(1)$, then $A = (a_{ij})$ has large quadratic small ball probability. The first author in [36] showed that sort of the converse is also true.

Theorem 11.4 (Inverse Littlewood–Offord theorem for quadratic forms). *Let $\varepsilon < 1, C$ be positive constants. Assume that*

$$\rho_q(A) \geq n^{-C}.$$

Then there exist index sets I_0, I of size $O_{C,\varepsilon}(1)$ and $(1 - \varepsilon)n$ respectively, with $I \cap I_0 = \emptyset$, and there exist integers k_{ii_0} (for any pair $i_0 \in I_0, i \in I$) of size bounded by $n^{O_{C,\varepsilon}(1)}$, and a structured set Q of the form

$$Q = \left\{ \sum_{h=1}^{O_C(1)} \frac{p_h}{q_h} g_h \mid p_h \in \mathbf{Z}, |p_h|, |q_h| = n^{O_{C,\varepsilon}(1)} \right\},$$

such that for all $i \in I$ the followings holds:

- (low rank decomposition) for any $j \in I$,

$$a_{ij} = a'_{ij} - \left(\sum_{i_0 \in I_0} k_{ii_0} a_{i_0j} + \sum_{i_0 \in I_0} k_{ji_0} a_{i_0i} \right);$$

- (common additive structure of small size) all but εn entries a'_{ij} belong to Q .

We remark that the common structure Q is not yet a GAP, as the coefficients are rational, instead of being integers. It is desirable to have an analogue of Theorem 7.6 with common structure as a genuine GAP with optimal parameters (see for instance [7, Conjecture 1] for a precise conjecture for bilinear forms.) For counting purposes, this inverse theorem is sufficiently strong.

12. APPLICATION: THE LEAST SINGULAR VALUE OF A RANDOM MATRIX

For a matrix A , let $\sigma_n(A)$ denote its smallest singular value. It is well known that $\sigma_n(A) \geq 0$ and the bound is strict if and only if A is non-singular. An important problem with many practical applications is to bound the least singular value of a non-singular matrix (see [17, 52, 53, 63, 47, 9] for discussions). The problem of estimating the least singular value of a random matrix was first raised by Goldstine and von Neumann [17] in the 1940s, with connection to their investigation of the complexity of inverting a matrix.

To answer Goldstine and von Neumann’s question, Edelman [9] computed the distribution of the LSV of the random matrix M_n^{Gau} of size n with iid standard gaussian entries, and showed that for all fixed $t > 0$

$$\begin{aligned} \mathbf{P}(\sigma_n(M_n^{Gau}) \leq tn^{-1/2}) &= \int_0^t \frac{1 + \sqrt{x}}{2\sqrt{x}} e^{-(x/2 + \sqrt{x})} dx + o(1) \\ &= t - \frac{1}{3}t^3 + O(t^4) + o(1). \end{aligned}$$

He conjectured that this distribution is universal (i.e., it must hold for other models of random matrices, such as Bernoulli).

More recently, in their study of smoothed analysis of the simplex method, Spielman and Teng [52, 53] showed that for any $t > 0$ (t can go to 0 with n)

$$(15) \quad \mathbf{P}(\sigma_n(M_n^{Gau}) \leq tn^{-1/2}) \leq t.$$

They conjectured that a slightly adjusted bound holds in the Bernoulli case [52]

$$(16) \quad \mathbf{P}(\sigma_n(M_n^{Ber}) \leq t) \leq tn^{1/2} + c^n,$$

where $0 < c < 1$ is a constant. The term c^n is needed as M_n^{Ber} can be singular with exponentially small ball probability.

Edelman's conjecture has been proved by Tao and the second author in [64]. This work also confirms Spielman and Teng's conjecture for the case t is fairly large; $t \geq n^{-\delta}$ for some small constant $\delta > 0$. For $t \geq n^{-3/2}$, Rudelson in [44], making use of Halász' machinery from [21], obtained a strong bound with an extra (multiplicative) constant factor. In many applications, it is important to be able to treat even smaller t . As a matter of fact, in applications what one usually needs is the probability bound to be very small, but this requires one to set t very small automatically.

In the last few years, thanks to the development of inverse theorems, one can now prove very strong bound for almost all range of t .

Consider a matrix M with row vectors X_i and singular values $\sigma_1 \geq \dots \geq \sigma_n$. Let d_i be the distance from X_i to the hyperplane formed by the other $n - 1$ rows. There are several ways to exhibit a direct relation between the d_i and σ_i . For instance, Tao and the second showed [58]

$$(17) \quad d_1^{-2} + \dots + d_n^{-2} = \sigma_1^{-2} + \dots + \sigma_n^{-2}.$$

A technical relation, but in certain applications more effective, is [45, Lemma 3.5].

From this, it is clear that if one can bound the d_i from below with high probability, then one can do the same for σ_n . Let $v = (a_1, \dots, a_n)$ be the normal vector of the hyperplane formed by X_2, \dots, X_n and ξ_1, \dots, ξ_n be the coordinates of X_1 , then

$$d_1 = |a_1 \xi_1 + \dots + a_n \xi_n|.$$

Thus, the probability that d_1 is small is exactly the small ball probability for the multi-set $A = \{a_1, \dots, a_n\}$. If this probability is large, then the inverse theorems tell us that the set A must have strong additive structure. However, A comes as the normal vector of a random hyperplane, so the probability that it has any special structure is very small (to quantify this we can use the counting theorems such as Theorem 9.4). Thus, we obtain, with high probability, a lower bound on all d_i . In principle, one can use this to deduce a lower bound for σ_n .

Carrying out the above plan requires certain extra ideas and some careful analysis. In [60], Tao and the second author managed to prove

Theorem 12.1. *For any constant $A > 0$, there is a constant $B > 0$ such that*

$$\mathbf{P}(\sigma_n(M_n^{Ber}) \leq n^{-B}) \leq n^{-A}.$$

The first inverse theorem, Theorem 7.5, was first proved in this paper, as a step in the proof of Theorem 12.1. In a consequent paper, Rudelson and Vershynin developed Theorem 10.1, and used it, in combination with [45, Lemma 3.5] and many other ideas to show

Theorem 12.2. *There is a constant $C > 0$ and $0 < c < 1$ such that for any $t > 0$,*

$$\mathbf{P}(\sigma_n(M_n^{Ber}) \leq tn^{-1/2}) \leq tn^{1/2} + c^n.$$

This bound is sharp, up to the constant C . It also gives a new proof of Kahn-Komlós-Szemerédi bound on the singularity probability of a random Bernoulli matrix (see Section 13). Both theorems hold in more general setting.

In practice, one often works with random matrices of the type $A + M_n$ where A is deterministic and M_n has iid entries. (For instance, in their works on smoothed analysis, Spielman and Teng used this to model a large data matrix perturbed by random noise.) They proved in [52]

Theorem 12.3. *Let A be an arbitrary n by n matrix. Then for any $t > 0$,*

$$\mathbf{P}(\sigma_n(A + M_n^{Gau}) \leq tn^{-1/2}) = O(t).$$

One may ask whether there is an analogue of Theorem 12.2 for this model. The answer is, somewhat surprisingly, negative. An analogue of the weaker Theorem 12.1 is, however, available, assuming that $\|A\|$ is bounded polynomially in n . For more discussion on this model, we refer to [63]. For applications in Random Matrix Theory (such as the establishment of the Circular Law) and many related results, we refer to [59, 65, 58, 18, 40, 2, 47] and the references therein.

13. APPLICATION: STRONG BOUNDS ON THE SINGULARITY PROBLEM – THE NON-SYMMETRIC CASE

We continue to discuss the singularity problem from Section 5. The first exponential bound on p_n was proved by Kahn, Komlós and Szemerédi [23], who showed that $p_n \leq .999^n$. In [56], Tao and the second author simplified the proof and got a slightly improved bound $.952^n$. A more notable improvement which pushed the bound to $(3/4 + o(1))^n$ was obtained in a subsequent paper [57], which combined Kahn et al. approach with an inverse theorem. The best current bound is $(1/\sqrt{2} + o(1))^n$ by Bourgain, Vu and Wood [3]. The proof of this bound still relied heavily on the approach

from [57] (in particular it used the same inverse theorem), but added a new twist which made the first part of the argument more effective.

In the following, we tried to present the approach from [23] and [57]. Similar to the proof in Appendix A, we first embed the problem in a finite field $\mathbf{F} = \mathbf{F}_p$, where p is a very large prime. Let $\{-1, 1\}^n \subset \mathbf{F}^n$ be the discrete unit cube in \mathbf{F}^n . We let X be the random variable taking values in $\{-1, 1\}^n$ which is distributed uniformly on this cube (thus each element of $\{-1, 1\}^n$ is attained with probability 2^{-n}). Let $X_1, \dots, X_n \in \{-1, 1\}$ be n independent samples of X . Then

$$p_n := \mathbf{P}(X_1, \dots, X_n \text{ linearly dependent}).$$

For each linear subspace V of \mathbf{F}^n , let A_V denote the event that X_1, \dots, X_n span V . Let us call a space V *non-trivial* if it is spanned by the set $V \cap \{-1, 1\}^n$. Note that $\mathbf{P}(A_V) \neq 0$ if and only if V is non-trivial. Since every collection of n linearly dependent vectors in \mathbf{F}^n will span exactly one proper subspace V of \mathbf{F}^n , we have

$$(18) \quad p_n = \sum_{V \text{ a proper non-trivial subspace of } \mathbf{F}^n} \mathbf{P}(A_V).$$

It is not hard to show that the dominant contribution to this sum came from the hyperplanes:

$$p_n = 2^{o(n)} \sum_{V \text{ a non-trivial hyperplane in } \mathbf{F}^n} \mathbf{P}(A_V).$$

Thus, if one wants to show $p_n \leq (3/4 + o(1))^n$, it suffices to show

$$\sum_{V \text{ a non-trivial hyperplane in } \mathbf{F}^n} \mathbf{P}(A_V) \leq (3/4 + o(1))^n.$$

The next step is to partition the non-trivial hyperplanes V into a number of classes, depending on the number of $(-1, 1)$ vectors in V .

Definition 13.1 (Combinatorial dimension). Let $D := \{d_{\pm} \in \mathbf{Z}/n : 1 \leq d_{\pm} \leq n\}$. For any $d_{\pm} \in D$, we define the *combinatorial Grassmannian* $\mathbf{Gr}(d_{\pm})$ to be the set of all non-trivial hyperplanes V in \mathbf{F}^n with

$$(19) \quad 2^{d_{\pm}-1/n} < |V \cap \{-1, 1\}^n| \leq 2^{d_{\pm}}.$$

We will refer to d_{\pm} as the *combinatorial dimension* of V .

It thus suffices to show that

$$(20) \quad \sum_{d_{\pm} \in D} \sum_{V \in \mathbf{Gr}(d_{\pm})} \mathbf{P}(A_V) \leq \left(\frac{3}{4} + o(1)\right)^n.$$

It is therefore of interest to understand the size of the combinatorial Grassmannians $\mathbf{Gr}(d_{\pm})$ and of the probability of the events A_V for hyperplanes V in those Grassmannians.

There are two easy cases, one when d_{\pm} is fairly small and one where d_{\pm} is fairly large.

Lemma 13.2 (Small combinatorial dimension estimate). *Let $0 < \alpha < 1$ be arbitrary. Then*

$$\sum_{d_{\pm} \in D: 2^{d_{\pm}-n} \leq \alpha^n} \sum_{V \in \mathbf{Gr}(d_{\pm})} \mathbf{P}(A_V) \leq n\alpha^n.$$

Proof (of Lemma 13.2). Observe that if X_1, \dots, X_n span V , then there are $n - 1$ vectors among the X_i which already span V . By symmetry, we thus have

$$(21) \quad \mathbf{P}(A_V) = \mathbf{P}(X_1, \dots, X_n \text{ span } V) \leq n\mathbf{P}(X_1, \dots, X_{n-1} \text{ span } V)\mathbf{P}(X \in V).$$

On the other hand, if $V \in \mathbf{Gr}(d_{\pm})$ and $2^{d_{\pm}-n} \leq \alpha^n$, then $\mathbf{P}(X \in V) \leq \alpha^n$ thanks to (19). Thus we have

$$\mathbf{P}(A_V) \leq n\alpha^n \mathbf{P}(X_1, \dots, X_{n-1} \text{ span } V).$$

Since X_1, \dots, X_{n-1} can span at most one space V , the claim follows. ■

Lemma 13.3 (Large combinatorial dimension estimate). *We have*

$$\sum_{d_{\pm} \in D: 2^{d_{\pm}-n} \geq 100/\sqrt{n}} \sum_{V \in \mathbf{Gr}(d_{\pm})} \mathbf{P}(A_V) \leq (1 + o(1))n^2 2^{-n}.$$

This proof uses Theorem 3.1 and is left as an exercise; consult [23, 57] for details. The heart of the matter is the following, somewhat more difficult, result.

Proposition 13.4 (Medium combinatorial dimension estimate). *Let $0 < \varepsilon_0 \ll 1$, and let $d_{\pm} \in D$ be such that $(\frac{3}{4} + 2\varepsilon_0)^n < 2^{d_{\pm}-n} < \frac{100}{\sqrt{n}}$. Then we have*

$$\sum_{V \in \mathbf{Gr}(d_{\pm})} \mathbf{P}(A_V) \leq o(1)^n,$$

where the rate of decay in the $o(1)$ quantity depends on ε_0 (but not on d_{\pm}).

Note that D has cardinality $|D| = O(n^2)$. Thus if we combine this proposition with Lemma 13.2 (with $\alpha := \frac{3}{4} + 2\varepsilon_0$) and Lemma 13.3, we see that we can bound the left-hand side of (20) by

$$n \left(\frac{3}{4} + 2\varepsilon_0\right)^n + n^2 o(1)^n + (1 + o(1))n^2 2^{-n} = \left(\frac{3}{4} + 2\varepsilon_0 + o(1)\right)^n.$$

Since ε_0 is arbitrary, the upper bound $(3/4 + o(1))^n$ follows.

We now informally discuss the proof of Proposition 13.4. We start with the trivial bound

$$(22) \quad \sum_{V \in \mathbf{Gr}(d_{\pm})} \mathbf{P}(A_V) \leq 1$$

that arises simply because any vectors X_1, \dots, X_n can span at most one space V . To improve upon this trivial bound, the key innovation in [23] is to replace X by another random variable Y which tends to be more concentrated on subspaces V than X is. Roughly speaking, one seeks the property

$$(23) \quad \mathbf{P}(X \in V) \leq c\mathbf{P}(Y \in V)$$

for some absolute constant $0 < c < 1$ and for all (or almost all) subspaces $V \in \mathbf{Gr}(d_{\pm})$. From this property, one expects (heuristically, at least)

$$(24) \quad \mathbf{P}(A_V) = \mathbf{P}(X_1, \dots, X_n \text{ span } V) \leq c^n \mathbf{P}(Y_1, \dots, Y_n \text{ span } V),$$

where Y_1, \dots, Y_n are iid samples of Y , and then by applying the trivial bound (22) with Y instead of X , we would then obtain a bound of the form $\sum_{V \in \mathbf{Gr}(d_{\pm})} \mathbf{P}(A_V) \leq c^n$, at least in principle. Clearly, it will be desirable to make c as small as possible; if we can make c arbitrarily small, we will have established Proposition 13.4.

The random variable Y can be described as follows. Let $0 \leq \mu \leq 1$ be a small absolute constant (in [23] the value $\mu = \frac{1}{108}e^{-1/108}$ was chosen), and

let $\eta^{(\mu)}$ be a random variable taking values in $\{-1, 0, 1\} \subset F$ which equals 0 with probability $1 - \mu$ and equals $+1$ or -1 with probability $\mu/2$ each. Then let $Y := (\eta_1^{(\mu)}, \dots, \eta_n^{(\mu)}) \in F^n$, where $\eta_1^{(\mu)}, \dots, \eta_n^{(\mu)}$ are iid samples of $\eta^{(\mu)}$. By using a Fourier-analytic argument of Halász [21], a bound of the form

$$\mathbf{P}(X \in V) \leq C\sqrt{\mu}\mathbf{P}(Y \in V)$$

was shown in [23], where C was an absolute constant (independent of μ), and V was a hyperplane which was *non-degenerate* in the sense that its combinatorial dimension was not too close to n . For μ sufficiently small, one then obtains (23) for some $0 < c < 1$, although one cannot make c arbitrarily small without shrinking μ also.

There are however some technical difficulties with this approach, arising when one tries to pass from (23) to (24). The first problem is that the random variable Y , when conditioned on the event $Y \in V$, may concentrate on a lower dimensional subspace on V , making it unlikely that Y_1, \dots, Y_n will span V . In particular, Y has a probability of $(1 - \mu)^n$ of being the zero vector, which basically means that one cannot hope to exploit (23) in any non-trivial way once $\mathbf{P}(X \in V) \leq (1 - \mu)^n$. However, in this case V has very low combinatorial dimension and Lemma 13.2 already gives an exponential gain.

Even when $(1 - \mu)^n < \mathbf{P}(X \in V) \leq 1$, it turns out that it is still not particularly easy to obtain (24), but one can obtain an acceptable substitute for this estimate by only replacing some of the X_j by Y_j . Specifically, one can try to obtain an estimate roughly of the form

$$(25) \quad \mathbf{P}(X_1, \dots, X_n \text{ span } V) \leq c^m \mathbf{P}(Y_1, \dots, Y_m, X_1, \dots, X_{n-m} \text{ span } V)$$

where m is equal to a suitably small multiple of n (we will eventually take $m \approx n/100$). Strictly speaking, we will also have to absorb an additional “entropy” loss of $\binom{n}{m}$ for technical reasons, though as we will be taking c arbitrarily small, this loss will ultimately be irrelevant.

The above approach (with some minor modifications) was carried out rigorously in [23] to give the bound $p_n = O(.999^n)$ which has been improved slightly to $O(.952^n)$ in [56], thanks to some simplifications. There are two main reasons why the final gain in the base was relatively small. Firstly, the chosen value of μ was small (so the $n(1 - \mu)^n$ error was sizeable), and secondly the value of c obtained was relatively large (so the gain of c^n or $c^{(1-\gamma)n}$ was relatively weak). Unfortunately, increasing μ also causes c to increase, and so even after optimizing μ and c one falls well short of the conjectured bound.

The more significant improvement to $(3/4 + o(1))^n$ relies on an inverse theorem. To reduce all the other losses to $(\frac{3}{4} + 2\varepsilon_0)^n$ for some small ε_0 , we increase μ up to $1/4 - \varepsilon_0/100$, at which point the arguments of Halász and [23, 56] give (23) with $c = 1$. The value $1/4$ for μ is optimal as it is the largest number satisfying the pointwise inequality

$$|\cos(x)| \leq (1 - \mu) + \mu \cos(2x) \text{ for all } x \in \mathbf{R},$$

which is the Fourier-analytic analogue of (23) (with $c = 1$). At first glance, the fact that $c = 1$ seems to remove any utility to (23), as the above argument relied on obtaining gains of the form c^n or $c^{(1-\gamma)n}$. However, we can proceed further by subdividing the collection of hyperplanes $\mathbf{Gr}(d_{\pm})$ into two classes, namely the *unexceptional* spaces V for which

$$\mathbf{P}(X \in V) < \varepsilon_1 \mathbf{P}(Y \in V)$$

for some small constant $0 < \varepsilon_1 \ll 1$ to be chosen later (it will be much smaller than ε_0), and the *exceptional* spaces for which

$$(26) \quad \varepsilon_1 \mathbf{P}(Y \in V) \leq \mathbf{P}(X \in V) \leq \mathbf{P}(Y \in V).$$

The contribution of the unexceptional spaces can be dealt with by the preceding arguments to obtain a very small contribution (at most δ^n for any fixed $\delta > 0$ given that we set $\varepsilon_1 = \varepsilon_1(\gamma, \delta)$ suitably small), so it remains to consider the exceptional spaces V .

The key technical step is to show that there are very few exceptional hyperplanes (and thus their contribution is negligible). This can be done using the following inverse theorem (the way the counting Theorem 8.1 was proved using the inverse Theorem 7.6).

Let $V \in \mathbf{Gr}(d_{\pm})$ be an exceptional space, with a representation of the form

$$(27) \quad V = \{(x_1, \dots, x_n) \in F^n : x_1 a_1 + \dots + x_n a_n = 0\}$$

for some elements $a_1, \dots, a_n \in F$. We shall refer to a_1, \dots, a_n as the *defining co-ordinates* for V .

Theorem 13.5. *There is a constant $C = C(\varepsilon_0, \varepsilon_1)$ such that the following holds. Let V be a hyperplane in $\mathbf{Gr}(d_{\pm})$ and a_1, \dots, a_n be its defining co-ordinates. Then there exist integers*

$$(28) \quad 1 \leq r \leq C$$

and $M_1, \dots, M_r \geq 1$ with the volume bound

$$(29) \quad M_1 \dots M_r \leq C 2^{n-d_{\pm}}$$

and non-zero elements $v_1, \dots, v_r \in F$ such that the following holds.

- (Defining coordinates lie in a progression) The symmetric generalized arithmetic progression

$$P := \{m_1 v_1 + \dots + m_r v_r : -M_j/2 < m_j < M_j/2 \text{ for all } 1 \leq j \leq r\}$$

is proper and contains all the a_i .

- (Bounded norm) The a_i have small P -norm:

$$(30) \quad \sum_{j=1}^n \|a_j\|_P^2 \leq C$$

- (Rational commensurability) The set $\{v_1, \dots, v_r\} \cup \{a_1, \dots, a_n\}$ is contained in the set

$$(31) \quad \left\{ \frac{p}{q} v_1 : p, q \in \mathbf{Z}; q \neq 0; |p|, |q| \leq n^{o(n)} \right\}.$$

14. APPLICATION: STRONG BOUNDS ON THE SINGULARITY PROBLEM – THE SYMMETRIC CASE

Similar to Conjecture 5.1, we raise

Conjecture 14.1.

$$p_n^{sym} = (1/2 + o(1))^n.$$

We are very far from this conjecture. Currently, no exponential upper bound is known. The first superpolynomial bound was obtained by the first author [36] very recently.

Theorem 14.2 [36]. *For any $C > 0$ and n sufficiently large*

$$p_n^{sym} \leq n^{-C}.$$

Shortly after, Vershynyn [69] proved the following better bound

Theorem 14.3. *There exists a positive constant c such that*

$$p_n^{sym} = O(\exp(-n^c)).$$

Both proofs made essential use of inverse theorems. The first author used the inverse quadratic Theorem 11.4 and Vershynyn’s proof used Theorem 10.1 several times.

In the following, we sketched the main ideas behind Theorem 14.2. Let $\mathbf{r} = (\xi_1, \dots, \xi_n)$ be the first row of M_n , and $a_{ij}, 2 \leq i, j \leq n$, be the cofactors of M_{n-1} obtained by removing \mathbf{r} and \mathbf{r}^T from M_n . We have

$$(32) \quad \det(M_n) = \xi_1^2 \det(M_{n-1}) + \sum_{2 \leq i, j \leq n} a_{ij} \xi_i \xi_j.$$

Recalling the proof of Theorem 5.6 (see Section 5). One first need to show that with high probability (with respect to M_{n-1}) a good fraction of the co-factors a_{ij} are nonzero. Theorem 4.1 then yields that

$$\mathbf{P}_{\mathbf{r}}(\det(M_n) = 0) \leq n^{-1/8+o(1)} = o(1).$$

To prove Theorem 14.2, we adapt the reversed approach, which, similar to the previous proofs, consists of an inverse statement and a counting step.

- (1) (Inverse step). If $\mathbf{P}_{\mathbf{r}}(\det(M_n) = 0 | M_{n-1}) \geq n^{-O(1)}$, then there is a strong additive structure among the cofactors a_{ij} .
- (2) (Counting step). With respect to M_{n-1} , a strong additive structure among the a_{ij} occurs with negligible probability.

By (32), one notices that the first step concentrates on the study of inverse Littlewood–Offord problem for quadratic forms $\sum_{i,j} a_{ij} \xi_i \xi_j$. Roughly speaking, Theorem 11.4 implies that most of the a_{ij} belong to a common structure. Thus, by extracting the structure on one row of the array $A = (a_{ij})$, we obtain a vector which is orthogonal to the remaining $n - 2$ rows of the matrix M_{n-1} . Executing the argument more carefully, we obtain the following lemma.

Lemma 14.4 (Inverse Step). *Let $\varepsilon < 1$ and C be positive constants. Assume that M_{n-1} has rank at least $n - 2$ and that*

$$\mathbf{P}_{\mathbf{r}} \left(\sum_{i,j} a_{ij} \xi_i \xi_j = 0 | M_{n-1} \right) \geq n^{-C}.$$

Then there exists a nonzero vector $\mathbf{u} = (u_1, \dots, u_{n-1})$ with the following properties.

- *All but n^ε elements of u_i belong to a proper symmetric generalized arithmetic progression of rank $O_{C,\varepsilon}(1)$ and size $n^{O_{C,\varepsilon}(1)}$.*
- *$u_i \in \{p/q : p, q \in \mathbf{Z}, |p|, |q| = n^{O_{C,\varepsilon}(n^\varepsilon)}\}$ for all i .*
- *\mathbf{u} is orthogonal to $n - O_{C,\varepsilon}(n^\varepsilon)$ rows of M_{n-1} .*

Let \mathcal{P} denote the collection of all \mathbf{u} satisfying the properties above. For each $\mathbf{u} \in \mathcal{P}$, let $\mathbf{P}_{\mathbf{u}}$ be the probability, with respect to M_{n-1} , that u is orthogonal to $n - O_{C,\varepsilon}(n^\varepsilon)$ rows of M_{n-1} . The following lemma takes care of our second step.

Lemma 14.5 (Counting Step). *We have*

$$\sum_{\mathbf{u} \in \mathcal{P}} \mathbf{P}_{\mathbf{u}} = O_{C,\varepsilon}((1/2)^{(1-o(1))n}).$$

The main contribution in the sum in Lemma 14.5 comes from those \mathbf{u} which have just a few non-zero components (i.e. compressible vectors). For incompressible vectors, we classify it into dyadic classes $\mathcal{C}_{\rho_1, \dots, \rho_{n-1}}$, where ρ_i is at most twice and at least half the probability $\mathbf{P}(\xi_1 u_1 + \dots + \xi_u u_i = 0)$. Assume that $\mathbf{u} \in \mathcal{C}_{\rho_1, \dots, \rho_{n-1}}$. Then by definition, as M_{n-1} is symmetric, the probability $\mathbf{P}_{\mathbf{u}}$ is bounded by $\prod O(\rho_i)$. On the other hand, by taking into account the structure of generalized arithmetic progressions, a variant of Theorem 8.1 shows that the size of each $\mathcal{C}_{\rho_1, \dots, \rho_{n-1}}$ is bounded by $\prod_i O(\rho_i) n^{-1/2+o(1)}$. Summing $P_{\mathbf{u}}$ over all classes \mathcal{C} , notice that the number of these classes are negligible, one obtains an upper bound of order $n^{-(1-o(1))n/2}$ for the compressible vectors.

We remark that it is in the Inverse Step that we obtain the final bound n^{-C} on the singular probability. In [69], Vershynin worked with a more general setting where one can assume a better bound. In this regime, he has been able to apply a variant of Theorem 10.1 to prove a very mild inverse-type result which is easy to be adapted for the Counting Step. As the details are complex, we invite the reader to consult [69].

15. APPLICATION: COMMON ROOTS OF RANDOM POLYNOMIALS

Let d be fixed. With $\vec{j}_d = (j_1, \dots, j_d), j_i \in \mathbf{Z}^+$ and $|\vec{j}_d| = \sum j_i$, let $\xi_{\vec{j}_d}$ be iid copies of a random variable ξ . Set $x^{\vec{j}_d} = \prod x_i^{j_i}$. Consider the random polynomial

$$P(x_1, \dots, x_d) = \sum_{\vec{j}_d, |\vec{j}_d| \leq n} \xi_{\vec{j}_d} x^{\vec{j}_d}$$

of degree n in d variables. (Here d is fixed and $n \rightarrow \infty$.) Random polynomials is a classical subject in analysis and probability and we refer to [4] for a survey.

In this section, we consider the following natural question. Let P_1, \dots, P_{d+1} be $d + 1$ independent random polynomials, each have d variables and degree n .

Question 15.1. *What is the probability that P_1, \dots, P_{d+1} have a common root?*

For short, let us denote the probability under consideration by $p(n, d)$

$$p(n, d) := \mathbf{P}(\exists x \in \mathbf{C}^d : P_i(x) = 0, i = 1, \dots, d + 1).$$

When ξ has continuous distribution, it is obvious that $p(n, d) = 0$. However, the situation is less clear when ξ has discrete distribution, even in the case $d = 1$. Indeed, when n is even and $P_1(x), P_2(x)$ are two independent random Bernoulli polynomials of one variable, then one has $\mathbf{P}(P_1(1) = P_2(1) = 0) = \Theta(1/n)$ and $\mathbf{P}(P_1(-1) = P_2(-1) = 0) = \Theta(1/n)$. Thus in this case $p(n, 1) = \Omega(1/n)$.

In a recent paper, Kozma and Zeitouni [32] proved $p(n, d) = O(1/n)$, answering Question 15.1 in the asymptotic sense.

Theorem 15.2. *For any fixed d there exists a constant $c(d)$ such that the following holds. Let P_1, \dots, P_{d+1} be $d + 1$ independent random Bernoulli polynomials in d variables and degree n .*

$$p(n, d) \leq c(d)/n.$$

In the sequel, we will focus on the case $d = 1$. This first case already captures some of the main ideas, especially the use of inverse theorems. The reader is invited to consult [32] for further details.

Theorem 15.3. *Let P_1, P_2 be two independent Bernoulli random polynomials in one variable of degree n . Then*

$$p(n, 1) = \begin{cases} O(n^{-1}) & n \text{ even} \\ O(n^{-3/2}) & n \text{ odd.} \end{cases}$$

Notice that the bounds in both cases are sharp. To start the proof, first observe that, because the coefficients of P_1 are ± 1 , all roots x of P_1 have magnitude $1/2 < |x| < 2$. Furthermore, x must be an algebraic integer. We will try to classify the common roots by their unique irreducible polynomial, relying on the following easy algebraic fact [32]:

Fact 15.4. *For every k there are only finitely many numbers whose irreducible polynomial has degree k that can be roots of a polynomial of arbitrary degree with coefficients ± 1 .*

Now we look at the event of having common roots. Assume that P_1 is fixed (i.e. condition on P_1) and let x_1, \dots, x_n be its n complex roots. For each x_i , we consider the probability that x_i is a root of $P_2(x)$. If $\mathbf{P}(P_2(x_i) = 0) \leq n^{-5/2}$ for all i , then $\mathbf{P}(\exists x \in \mathbf{C} : P_1(x) = P_2(x)) = O(n^{-3/2})$, and there is nothing to prove. We now consider the case $\mathbf{P}(P_2(x_i) = 0) \geq n^{-5/2}$ for some root x_i of $P_1(x)$. Notice that

$$\mathbf{P}(P_2(x_i) = 0) = \mathbf{P}_{\xi_0, \dots, \xi_n} \left(\sum_{j=0}^n \xi_j x_i^j = 0 \right) = \rho(X),$$

where X is the geometric progression $X = \{1, x_i, \dots, x_i^n\}$.

Now Theorem 7.6 comes into play. As $\rho(X) \geq n^{-5/2}$, most of the terms of X are additively correlated. On the other hand, as X is a geometric progression, this is the case only if x_i is a root of a bounded degree polynomial with well-controlled rational coefficients.

Lemma 15.5. *For any $C > 0$, there exists n_0 such that if $n > n_0$, and if*

$$\rho(X) \geq n^{-C},$$

where $X = \{1, x, \dots, x^n\}$. Then x is an algebraic number of degree at most $2C$.

Proof (of Lemma 15.5). Set $\varepsilon = 1/(2C + 2)$. Theorem 7.6, applied to the set X , implies that there exists a GAP Q of rank r and size $|Q| = O_C(n^{C-r/2})$ which contains at least $(2C + 1)/(2C + 2)$ -portion of the elements of X . By pigeon-hole principle, there exists $2C + 1$ consecutive terms of X , say $x^{i_0}, \dots, x^{i_0+2C}$, all of which belong to Q .

As $|Q| \geq 1$, the rank r of Q must be at most $2C$. Thus there exist integral coefficients m_1, \dots, m_{2C+1} , all of which are bounded by $n^{O_C(1)}$, such that the linear combination $\sum_{i=0}^{2C} m_i x^{i_0+i}$ vanishes. In particular, it follows that x is an algebraic number of degree at most $2C$. ■

We now prove Theorem 15.3. Write

$$\begin{aligned} p(n, 1) &= \mathbf{P}(\exists x \in \mathbf{C} : P_1(x) = P_2(x) = 0) \\ &\leq \mathbf{P}(P_1(1) = P_2(1) = 0) + \mathbf{P}(P_1(-1) = P_2(-1) = 0) \end{aligned}$$

$$\begin{aligned}
 &+ \mathbf{P}(\exists x \text{ of algebraic degree } 2, 3, 4, 5 : P_1(x) = P_2(x) = 0) \\
 &+ \mathbf{P}(\exists x \text{ of algebraic degree } \geq 6 : P_1(x) = P_2(x) = 0) \\
 &= S_1 + S_2 + S_3.
 \end{aligned}$$

For the first term, it is clear that $S_1 = \Theta(n^{-1})$ if n is even, and $S_1 = 0$ otherwise. For the second term S_2 , by Lemma 15.4, the number of possible common roots x of algebraic degree at most 5 is $O(1)$, so it suffices to show that $\mathbf{P}(P_1(x) = P_2(x)) = n^{-3/2}$ for each such x . On the other hand, by Lemma 15.5 we must have $\mathbf{P}(P_i(x) = 0) \leq n^{-3/4}$ because x cannot be a rational number (i.e. algebraic number of degree one). Thus we have

$$\mathbf{P}(P_1(x) = P_2(x) = 0) = \mathbf{P}(P_1(x) = 0)\mathbf{P}(P_2(x) = 0) \leq n^{-3/2}.$$

Lastly, in order to bound S_3 we first fix $P_1(x)$. It has at most n roots x of algebraic degree at least 6. For each of these roots, by Lemma 15.5, $\mathbf{P}(P_2(x) = 0) = O(n^{-5/2})$. Thus the probability that P_2 has at least a common root with P_1 which is an algebraic number of degree at least 6 is bounded by $n \times O(n^{-5/2}) = O(n^{-3/2})$. As a result, $S_3 = O(n^{-3/2})$.

16. APPLICATION: LITTLEWOOD–OFFORD TYPE BOUND FOR MULTILINEAR FORMS AND BOOLEAN CIRCUITS

Let k be a fixed positive integer, and $p(\xi_1, \dots, \xi_n) = \sum_{S \in [n]^{\leq k}} c_S \xi_S$ be a random multi-linear polynomial of degree at most k , where ξ_i are iid Bernoulli variables (taking values $\{0, 1\}$ with equal probability) and $\xi_S = \prod_{i \in S} \xi_i$. As mentioned in Section 4, by generalizing the proof of Theorem 4.1, Costelo, Tao and the second author proved the following

Theorem 16.1. *Let K denote the number of non-zero coefficients c_S , and set $m := K/n^{k-1}$. Then for any real number x we have*

$$\mathbf{P}(p = x) = O\left(m^{-\frac{1}{2^{(k^2+k)/2}}}\right).$$

Using a finer analysis, Razborov and Viola [42] improved the exponent $\frac{1}{2^{(k^2+k)/2}}$ to $\frac{1}{2k2^k}$.

Theorem 16.2. Let $p(\xi_1, \dots, \xi_n) = \sum_{S \in [n]^{\leq k}} c_S \xi_S$ be a multi-linear polynomial of degree k , and assume that there exist r terms $\xi_{S_1}, \dots, \xi_{S_r}$ of degree k each where the S_i are mutually disjoint and $c_{S_i} \neq 0$. Then for any real number x we have

$$\mathbf{P}(p = x) = O(r^{-b_k}),$$

where $b_k = (2k2^k)^{-1}$.

One observes that $r = \Omega(m/k)$, where m was defined in Theorem 16.1. Indeed, assume that the collection $\{S_1, \dots, S_r\}$ is maximal (with respect to disjointness). Then every set S with $c_S \neq 0$, either ξ_S has degree less than k or S intersects one of the S_i . Thus $K = O(rkn^{k-1})$, and so $r = \Omega(m/k)$.

It is a very interesting question (in its own right and for applications) to improve the exponent further. In the rest of this section, we are going to discuss Razborov and Viola’s main application of Theorem 16.2.

For two functions $f, g : \{0, 1\}^n \rightarrow \mathbf{R}$, one defines their correlation as

$$\mathbf{Cor}_n(f, g) := \mathbf{P}(f(\xi_1, \dots, \xi_n) = g(\xi_1, \dots, \xi_n)) - 1/2,$$

where ξ_i are iid Bernoulli variables taking values $\{0, 1\}$ with equal probability.

Most of the research in Complexity Theory has so far concentrated on the case in which both f and g are Boolean functions (that is $f(x), g(x) \in \{0, 1\}$). To incorporate into this framework arbitrary multivariate polynomials, one converts them to Boolean functions. There are two popular ways of doing this. For a polynomial p with integer coefficients, define a Boolean function $b(x) = 1$ if $m|p(x)$, where m is a given integer, and 0 otherwise. These functions b are called *modular polynomials*. For arbitrary p , one can set $b(x) = 1$ if $p(x) > t$ for some given threshold t , and 0 otherwise. We refer to these functions b as *threshold polynomials*. For further discussion on these polynomials, we refer the reader to [34, 35].

It is an open problem to exhibit an explicit Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ such that $\mathbf{Cor}_n(b, f) = o(1/\sqrt{n})$ for any modular polynomial b whose underlying polynomial p has degree $\log_2 n$ (see [70]). The same problem is also open for threshold polynomials.

In [42], Razborov and Viola initiated a similar study for the correlation of multi-variable polynomials where any output outside of $\{0, 1\}$ is counted as an error. They highlighted the following problem.

Problem 16.3. Exhibit an explicit Boolean function f such that $\mathbf{Cor}_n(p, f) = o(1/\sqrt{n})$ for any real polynomial $p : \{0, 1\}^n \rightarrow \mathbf{R}$ of degree $\log_2 n$.

It is well-known that analogies between polynomial approximations and matrix approximations are important and influential in theory and other areas like Machine Learning (see for instance [51]). Viewed under this angle, Razborov and Viola's model is a straightforward analogy of matrix rigidity [68] that still remains one of the main unresolved problems in the modern Complexity Theory. For further discussion and motivation, we refer to [42] and the references therein. It is noted that solving Problem 16.3 is a pre-requisite for solving the corresponding open problem for threshold polynomials. Similarly, the special case of Problem 16.3 when the polynomials have integer coefficients is a pre-requisite for solving the corresponding open problem for modular polynomials. As a quick application of Theorem 16.2, we demonstrate here a result addressing the question for lower degree polynomials.

Theorem 16.4 [42, Theorem 1.2]. *We have $\mathbf{Cor}_n(p, \text{parity}) \leq 0$ for every sufficiently large n and every real polynomial $p : \{0, 1\}^n \rightarrow \mathbf{R}$ of degree at most $\log_2 \log_2 n/2$.*

Proof (of Theorem 16.4). First we suppose that the hypothesis of Theorem 16.2 is satisfied with $r = \sqrt{n}$. Then the probability that the polynomial outputs a Boolean value is bounded by

$$2 \times O\left(\left(1/\sqrt{n}\right)^{\frac{1}{2k^2}}\right) \leq 1/2,$$

where $k \leq \frac{1}{2} \log_2 \log_2 n$.

Otherwise, we can cover all the terms of degree k by $k\sqrt{n}$ variables. Freeze these variables and iterate. After at most k iterations, either the hypothesis of Theorem 16.2 is satisfied with $r = \sqrt{n}$ (and with smaller degree), in which case we would be done, or else we end up with a degree-one polynomial with $n - O(k^2)\sqrt{n} \geq 1$ variables, in which case the statement is true by comparison with the parity function. ■

17. APPLICATION: SOLVING FRANKL AND FÜREDI'S CONJECTURE

In this section, we return to the discussion in Section 2 and give a proof of Conjecture 2.4 and a new proof for Theorem 2.2. Both proofs are based on the following inverse theorem.

Theorem 17.1. *For any fixed d there is a constant C such that the following holds. Let $A = \{a_1, \dots, a_n\}$ be a multi-set of vectors in \mathbf{R}^d such*

that $p_{d,1,Ber}(A) \geq Ck^{-d/2}$. Then A is "almost" flat. Namely, there is a hyperplane H such that $\text{dist}(a_i, H) \geq 1$ for at most k values of $i = 1, \dots, n$.

The proof of this theorem combines Esseén's bound (Lemma 6.2) together with some geometric arguments. For details, see [66]; $\text{dist}(a, H_i)$, of course, means the distance from a to H_i .

We first prove Theorem 2.2 by induction on the dimension d . The case $d = 1$ follows from Theorem 2.1, so we assume that $d \geq 2$ and that the claim has already been proven for smaller values of d . It suffices to prove the upper bound

$$p(d, R, Ber, n) \leq (1 + o(1))2^{-n}S(n, s).$$

Fix R , and let $\varepsilon > 0$ be a small parameter to be chosen later. Suppose the claim failed, then there exists $R > 0$ such that for arbitrarily large n , there exist a multi-set $A = \{a_1, \dots, a_n\}$ of vectors in \mathbf{R}^d of length at least 1 and a ball B of radius R such that

$$(33) \quad \mathbf{P}(S_A \in B) \geq (1 + \varepsilon)2^{-n}S(n, s).$$

In particular, from Stirling's approximation one has

$$\mathbf{P}(S_A \in B) \gg n^{-1/2}.$$

Applying the pigeonhole principle, we can find a ball B_0 of radius $\frac{1}{\log n}$ such that

$$\mathbf{P}(S_A \in B_0) \gg n^{-1/2} \log^{-d} n.$$

Set $k := n^{2/3}$. Since $d \geq 2$ and n is large, we have

$$\mathbf{P}(S_A \in B_0) \geq Ck^{-d/2}$$

for some fixed constant C . Applying Theorem 17.1 (rescaling by $\log n$), we conclude that there exists a hyperplane H such that $\text{dist}(v_i, H) \leq 1/\log n$ for at least $n - k$ values of $i = 1, \dots, n$.

Let V' denote the orthogonal projection to H of the vectors v_i with $\text{dist}(v_i, H) \leq 1/\log n$. By conditioning on the signs of all the ξ_i with $\text{dist}(v_i, H) > 1/\log n$, and then projecting the sum X_V onto H , we conclude from (33) the existence of a $d - 1$ -dimensional ball B' in H of radius R such that

$$\mathbf{P}(X_{V'} \in B') \geq (1 + \varepsilon)2^{-n}S(n, s).$$

On the other hand, the vectors in V' have magnitude at least $1 - 1/\log n$. If n is sufficiently large depending on d, ε this contradicts the induction

hypothesis (after rescaling the V' by $1/(1 - 1/\log n)$ and identifying H with \mathbf{R}^{n-1} in some fashion; notice that the scaling changes R slightly but does not change s , and also that the function $2^{-n}S(n, s)$ is decreasing with n). This concludes the proof of (4).

Now we turn to the proof of Conjecture 2.4. We can assume $s \geq 3$, as the remaining cases have already been treated (see Section 2). If the conjecture failed, then there exist arbitrarily large n for which there exist a multi-set $A = \{a_1, \dots, a_n\}$ of vectors in \mathbf{R}^d of length at least 1 and a ball B of radius R such that

$$(34) \quad \mathbf{P}(S_A \in B) > 2^{-n}S(n, s).$$

By iterating the argument used to prove (4), we may find a one-dimensional subspace L of \mathbf{R}^d such that $\text{dist}(v_i, L) \ll 1/\log n$ for at least $n - O(n^{2/3})$ values of $i = 1, \dots, n$. By reordering, we may assume that $\text{dist}(v_i, L) \ll 1/\log n$ for all $1 \leq i \leq n - k$, where $k = O(n^{2/3})$.

Let $\pi : \mathbf{R}^d \rightarrow L$ be the orthogonal projection onto L . We divide into two cases. The first case is when $|\pi(v_i)| > \frac{R}{s}$ for all $1 \leq i \leq n$. We then use the trivial bound

$$\mathbf{P}(S_A \in B) \leq \mathbf{P}(S_{\pi(V)} \in \pi(B)).$$

If we rescale Theorem 2.1 by a factor slightly less than s/R , we see that

$$\mathbf{P}(S_{\pi(V)} \in \pi(B)) \leq 2^{-n}S(n, s)$$

which contradicts (34).

In the second case, we assume $|\pi(v_n)| \leq R/s$. We let A' be the multi-set $\{a_1, \dots, a_{n-k}\}$, then by conditioning on the $\xi_{n-k+1}, \dots, \xi_{n-1}$ we conclude the existence of a unit ball B' such that

$$\mathbf{P}(S_{A'} + \xi_n a_n \in B') \geq \mathbf{P}(S_A \in B).$$

Let $x_{B'}$ be the center of B' . Observe that if $S_{V'} + \xi_n a_n \in B'$ (for any value of ξ_n) then $|S_{\pi(V')} - \pi(x_{B'})| \leq R + \frac{R}{s}$. Furthermore, if $|S_{\pi(V')} - \pi(x_{B'})| > \sqrt{R^2 - 1}$, then the parallelogram law shows that $S_{V'} + a_n$ and $S_{V'-n}$ cannot both lie in B' , and so conditioned on $|S_{\pi(V')} - \pi(x_{B'})| > \sqrt{R^2 - 1}$, the probability that $S_{V'} + \xi_n a_n \in B'$ is at most $1/2$.

We conclude that

$$\begin{aligned} \mathbf{P}(S_{A'} + \xi_n a_n \in B') &\leq \mathbf{P}(|A_{\pi(A')} - \pi(x_{B'})| \leq \sqrt{R^2 - 1}) \\ &\quad + \frac{1}{2} \mathbf{P}\left(\sqrt{R^2 - 1} < |S_{\pi(V')} - \pi(x_{B'})| \leq R + \frac{R}{s}\right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{2} \left(\mathbf{P}(|A_{\pi(A')} - \pi(x_{B'})| \leq \sqrt{R^2 - 1}) \right. \\
 &\quad \left. + \mathbf{P} \left(|S_{\pi(A')} - \pi(x_{B'})| \leq R + \frac{R}{s} \right) \right).
 \end{aligned}$$

However, note that all the elements of $\pi(A')$ have magnitude at least $1 - 1/\log n$. Assume, for a moment, that R satisfies

$$(35) \quad \sqrt{R^2 - 1} < s - 1 \leq R < R + \frac{R}{s} < s.$$

From Theorem 2.1 (rescaled by $(1 - 1/\log n)^{-1}$), we conclude that

$$\mathbf{P}(|S_{\pi(A')} - \pi(x_{B'})| \leq \sqrt{R^2 - 1}) \leq 2^{-(n-k)} S(n - k, s - 1)$$

and

$$\mathbf{P} \left(|\pi(S_{A'}) - \pi(x_{B'})| \leq R + \frac{R}{s} \right) \leq 2^{-(n-k)} S(n - k, s).$$

On the other hand, by Stirling’s formula (if n is sufficiently large) we have

$$\frac{1}{2} (2^{-(n-k)} S(n - k, s - 1)) + \frac{1}{2} 2^{-(n-k)} S(n - k, s) = \sqrt{\frac{2}{\pi} \frac{s - 1/2 + o(1)}{n^{1/2}}}$$

while

$$2^{-n} S(n, s) = \sqrt{\frac{2}{\pi} \frac{s + o(1)}{n^{1/2}}}$$

and so we contradict (34).

An inspection of the above argument shows that all we need on R are the conditions (35). To satisfy the first inequality in (35), we need $R < \sqrt{(s - 1)^2 + 1}$. Moreover, once $s - 1 \leq R < \sqrt{(s - 1)^2 + 1}$, one can easily check that $R + \frac{R}{s} < s$ holds automatically for any $s \geq 3$, concluding the proof.

APPENDIX A. PROOF OF THEOREM 7.6

In this section, we sketch the proof of Theorem 7.6.

Embedding. The first step is to embed the problem into a finite field \mathbf{F}_p for some prime p . In the case when the a_i are integers, we simply take p to be a large prime (for instance $p \geq 2^n \left(\sum_{i=1}^n |a_i| + 1 \right)$ suffices).

If A is a subset of a general torsion-free group G , we rely on the concept of Freiman isomorphism. Two sets A, A' of additive groups G, G' (not necessarily torsion-free) are *Freiman-isomorphism of order k* (in generalized form) if there is a bijective map f from A to A' such that $f(a_1) + \dots + f(a_k) = f(a'_1) + \dots + f(a'_k)$ in G' if and only if $a_1 + \dots + a_k = a'_1 + \dots + a'_k$ in G , for any subsets $\{a_1, \dots, a_k\} \subset A; \{a'_1, \dots, a'_k\} \subset A'$.

The following theorem allows us to pass from an arbitrary torsion-free group to \mathbf{Z} or cyclic groups of prime order (see [67, Lemma 5.25]).

Theorem A.1. *Let A be a finite subset of a torsion-free additive group G . Then for any integer k the following holds.*

- there is a Freiman isomorphism $\phi : A \rightarrow \phi(A)$ of order k to some finite subset $\phi(A)$ of the integers \mathbf{Z} ;
- more generally, there is a map $\phi : A \rightarrow \phi(A)$ to some finite subset $\phi(A)$ of the integers \mathbf{Z} such that

$$a_1 + \dots + a_i = a'_1 + \dots + a'_j \Leftrightarrow \phi(a_1) + \dots + \phi(a_i) = \phi(a'_1) + \dots + \phi(a'_j)$$

for all $i, j \leq k$.

The same is true if we replace \mathbf{Z} by \mathbf{F}_p , if p is sufficiently large depending on A .

Thus instead of working with a subset A of a torsion-free group, it is sufficient to work with subset of \mathbf{F}_p , where p is large enough. From now on, we can assume that a_i are elements of \mathbf{F}_p for some large prime p . We view elements of \mathbf{F}_p as integers between 0 and $p - 1$. We use the short hand ρ to denote $\rho(A)$. The next few steps are motivated by Halász' analysis in [21].

Fourier Analysis. The main advantage of working in \mathbf{F}_p is that one can make use of discrete Fourier analysis. Assume that

$$\rho = \rho(A) = \mathbf{P}(S = a),$$

for some $a \in \mathbf{F}_p$. Using the standard notation $e_p(x)$ for $\exp(2\pi\sqrt{-1}x/p)$, we have

$$(36) \quad \rho = \mathbf{P}(S = a) = \mathbf{E} \frac{1}{p} \sum_{t \in \mathbf{F}_p} e_p(t(S - a)) = \mathbf{E} \frac{1}{p} \sum_{t \in \mathbf{F}_p} e_p(tS) e_p(-ta).$$

By independence

$$(37) \quad \mathbf{E}e_p(tS) = \prod_{i=1}^n e_p(t\xi_i a_i) = \prod_{i=1}^n \cos \frac{2\pi t a_i}{p}.$$

It follows that

$$(38) \quad \rho \leq \frac{1}{p} \sum_{t \in \mathbf{F}_p} \prod_i \left| \cos \frac{2\pi a_i t}{p} \right| = \frac{1}{p} \sum_{t \in \mathbf{F}_p} \prod_i \left| \frac{\cos \pi a_i t}{p} \right|,$$

where we made the change of variable $t \rightarrow t/2$ (in \mathbf{F}_p) to obtain the last identity.

By convexity, we have that $|\sin \pi z| \geq 2\|z\|$ for any $z \in \mathbf{R}$, where $\|z\| := \|z\|_{\mathbf{R}/\mathbf{Z}}$ is the distance of z to the nearest integer. Thus,

$$(39) \quad \left| \cos \frac{\pi x}{p} \right| \leq 1 - \frac{1}{2} \sin^2 \frac{\pi x}{p} \leq 1 - 2 \left\| \frac{x}{p} \right\|^2 \leq \exp \left(-2 \left\| \frac{x}{p} \right\|^2 \right),$$

where in the last inequality we used that fact that $1 - y \leq \exp(-y)$ for any $0 \leq y \leq 1$.

Consequently, we obtain a key inequality

$$(40) \quad \rho \leq \frac{1}{p} \sum_{t \in \mathbf{F}_p} \prod_i \left| \cos \frac{\pi a_i t}{p} \right| \leq \frac{1}{p} \sum_{t \in \mathbf{F}_p} \exp \left(-2 \sum_{i=1}^n \left\| \frac{a_i t}{p} \right\|^2 \right).$$

Large level sets. Now we consider the level sets

$$S_m := \left\{ t \mid \sum_{i=1}^n \|a_i t/p\|^2 \leq m \right\}.$$

We have

$$n^{-C} \leq \rho \leq \frac{1}{p} \sum_{t \in \mathbf{F}_p} \exp \left(-2 \sum_{i=1}^n \left\| \frac{a_i t}{p} \right\|^2 \right) \leq \frac{1}{p} + \frac{1}{p} \sum_{m \geq 1} \exp(-2(m-1)) |S_m|.$$

Since $\sum_{m \geq 1} \exp(-m) < 1$, there must be is a large level set S_m such that

$$(41) \quad |S_m| \exp(-m+2) \geq \rho p.$$

In fact, since $\rho \geq n^{-C}$, we can assume that $m = O(\log n)$.

Double counting and the triangle inequality. By double counting we have

$$\sum_{i=1}^n \sum_{t \in S_m} \left\| \frac{a_i t}{p} \right\|^2 = \sum_{t \in S_m} \sum_{i=1}^n \left\| \frac{a_i t}{p} \right\|^2 \leq m |S_m|.$$

So, for most a_i

$$(42) \quad \sum_{t \in S_m} \left\| \frac{a_i t}{p} \right\|^2 \leq \frac{m}{n'} |S_m|.$$

By averaging, the set of a_i satisfying (42) has size at least $n - n'$. We call this set A' . The set $A \setminus A'$ has size at most n' and this is the exceptional set that appears in Theorem 7.6. In the rest of the proof, we are going to show that A' is a dense subset of a proper GAP.

Since $\|\cdot\|$ is a norm, by the triangle inequality, we have for any $a \in kA'$

$$(43) \quad \sum_{t \in S_m} \left\| \frac{at}{p} \right\|^2 \leq k^2 \frac{m}{n'} |S_m|.$$

More generally, for any $l \leq k$ and $a \in lA'$

$$(44) \quad \sum_{t \in S_m} \left\| \frac{at}{p} \right\|^2 \leq k^2 \frac{m}{n'} |S_m|.$$

Dual sets. Define $S_m^* := \left\{ a \mid \sum_{t \in S_m} \left\| \frac{at}{p} \right\|^2 \leq \frac{1}{200} |S_m| \right\}$ (the constant 200 is ad hoc and any sufficiently large constant would do). S_m^* can be viewed as some sort of a *dual* set of S_m . In fact, one can show as far as cardinality is concerned, it does behave like a dual

$$(45) \quad |S_m^*| \leq \frac{8p}{|S_m|}.$$

To see this, define $T_a := \sum_{t \in S_m} \cos \frac{2\pi at}{p}$. Using the fact that $\cos 2\pi z \geq 1 - 100\|z\|^2$ for any $z \in \mathbf{R}$, we have, for any $a \in S_m^*$

$$T_a \geq \sum_{t \in S_m} \left(1 - 100 \left\| \frac{at}{p} \right\|^2 \right) \geq \frac{1}{2} |S_m|.$$

One the other hand, using the basic identity $\sum_{a \in \mathbf{F}_p} \cos \frac{2\pi ax}{p} = p\mathbf{1}_{x=0}$, we have

$$\sum_{a \in \mathbf{F}_p} T_a^2 \leq 2p|S_m|.$$

(45) follows from the last two estimates and averaging.

Set $k := c_1 \sqrt{\frac{n'}{m}}$, for a properly chosen constant c_1 . By (44) we have $\cup_{l=1}^k lA' \subset S_m^*$. Set $A'' = A' \cup \{0\}$; we have $kA'' \subset S_m^* \cup \{0\}$. This results in the critical bound

$$(46) \quad |kA''| = O\left(\frac{p}{|S_m|}\right) = O(\rho^{-1} \exp(-m + 2)).$$

The role of \mathbf{F}_p is now no longer important, so we can view the a_i as integers. Notice that (46) leads us to a situation similar to that of Freiman’s inverse result (Theorem 7.3). In that theorem, we have a bound on $|2A|$ and conclude that A has a strong additive structure. In the current situation, 2 is replaced by k , which can depend on $|A|$. We can, however, finish the job by applying the following variant of Freiman’s inverse theorem.

Theorem A.2 (Long range inverse theorem, [39]). *Let $\gamma > 0$ be constant. Assume that X is a subset of a torsion-free group such that $0 \in X$ and $|kX| \leq k^\gamma |X|$ for some integer $k \geq 2$ that may depend on $|X|$. Then there is proper symmetric GAP Q of rank $r = O(\gamma)$ and cardinality $O_\gamma(k^{-r} |kX|)$ such that $X \subset Q$.*

One can prove Theorem A.2 by combining Freiman theorem with some extra combinatorial ideas and several facts about GAPs. For full details we refer to [39].

The proof of the continuous version, Theorem 9.2, is similar. Given a real number w and a variable ξ , we define the ξ -norm of w by $\|w\|_\xi := (\mathbf{E}\|w(\xi_1 - \xi_2)\|^2)^{1/2}$, where ξ_1, ξ_2 are two iid copies of ξ . We have the following variant of Lemma 6.2.

$$(47) \quad \rho_{r,\xi}(A) \leq \exp(\pi r^2) \int_{\mathbf{R}^d} \exp\left(-\sum_{i=1}^n \|\langle a_i, z \rangle\|_\xi^2 / 2 - \pi \|z\|_2^2\right) dz.$$

This will play the role of (38) in the previous proof. The next steps are similar and we refer the reader to [39] for more details.

APPENDIX B. PROOF OF THEOREM 10.2

We provide here a proof from [46] (see also [16]). This proof is also influenced by Halász’ analysis from [21]. The starting point is again Esseen’s bound. Applying Lemma 6.2, we obtain

$$(48) \quad \rho_{d,\beta\sqrt{d},\xi}(A) \leq C^d \int_{B(0,\sqrt{d})} \prod_{k=1}^n |\phi(\langle \theta, a_k \rangle / \beta)| d\theta,$$

where ϕ is the characteristic function.

Let ξ' be an independent copy of ξ and denote by $\bar{\xi}$ the symmetric random variable $\xi - \xi'$. Then we easily have $|\phi(t)| \leq \exp(-\frac{1}{2}(1 - \mathbf{E} \cos(2\pi t \bar{\xi}))$.

Conditioning on ξ' , the assumption $\sup_a \mathbf{P}(\xi \in B(a, 1)) \leq 1 - b$ implies that $\mathbf{P}(|\bar{\xi}| \geq 1) \geq b$. Thus,

$$\begin{aligned} 1 - \mathbf{E} \cos(2\pi t \bar{\xi}) &\geq \mathbf{P}(|\bar{\xi}| \geq 1) \cdot \mathbf{E}(1 - \cos(2\pi t \bar{\xi}) \mid |\bar{\xi}| \geq 1) \\ &\geq b \cdot \frac{4}{\pi^2} \mathbf{E}\left(\min_{q \in \mathbf{Z}} |2\pi t \bar{\xi} - 2\pi q|^2 \mid |\bar{\xi}| \geq 1\right) \\ &= 16b \cdot \mathbf{E}\left(\min_{q \in \mathbf{Z}} |t \bar{\xi} - q|^2 \mid |\bar{\xi}| \geq 1\right). \end{aligned}$$

Substituting of this into (48) and using Jensen’s inequality, we get

$$\begin{aligned} &\rho_{d,\beta\sqrt{d},\xi}(A) \\ &\leq C^d \int_{B(0,\sqrt{d})} \exp\left(-8b \mathbf{E}\left(\sum_{k=1}^n \min_{q \in \mathbf{Z}} |\bar{\xi} \langle \theta, \mathbf{a}_k \rangle / \beta - q|^2 \mid |\bar{\xi}| \geq 1\right)\right) d\theta \\ &\leq C^d \mathbf{E}\left(\int_{B(0,\sqrt{d})} \exp\left(-8b \min_{p \in \mathbf{Z}^n} \left\| \frac{\bar{\xi}}{\beta} \theta \cdot \mathbf{a} - p \right\|_2\right) d\theta \mid |\bar{\xi}| \geq 1\right) \\ &\leq C^d \sup_{z \geq 1} \int_{B(0,\sqrt{d})} \exp(-8b f^2(\theta)) d\theta, \end{aligned}$$

where $f(\theta) = \min_{p \in \mathbf{Z}^n} \left\| \frac{z}{\beta} \theta \cdot \mathbf{a} - p \right\|_2$.

The crucial step is to bound the size of the *recurrence set*

$$I(t) := \left\{ \theta \in B(0, \sqrt{d}) : f(\theta) \leq t \right\}.$$

Lemma B.1. *We have*

$$\mu(I(t)) \leq \left(\frac{Ct\beta}{\gamma\sqrt{d}} \right)^d, \quad t < \alpha/2.$$

Proof (of Lemma B.1). Fix $t < \alpha/2$. Consider two points $\theta', \theta'' \in I(t)$. There exist $p', p'' \in \mathbf{Z}^n$ such that

$$\left\| \frac{z}{\beta} \theta' \cdot \mathbf{a} - p' \right\|_2 \leq t, \quad \left\| \frac{z}{\beta} \theta'' \cdot \mathbf{a} - p'' \right\|_2 \leq t.$$

Let

$$\tau := \frac{z}{\beta} (\theta' - \theta''), \quad p := p' - p''.$$

Then, by the triangle inequality,

$$(49) \quad \|\tau \cdot \mathbf{a} - p\|_2 \leq 2t.$$

Recall that by the assumption of the theorem, $\mathbf{LCD}_{\alpha,\gamma}(\mathbf{a}) \geq \frac{\sqrt{d}}{\beta}$. Thus, by the definition of the least common denominator, either $\|\tau\|_2 \geq \frac{\sqrt{d}}{\beta}$ or

$$(50) \quad \|\tau \cdot \mathbf{a} - p\|_2 \geq \min(\gamma\|\tau \cdot \mathbf{a}\|_2, \alpha).$$

In the latter case, since $2t < \alpha$, (49) and (50) imply

$$2t \geq \gamma\|\tau \cdot \mathbf{a}\|_2 \geq \gamma\|\tau\|_2,$$

where the last inequality follows from (14).

Thus we have proved that every pair of points $\theta', \theta'' \in I(t)$ satisfies:

$$\text{either } \|\theta' - \theta''\|_2 \geq \frac{\sqrt{d}}{z} =: R \quad \text{or} \quad \|\theta' - \theta''\|_2 \leq \frac{2t\beta}{\gamma z} =: r.$$

It follows that $I(t)$ can be covered by Euclidean balls of radii r , whose centers are R -separated in the Euclidean distance. Since $I(t) \subset B(0, \sqrt{d})$, the number of such balls is at most

$$\frac{\mu(B(0, \sqrt{d} + R/2))}{\mu(B(0, R/2))} = \left(\frac{2\sqrt{m}}{R} + 1 \right)^d \leq \left(\frac{3\sqrt{d}}{R} \right)^d.$$

Summing these volumes, we obtain $\mu(I(t)) \leq \left(\frac{3Cr}{R} \right)^m$. ■

Proof (of Theorem 10.2). First, by the definition of $I(t)$ and as $\mu(B(0, \sqrt{d}) \leq C^d$, we have

$$(51) \quad \int_{B(0, \sqrt{m}) \setminus I(\alpha/2)} \exp(-8bf^2(\theta)) \, d\theta \leq \int_{B(0, \sqrt{d})} \exp(-2b\alpha^2) \, d\theta \\ \leq C^d \exp(-2b\alpha^2).$$

Second, by using Lemma B.1, we have

$$(52) \quad \int_{I(\alpha/2)} \exp(-8bf^2(\theta)) \, d\theta = \int_0^{\alpha/2} 16bt \exp(-8bt^2) \mu(I(t)) \, dt \\ \leq 16b \left(\frac{C\beta}{\gamma\sqrt{d}} \right)^d \int_0^\infty t^{d+1} \exp(-8bt^2) \, dt \\ \leq \left(\frac{C'\beta}{\gamma\sqrt{b}} \right)^d \sqrt{d} \leq \left(\frac{C''\beta}{\gamma\sqrt{b}} \right)^d.$$

Combining (51) and (52) completes the proof of Theorem 10.2. ■

REFERENCES

- [1] B. Bollobás, *Random Graphs*, Academic Press, New York.
- [2] C. Bordenave and D. Chafai, *Around the circular law*, Probab. Surveys 9 (2012), 1–89.
- [3] J. Bourgain, V. Vu and P. M. Wood, *On the singularity probability of discrete random matrices*, Journal of Functional Analysis 258 (2010), no.2, 559–603.
- [4] A. T. Bharucha-Reid and M. Sambandham, *Random polynomials*, Academic Press, Orlando, 1986.
- [5] E. Breuillard, B. Green and Terence Tao, *The structure of approximate groups*, to appear in Pub. IHES, <http://arxiv.org/abs/1110.5008>.
- [6] D. Conlon, J. Fox and B. Sudakov, *Essays in extremal combinatorics*, submitted, <http://arxiv.org/abs/1212.1300>.
- [7] K. Costello, *Bilinear and quadratic variants on the Littlewood-Offord problem*, to appear in Israel of Mathematics, <http://arxiv.org/abs/0902.1538>.
- [8] K. Costello, T. Tao and V. Vu, *Random symmetric matrices are almost surely non-singular*, Duke Math. J. 135 (2006), 395–413.
- [9] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. 9 (1988), no. 4, 543–560.

- [10] P. Erdős, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. 51 (1945), 898–902.
- [11] P. Erdős and L. Moser, *Elementary Problems and Solutions*, Amer. Math. Monthly, 54 (1947), no. 4, 229–230.
- [12] C. G. Esséen, *On the Kolmogorov-Rogozin inequality for the concentration function*, Z. Wahrsch. Verw. Gebiete 5 (1966), 210–216.
- [13] W. Feller, *An introduction to probability and its applications*, Wiley series in probability and mathematical statistics.
- [14] P. Frankl and Z. Füredi, *Solution of the Littlewood-Offord problem in high dimensions*, Ann. of Math. (2) 128 (1988), no. 2, 259–270.
- [15] G. Freiman, *Foundations of a Structural Theory of Set Addition*, Translations of Mathematical Monographs 37, Amer. Math. Soc., Providence, RI, USA, 1973.
- [16] O. Friedland and S. Sodin, *Bounds on the concentration function in terms of Diophantine approximation*, C. R. Math. Acad. Sci. Paris 345 (2007), no. 9, 513–518.
- [17] H. Goldstine and J. von Neumann, *Numerical inverting of matrices of high order*, Bull. Amer. Math. Soc. 53 (1947), 1021–1099.
- [18] F. Götze and A. Tikhomirov, *The circular law for random matrices*, Ann. Probab. 38 (2010), no. 4, 1444–1491.
- [19] J. Griggs, *The Littlewood-Offord problem: tightest packing and an M -part Sperner theorem*, Europ. J. Combin. 1 (1980), 225–234.
- [20] D. S. Gunderson, V. Rödl and A. Sidorenko, *Extremal problems for sets forming Boolean algebras and complete partite hypergraphs*, J. Combin. Theory Ser. A 88 (1999), 342–367.
- [21] G. Halász, *Estimates for the concentration function of combinatorial number theory and probability*, Period. Math. Hungar. 8 (1977), no. 3–4, 197–211.
- [22] D. Hilbert, *Über die Irreduzibilität ganzer rationaler Funktionen mit ganzzahligen Koeffizienten*, J. Reine Angew. Math. 110 (1892), 104–129.
- [23] J. Kahn, J. Komlós and E. Szemerédi, *On the probability that a random ± 1 matrix is singular*, J. Amer. Math. Soc. 8 (1995), 223–240.
- [24] G. Katona, *On a conjecture of Erdős and a stronger form of Sperner’s theorem*, Studia Sci. Math. Hungar 1 (1966), 59–63.
- [25] D. Kleitman, *On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors*, Advances in Math. 5 1970 155–157 (1970).
- [26] D. Kleitman, *Some new results on the Littlewood-Offord problem*, J. Combinatorial Theory Ser. A 20 (1976), no. 1, 89–113.
- [27] D. Kleitman, *On a lemma of Littlewood and Offord on the distribution of certain sums*, Math. Z. 90 1965 251–259.
- [28] J. Komlós, *On the determinant of $(0, 1)$ matrices*, Studia Sci. Math. Hungar. 2 (1967), 7–22.
- [29] J. Komlós, *On the determinant of random matrices*, Studia Sci. Math. Hungar. 3 (1968), 387–399.
- [30] A. Kolmogorov, *Two uniform limit theorems for sums of independent random variables*, Theor. Probab. Appl. 1 (1956), 384–394.
- [31] A. Kolmogorov, *Sur les propriétés des fonctions de concentrations de M. P. Lévy*, Ann. Inst. H. Poincaré 16 (1958), 27–34.

- [32] G. Kozma and O. Zeitouni, *On common roots of random Bernoulli polynomials*, to appear in Int. Math. Res. Not., <http://arxiv.org/abs/1109.2316>.
- [33] J. E. Littlewood and A. C. Offord, *On the number of real roots of a random algebraic equation*. III. Rec. Math. Mat. Sbornik N.S. 12, (1943). 277–286.
- [34] S. Muroga, I. Toda, and S. Takasu, *Theory of majority decision elements*, J. Franklin Inst., 271, 376–418, 1961.
- [35] S. Muroga, *Threshold logic and its applications*, Wiley-Interscience, New York, 1971.
- [36] H. Nguyen, *Inverse Littlewood-Offord problems and the singularity of random symmetric matrices*, Duke Mathematics Journal Vol. 161, 4 (2012), 545–586.
- [37] H. Nguyen, *A new approach to an old problem of Erdős and Moser*, Journal of Combinatorial Theory, Series A 119 (2012) 977–993.
- [38] H. Nguyen, *Singularity of random combinatorial matrices*, to appear in SIAM J. Discrete Mathematics, <http://arxiv.org/abs/1112.0753>.
- [39] H. Nguyen and V. Vu, *Optimal Littlewood-Offord theorems*, Advances in Math., Vol. 226 6 (2011), 5298–5319.
- [40] A. Pajor and L. Pastur, *On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution*, Studia Math. 195 (2009), no. 1, 11–29.
- [41] R. A. Proctor, *Solution of two difficult combinatorial problems with linear algebra*, Amer. Math. Monthly 89 (1982), no. 10, 721–734.
- [42] A. Razborov and E. Viola, *Real Advantage*, submitted, <http://eccc.hpi-web.de/report/2012/134/>.
- [43] B. A. Rogozin, *An estimate for concentration functions*, Theor. Probab. Appl. 6 (1961), 94–97.
- [44] M. Rudelson, *Invertibility of random matrices: Norm of the inverse*, Annals of Mathematics, 168 (2008), no. 2, 575–600.
- [45] M. Rudelson and R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, Advances in Mathematics 218 (2008), 600–633.
- [46] M. Rudelson and R. Vershynin, *Smallest singular value of a random rectangular matrix*, Communications on Pure and Applied Mathematics 62 (2009), 1707–1739.
- [47] M. Rudelson and R. Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, Proceedings of the International Congress of Mathematicians. Volume III, 1576–1602, Hindustan Book Agency, New Delhi, 2010.
- [48] A. Sali, *Strong form of an M -part Sperner theorem*, European J. Combinatorics 4 (1983), 179–183.
- [49] A. Sali, *A Sperner type theorem*, Order 2 (1985), 13–127.
- [50] A. Sárközy and E. Szemerédi, *Über ein Problem von Erdős und Moser*, Acta Arithmetica 11 (1965), 205–208.
- [51] A. A. Sherstov, *Communication lower bounds using dual polynomials*, Bulletin of the EATCS, 95, 59–93, 2008.
- [52] D. A. Spielman and S. H. Teng, *Smoothed analysis of algorithms*, Proceedings of the International Congress of Mathematicians, Vol. I, 597–606, Higher Ed. Press, Beijing, 2002.
- [53] D. A. Spielman and S. H. Teng, *Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time*, J. ACM 51 (2004), no. 3, 385–463.

- [54] R. Stanley, *Weyl groups, the hard Lefschetz theorem, and the Sperner property*, SIAM J. Algebraic Discrete Methods 1 (1980), no. 2, 168–184.
- [55] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. 20 (1969), 199–245.
- [56] T. Tao and V. Vu, *On random ± 1 matrices: singularity and determinant*, Random Structures Algorithms 28 (2006), 1–23.
- [57] T. Tao and V. Vu, *On the singularity probability of random Bernoulli matrices*, Journal of the A. M. S 20 (2007), 603–673.
- [58] T. Tao and V. Vu, *Random matrices: The Circular Law*, Communication in Contemporary Mathematics 10 (2008), 261–307.
- [59] T. Tao and V. Vu, *From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices*, Bull. Amer. Math. Soc. (N.S.) 46 (2009), no. 3, 377–396.
- [60] T. Tao and V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random matrices*, Annals of Mathematics (2) 169 (2009), no. 2, 595–632.
- [61] T. Tao and V. Vu, *On the permanent of random Bernoulli matrices*, Adv. Math. 220 (2009), 657–669.
- [62] T. Tao and V. Vu, *A sharp inverse Littlewood-Offord theorem*, Random Structures Algorithms 37 (2010), no. 4, 525–539.
- [63] T. Tao and V. Vu, *Smooth analysis of the condition number and the least singular value*, Mathematics of Computation 79 (2010), 2333–2352.
- [64] T. Tao and V. Vu, *Random matrices: the distribution of the smallest singular values*, Geom. Funct. Anal. 20 (2010), no. 1, 260–297.
- [65] T. Tao and V. Vu, *Random matrices: universality of ESDs and the circular law*, Ann. Probab. 38 (2010), no. 5p. 2023–2065, with an appendix by M. Krishnapur.
- [66] T. Tao and V. Vu, *The Littlewood-Offord problem in high dimensions and a conjecture of Frankl and Füredi*, Combinatorica 32 (2012), no. 3, 363–372.
- [67] T. Tao and V. Vu, *Additive Combinatorics*, Cambridge Univ. Press, 2006.
- [68] L. G. Valiant, *Graph-theoretic arguments in low-level complexity*, In Proceedings of the 6th MFCS, Lecture Notes in Computer Science, 53, p. 162–176, New York/Berlin, 1977, Springer-Verlag.
- [69] R. Vershynin, *Invertibility of symmetric random matrices*, to appear in Random Structures and Algorithms, <http://arxiv.org/abs/1102.0300>.
- [70] E. Viola, *On the power of small depth computation*, Foundations and Trends in Theoretical Computer Science, 5(1), 1–72, 2009.

Hoi H. Nguyen

*Department of Mathematics,
Yale University,
10 Hillhouse Ave.,
New Haven,
CT 06511*

e-mail: hoi.nguyen@yale.edu

Van H. Vu

*Department of Mathematics,
Yale University,
10 Hillhouse Ave.,
New Haven,
CT 06511*

e-mail: van.vu@yale.edu

THE BEGINNINGS OF GEOMETRIC GRAPH THEORY

JÁNOS PACH*

*“... to ask the right question and
to ask it of the right person.”
(Richard Guy)*

Geometric graphs (topological graphs) are graphs drawn in the plane with possibly crossing straight-line edges (resp., curvilinear edges). Starting with a problem of Heinz Hopf and Erika Pannwitz from 1934 and a seminal paper of Paul Erdős from 1946, we give a biased survey of Turán-type questions in the theory of geometric and topological graphs. What is the maximum number of edges that a geometric or topological graph of n vertices can have if it contains no forbidden subconfiguration of a certain type? We put special emphasis on open problems raised by Erdős or directly motivated by his work.

1. INTRODUCTION

The term “geometric graph theory” is often used to refer to a large, amorphous body of research related to graphs defined by geometric means. Here we take a narrower view: by a *geometric graph* we mean a graph G drawn in the *plane* with possibly intersecting straight-line edges. If the edges are allowed to be arbitrary continuous curves connecting the vertices (points), then G is called a *topological graph*. Disregarding the particular way the graph is drawn, we obtain the “abstract” underlying graph of G , which is usually also denoted by G . We use the term *geometric graph theory* as a short form for “the theory of geometric and topological graphs.”

*Supported by NSF Grant CCF-08-30272, by OTKA under EUROGIGA projects GraDR and ComPoSe 10-EuroGIGA-OP-003, and by Swiss National Science Foundation Grants 200020-144531 and 200021-137574.

In the past few decades, a number of exciting discoveries have been made in this field. Some of them have found interesting applications in graph drawing, in combinatorial and computational geometry, in additive number theory, and elsewhere. See, e.g., [5], [68], [95], [101], [27]. Many related contributions can be found in the proceedings of the annual symposia on graph drawing, published in Springer's Lecture Notes series in Computer Science (for instance, in [64]) and in two collections of papers [78], [79]. For surveys, see Chapter 14 in [80], Chapter 10 in [51], and Chapters 1 and 3 in [42].

Paul Erdős had a profound influence on the subject. On the occasion of his 100th birthday, we review the beginnings of geometric graph theory in the 1930s and 40s, which were also formative years in Erdős's personal and mathematical life. We use this as a starting point to give a short and biased survey of some research directions that can be traced back more or less directly to these early developments. We put special emphasis on open problems raised by Erdős and others, which had a large impact on the evolution of geometric graph theory.

2. A PROBLEM IN JAHRESBERICHT – GERMAN MATHEMATICS

In 1934, *Heinz Hopf* and *Erika Pannwitz*, Hopf's student at Friedrich Wilhelms University (today Humboldt University) in Berlin, posed the following problem in the problem section of *Jahresbericht der Deutschen Mathematiker-Vereinigung*.

Problem 1 [57]. Let $p_0, p_1, \dots, p_{n-1}, p_n = p_0$ be n distinct points in the plane such that the distance conditions

$$\begin{aligned} d(p_i, p_j) &\leq 1 \quad (0 \leq i < j < n), \\ d(p_i, p_{i+1}) &= 1 \quad (i = 0, \dots, n-1) \end{aligned}$$

are satisfied. Prove that this is possible if and only if n is odd or $n = 2$.

Three solutions were subsequently published in 1935: by *W. Fenchel* (Copenhagen), by *J. W. Sutherland* (Cambridge) [43], and in the next issue of the journal, by *H. Baron* (Berlin) [9]. Other correct solutions were submitted by *A. E. Mayer* (Wien), *H. Baer* (Frankfurt a. M.), *L. Ehrlich* (Berlin), *J. Fox* (Brooklyn), *R. Frucht* (Triest), *L. Goeritz* (Rostock), *F. Gruber* (Vienna), *J. Juilfs* (Berlin), *R. Lauffer* (Graz), *E. Linés Escardó* (Madrid), *B. Neumann* (Cambridge), *L. Rédei* (Mezőtúr), *L. A. Santaló* (Madrid), *P. Scherk* (Göttingen), and *W. Schulz* (Berlin).

The “Annual Reports” of the German Mathematical Society were published, of course, in German. However, many solutions and articles were sent by mathematicians from other, non German speaking countries, mostly from Europe and from the United States. In the 1930s, German universities played a leading role in mathematics. From all over the world, many young talents (like Fox, Rédei, and Santaló) came to study in Berlin, München, Hamburg, Göttingen, and elsewhere. At the 1936 International Congress of Mathematicians held in Oslo, half of the plenary lectures were delivered in German [77]. When after a 14-year recess due to the war the next congress was held at Harvard University, only one of the 21 main lectures had a German title: it was the talk of Hopf, one of the original proposers of Problem 1. However, this time he did not arrive from Berlin, he was Professor at ETH Zürich. Fenchel, Frucht, Neumann, and Santaló had also fled Germany and built distinguished academic careers in Copenhagen, Valparaiso, Canberra, and Buenos Aires. They became leading experts in convexity, graph theory, group theory, and integral geometry. The lives of many of those who stayed in Germany were sidetracked: Pannwitz worked for the German Cryptography Service during the war and Juilfs became an SS Obersturmsführer. Between 1944 and 1951 the publication of *Jahresbericht* was halted.

Fenchel’s elegant solution to Problem 1 was based on the following observation [43]. Connect two points, p_i and p_j , by a segment if their distance is equal to the diameter of the point set $P = \{p_0, \dots, p_{n-1}\}$ (which is, in our case, equal to 1). The resulting geometric graph is called the *diameter graph* (or the graph of diameters) associated with P . It follows from the triangle inequality that any two edges of the diameter graph either share an endpoint or cross each other. Suppose now that $n > 2$ and that P satisfies the properties in Problem 1. Since the diameter graph has no two disjoint edges, the segments p_0p_1 and p_2p_3 must lie in the same half-plane bounded by the line p_1p_2 . Thus, p_0 and p_3 lie in the same half-plane. For the same reason, all edges $p_3p_4, p_4p_5, \dots, p_{n-1}p_0$ must cross the line p_1p_2 , hence the elements of the sequence $p_3, p_4, \dots, p_n = p_0$ lie on alternating sides of the line p_1p_2 . This is possible only if n is odd.

3. A PAPER IN THE MONTHLY – PAUL ERDŐS ENTERS THE SCENE

Erdős was one of the most successful problem solvers of *Középiskolai Matematikai Lapok*, an excellent Hungarian journal for high school students, founded in 1893. He had a lifelong passion for mathematical puzzles and spoke fluent German. In 1934, the same year, when the Hopf-Pannwitz

problem appeared, Erdős received his doctorate at Péter Pázmány University (today Loránd Eötvös University), Budapest. Because of the increasingly anti-semitic atmosphere in Hungary, he accepted a fellowship arranged by *Louis J. Mordell*, and moved first to Manchester and four years later to Princeton. He had access to the *Jahresbericht*, and it is almost certain that he came across Problem 1 shortly after it was published. We will see in the sequel that it inspired him to create a whole new area of research in discrete geometry.

The argument of Fenchel described in the previous section can be easily modified to yield the following statement. It first appeared in a classic paper of Erdős [29] published in the *American Mathematical Monthly* in 1946. He generously attributed the result to Hopf and Pannwitz, although in this form it does not appear in [57]: it was first formulated by him.

Theorem 2 [29]. *The number of edges of the graph of diameters induced by a set of n points in the plane is at most n . This bound can be attained for every $n > 2$.*

In the same paper, Erdős quoted Andrew Vázsonyi's conjecture from the mid-1930s (see also [31]), according to which the number of times the diameter (the maximum distance) can occur among n points in 3-space is at most $2n - 2$. This statement was proved independently by Grünbaum [52], Heppes [54], and Straszewicz [97]. All of these proofs used the notion of *ball polytopes*, that is, convex bodies obtained by taking the intersection of balls of equal radii. However, as was pointed out by Kupitz, Martini, and Perles [66], ball polytopes have some unpleasant features different from the properties of convex polytopes. In particular, their edge-skeletons need not be 3-connected. Therefore, making the above proofs precise requires a lengthy analysis. Half a century later, simpler proofs were found by Perlstein and Pinchasi [92] and by Swanepoel [99].

Theorem 3 ([52], [54], [97]). *The number of edges of the graph of diameters induced by a set of n points in 3-dimensional space is at most $2n - 2$. This bound can be attained for every $n > 3$.*

Erdős [29] also remarked that this statement has an interesting geometric corollary.

Corollary 4. *Every (finite) set of points in 3-dimensional space can be decomposed into 4 sets of smaller diameter.*

Indeed, it follows from Theorem 3 that the diameter graph associated with any finite set of points has a vertex of degree at most 3. Removing such a vertex, one can show by induction that the chromatic number of the

diameter graph is at most 4. This is equivalent to Corollary 4. See also [26] and [55].

Corollary 4 is the $d = 3$ special case of Borsuk's conjecture [12] which states that any d -dimensional set of points can be decomposed into $d + 1$ sets of smaller diameter. In 1993, Kahn and Kalai [59] (see also [76]) disproved Borsuk's conjecture for large values of d . Today the conjecture is known to fail in all dimensions $d \geq 65$. See [56] and [93] for a survey and [11] for a recent improvement.

As was reported by Erdős [32], a simple construction due to Lenz (1955) shows that, for a fixed $d \geq 4$, the number of times the diameter can occur among n points in d -dimensional space can grow quadratically in n . Indeed, let $k = \lfloor d/2 \rfloor$, and take k concentric unit circles in \mathbb{R}^d , in pairwise orthogonal planes. On each of these circles, pick $\lfloor n/k \rfloor$ or $\lceil n/k \rceil$ points very close to each other, so that their total number is n . The diameter of the resulting point set is $\sqrt{2}$, and the distance $\sqrt{2}$ occurs $\frac{1}{2}(1 - \frac{1}{k} + o(1))n^2$ times. Using the Erdős-Stone theorem [41], a cornerstone of extremal graph theory, Erdős proved that this construction is asymptotically best possible.

Theorem 5 [32]. *For a fixed $d \geq 4$, the maximum number of edges of the diameter graph of a set of n points in d -dimensional space is*

$$\frac{1}{2} \left(1 - \frac{1}{\lfloor d/2 \rfloor} + o(1) \right) n^2.$$

Erdős suggested that instead of estimating the number of occurrences of the largest distance, one can also investigate the frequency of the 2nd largest, 3rd largest, etc. distances determined by a set of n points. In particular, it was shown by Vesztergombi [111] (see also [38]) that the i -th largest distance among n points in the plane cannot occur more than $2in$ times. Morić and Pach [73] showed that for a fixed i , the number of times the i -th largest distance can occur among n points in 3-dimensional space is $O(n)$. The constant provided by the proof, hidden in the big- O notation, grows exponentially in i , which can probably be much improved. The nature of the problem again changes in dimension d larger than 3: the i -th largest distance can occur $\Omega(n^2)$ times.

Perhaps the most important contribution of Erdős's paper [29] in the *Monthly* was that he modified the Hopf-Pannwitz problem, as follows. Let $f_d(n)$ denote the the maximum number of times that *any* distance can occur among n points in d -dimensional space. Erdős [32] proved that for any $d \geq 4$, $f_d(n)$ is asymptotically equal to the maximum number of occurrences of the *diameter*, given in Theorem 5. The exact value of $f_4(n)$ for every n was determined by Brass [14]. Swanepoel [100] extended this result to every

even $d \geq 4$, provided that n is sufficiently large depending on d . He also found the maximum number of times the diameter can occur among n points in d -dimensional space, for every $d \geq 4$ and for all sufficiently large n . For some other extensions of these results, see [39] and [7].

The asymptotic behavior of the functions $f_2(n)$ and $f_3(n)$ is still a mystery. Erdős [29] proved that $f_2(n) > n^{1+c/\log \log n}$ for a suitable constant $c > 0$, and conjectured that this bound is not far from being tight. However, the best known upper bound is still $f_2(n) = O(n^{4/3})$, which was established by Spencer, Szemerédi, and Trotter [96] thirty years ago. For alternative proofs, see [22], [101], and [88]. In 3-dimensional space, we have

$$cn^{4/3} \log \log n < f_3(n) < n^{3/2},$$

where $c > 0$ is a constant and $\alpha(n)$ is an extremely slowly growing function, closely related to the inverse of Ackermann's function. The lower and upper bounds were proved in [32] and [22], respectively. (With no danger of confusion, in different formulas we use the same letter c to denote different unrelated constants.)

Obviously, the number of distinct distances determined by n points in the plane is at least $\binom{n}{2}/f_2(n) > cn^{2/3}$. "Though I have thought to improve this result for many years – wrote Erdős in [29] – I have not been able to do so." After many small improvements ([75], [20], [21], [95], [102], [60], [61]), 65 years later Guth and Katz [53] got very close to verifying Erdős's conjecture:

Conjecture 6 (Erdős [29]). *The number of distinct distances determined by n points in the plane is at least $cn/\sqrt{\log n}$, for a suitable constant $c > 0$.*

If true, the order of magnitude of this bound cannot be improved, as shown by a $\sqrt{n} \times \sqrt{n}$ piece of the integer grid. In their breakthrough paper, using a framework set up by Elekes [28], Guth and Katz have established a $cn/\log n$ lower bound. In fact, Erdős [31], [33], [34], [35] also made a stronger conjecture, stating that any set of n points in the plane has an element from which there are at least $cn/\sqrt{\log n}$ distinct distances to the other points. It does not seem to be an easy task to adapt the Guth-Katz proof to estimate this quantity. So far the best lower bound is $cn^{0.864\dots}$, due to Katz and Tardos [61].

We close this section by another possible generalization of Theorem 2 to higher dimensions, different from Theorems 3 and 5.

Conjecture 7 (Z. Schur [94]). *For any positive integers d and n ($n > d$), the graph of diameters induced by a set of n points in d -dimensional space contains at most n complete subgraphs with d vertices.*

For $d = 3$, Schur's conjecture has been proved by Schur, Perles, Martini, and Kupitz [94]. In [74], it was shown Conjecture 7 would follow from the following statement.

Conjecture 8 [74]. *For any positive integers d and n ($n > d > 2$), any two complete subgraphs of size d of the graph of diameters induced by a set of n points in d -dimensional space share at least $d - 2$ vertices.*

For $d = 3$, Conjecture 8 is true. In fact, Dolnikov [25] proved the stronger statement that the graph of diameters of a 3-dimensional point set contains no two disjoint odd cycles. For larger values of d , we have been unable to verify even the weaker conjecture that the graph of diameters contains no two vertex-disjoint cliques of size d .

For more results and open problems related to the subject of this section, see [15] and [40].

4. DROPPING THE METRIC RESTRICTIONS – GEOMETRIC GRAPHS

Fenchel's solution [43] for the Hopf-Pannwitz problem (Problem 1) can be easily modified to establish a statement, a bit stronger than Theorem 2. Recall that a *geometric graph* G is a graph drawn in the plane by possibly crossing straight line edges. For simplicity, we assume throughout that no 3 vertices (points) of G are collinear. An *edge* of G is a *closed* segment connecting a pair of vertices. Therefore, the condition that no 2 edges are disjoint is equivalent to saying that any pair of edges share either an endpoint or an interior point. Of course, they cannot share more than one point, because of the assumption that no 3 vertices are collinear.

Theorem 9 (Erdős, Avital-Hanani [8], Kupitz [65], Perles). *Every geometric graph of n vertices that does not contain 2 disjoint edges has at most n edges. This bound can be attained for every $n > 2$.*

This statement first appeared in print as Problem 3 at the end of a paper written by Shmuel Avital and Haim Hanani [8], which was published in *Gilyonot Le'matematika*, an Israeli journal for high school students and amateurs, edited by Joseph Gillis at Weizmann Institute, Rehovot. It is very likely that the authors heard the question from Paul Erdős. After being banned from entering the United States for 9 years, as an "undesirable alien," in 1955 Erdős was appointed a "Permanent Visiting Professor" at Technion, Haifa. Every year he spent at least one month in Israel, and Hanani was one of his close friends and collaborators.

When Micha Perles (Hebrew University) was told about Theorem 9 roughly ten years after the publication of the Avital-Hanani paper, he found the following “proof from the Book:” Suppose that there is a spider sitting at each vertex v of the graph (web). It looks around and if it finds an edge e incident to v with the property that within the next 180-degree range in the clockwise direction there is no other edge, it walks to the middle of e and lays an egg. Otherwise, the spider stays at v and does not lay an egg. Notice that if G has no 2 disjoint edges, there will be no edge left without an egg. Therefore, the number of edges cannot exceed the number of spiders. Inspired by Perles, Yaakov Kupitz fully characterized all geometric graphs and point configurations for which equality holds in Theorems 9 and 2. (See also [67].) He has also found some interesting generalizations of Theorem 9, and these results constituted his master thesis [65].

It is a natural question to ask whether Theorem 9 can be generalized to topological graphs, that is, to graphs G drawn in the plane by possibly crossing curvilinear edges. It is clear that we need some additional assumptions on G , because it is easy to draw a complete topological graph in which every pair of edges intersect. We call a topological graph *simple* if every pair of edges have at most one point in common, which is either a common endpoint or a proper crossing. Two edges are not allowed to touch each other.

In the late 1960s, independently of the above developments, John Conway defined a *thrackle* as a simple topological graph, in which every pair of edges share precisely one point: an endpoint or a proper crossing. This term may have been first used in a commercial: fishermen referred to their entangled nets as being thrackled.

Conjecture 10 (Conway’s thrackle conjecture [114]). *Every thrackle of n vertices has at most n edges. This bound can be attained for every $n > 2$.*

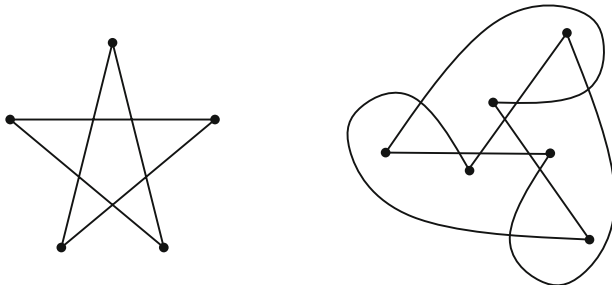


Fig. 1. C_5 and C_6 drawn as thrackles

The first linear upper bound on the number of edges of a thrackle of n vertices was established in [69]. It was improved by Cairns and Nikolayevsky [16]. The best known upper bound, $1.428n$, was proved in [48]. Apart from the case of straight-line thrackles (Theorem 9), Conway's conjecture is known to be true for *x-monotone* thrackles (for which any vertical line intersects every edge in at most one point) [87] and for *outerplanar* thrackles (whose vertices lie on a circle and all edges in its interior) [17]. Perhaps the next step would be to verify the conjecture for thrackles in which every edge is the union of at most 2 (or at most a bounded number of) *x-monotone* pieces.

Avital and Hanani [8] asked the question that at most how many edges can a geometric graph of n vertices have if it contains no k pairwise disjoint edges. For *convex geometric graphs*, that is, for geometric graphs whose vertices lie on a closed convex curve, Kupitz [65] proved that this maximum is equal to $(k - 1)n$, for all $n > 2(k - 1)$. For arbitrary geometric graphs, in the special case $k = 3$, the first linear upper bound (of roughly $6n$) was established by Alon and Erdős [6]. It was subsequently improved by O'Donnell and Perles (unpublished) and by Goddard, Katchalski, and Kleitman [50]. The following asymptotically tight bound was found by Černý [19].

Theorem 11 (Černý [19]). *Every geometric graph of n vertices which does not contain 3 disjoint edges has at most $2.5n$ edges. This bound is tight up to an additive constant.*

For larger values of k , the first linear upper bound, $O(k^4n)$, for the number of edges of a geometric graph G with no k disjoint edges was given by Pach and Törőcsik [90]. After an initial improvement by G. Tóth and Valtr [107], Tóth [106] established the upper bound $|E(G)| \leq O(k^2n)$; see also [112]. The following conjecture is perhaps too optimistic.

Conjecture 12. *The maximum number of edges of a geometric graph of n vertices that contains no k disjoint edges is $O(kn)$.*

It is perfectly possible that this conjecture remains true for simple topological graphs. However, in this case, even for $k = 3$, we do not have a linear upper bound in n on the number of edges. All we know is that, according to [89], the maximum number of edges of a simple topological graph with n vertices that contains no k disjoint edges is $n(\log n)^{O(k)}$. In particular, it follows that a *complete* simple topological graph with n vertices has $\Omega(\frac{\log n}{\log \log n})$ pairwise disjoint edges. Fox and Sudakov [47] improved this bound to $\Omega(\log^{1+\varepsilon} n)$, for a suitable $\varepsilon > 0$. Presently, the best known result in this direction is due to Suk [98].

Theorem 13 (Suk [98]). *Every complete simple topological graph of n vertices has $\Omega(n^{1/3})$ disjoint edges.*

An alternative proof of this bound was found by Fulek and Ruiz-Vargas [49]. If the strengthening of Conjecture 12 to all simple topological graphs is true, it immediately implies

Conjecture 14. *Every complete simple topological graph of n vertices has $\Omega(n)$ disjoint edges.*

For geometric graphs G (in fact, for topological graphs drawn with x -monotone edges), Conjecture 14 is obviously true. Ordering the vertices with respect to their x -coordinates and taking all edges between consecutive vertices, we obtain a non-selfintersecting Hamilton path in G . Taking every other edge of this path, we get a set of $\lfloor n/2 \rfloor$ pairwise disjoint edges. As far as I know, for complete simple topological graphs we do not have any lower bound for the size of the longest non-selfintersecting path, comparable to the one given by Suk's theorem (Theorem 13). The best bound I am aware of is $\Omega(\log^{1/6} n)$; see [86].

Conjecture 15. *There exists $\varepsilon > 0$ such that every complete simple topological graph on n vertices has a non-selfintersecting path of length at least n^ε .*

No example is known in which the size of the longest non-selfintersecting path is $o(n)$.

5. RELAXATIONS OF PLANARITY

For more than two decades starting from the 1940s, one of Erdős' contemporaries, György Hajós, made persistent efforts to settle the 4-color conjecture for planar graphs. He conjectured that every graph of chromatic number k contains a subdivision ("topological subgraph") of a complete graph with k vertices. For $k = 5$, this would of course imply the 4-color theorem. Unfortunately, we still do not know if Hajós' conjecture is true in this case. However, for $k \geq 7$, the conjecture was disproved by Catlin [18], and shortly after Erdős and Fajtlowicz [36] discovered that the conjecture combined with Turán's theorem [108] would imply that every graph G with at least constant times k^3 vertices has k vertices that induce either a complete subgraph or an empty subgraph in G . (See also [105].) However, in his classic note [30] written 30 years earlier, Erdős used the "probabilistic

method” to prove the existence of graphs with $2^{k/2}$ vertices that do not have this property.

Nevertheless, a result much weaker than Hajós’ conjecture, first proposed in the doctoral dissertation of Rudolf Halin, turned out to be true. Dirac [24] and Jung [58] observed that an idea of Wagner [113] can be used to establish the existence of a function $f(k)$ with the property that every graph with chromatic number at least $f(k)$ contains a subdivision of a complete graph K_k with k vertices. Surprisingly, Mader [70] found a much stronger result with a much simpler proof: There also exists a function $g(k)$ such that every graph of n vertices and more than $g(k)n$ edges contains a subdivision of K_k . (Every graph of chromatic number $f(k)$ contains a subgraph in which every vertex has degree at least $f(k) - 1$.) The correct order of magnitude of the function $g(k)$ was determined 30 years later by Komlós and Szemerédi [63] and by Bollobás and Thomason [10]: $g(k) = \Theta(k^2)$. This settled a conjecture of Erdős and Hajnal [37] and Mader [70]. Another famous result of this kind was conjectured by Dirac [23].

Theorem 16 (Mader [71]). *For every $n \geq 3$, the maximum number of edges that a graph with n vertices can have without containing a subdivision of K_5 is $3n - 6$.*

The above statements are usually discussed in the framework of “topological graph theory” (see [72]). They do not depend on the particular drawing of G . They describe “global” properties of graphs G with more edges than how many planar graphs can have, and one does not have much control of the size of the forced subdivisions. In what follows, we would like to discuss some problems related to “local” properties of geometric or topological graphs.

By Euler’s theorem, if a geometric or topological graph G has more than $3n - 6$ edges, two of its edges must cross each other. (A *crossing* occurs when two edges share a common interior point.) In fact, if G has much more than $3n - 6$ edges, the number of crossings increases dramatically. Erdős and Guy conjectured, and Ajtai, Chvátal, Newborn, Szemerédi [5] and, independently, Leighton [68] proved that, if the number of edges, e , satisfies $e > 3n - 6$, there are at ce^3/n^2 crossings, where c is a suitable positive constant. The best known value of the constant $c > \frac{1024}{31827} > 0.032$ was found in [84].

What happens if, instead of a crossing pair of edges, we want to guarantee the existence of some larger configurations involving several crossings? What kind of *unavoidable* substructures must occur in every geometric or topological graph G having n vertices and more than Cn edges, for an appropriately large constant $C > 0$?

A geometric or topological graph is called k -quasiplanar if it contains no k pairwise crossing edges.

Conjecture 17. *For any positive integer k , there is a constant C_k such that the number of edges of any k -quasiplanar topological graph with n vertices is at most $C_k n$.*

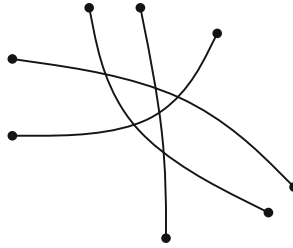


Fig. 2. Four pairwise crossing edges in a topological graph

For $k = 3$, for simple topological graphs (i.e., where every pair of edges cross at most once), Conjecture 17 was proved in [4]. Without the simplicity condition, the statement was first proved in [83]. The best known upper bound of roughly $8n$ was established by Ackerman and Tardos [3], who also proved that the maximum number of edges that a simple 3-quasiplanar topological graph can have is $6.5n - O(1)$. For $k = 4$, the conjecture has been verified by Ackerman [1].

For larger values of k , Conjecture 17 is still open. The upper bound $n(\log n)^{O(k)}$ for the number of edges of a simple k -quasiplanar topological graph was first proved in [85], and then for all k -quasiplanar topological graphs in [83]. This was further improved to $n(\log n)^{O(\log k)}$ by Fox and Pach [44]. For simple topological graphs, presently the best known upper bound is $(n \log n)\alpha_k(n)$, where $\alpha_k(n)$ denotes an extremely slowly growing function related to the inverse of the Ackermann function. It was established in [45]. For k -quasiplanar geometric graphs and, more generally, for simple topological graphs whose edges are represented by x -monotone arcs, Valtr [109], [110] showed that the number of edges cannot exceed $c_k n \log n$. Extending Valtr's ideas, Fox, Pach, and Suk proved the following.

Theorem 18 [45]. *The number of edges of a k -quasiplanar topological graph with n vertices, the edges of which are represented by x -monotone arcs, is at most $2^{c k^6} n \log n$, for a suitable absolute constant c .*

Erdős raised the question whether every system of continuous arcs in the plane with no k pairwise intersecting members can be split into a constant number, c_k , of subsystems such that no two arcs belonging to the

same subsystem intersect. He emphasized the first interesting special case, where $k = 3$ and the arcs are straight-line segments. A positive answer to Erdős' question would imply that Conjecture 17 is true. To see this, observe that no k members of the system of edges (open arcs) of a k -quasiplanar topological graph G intersect. If this system can be decomposed into c_k subsystems consisting of disjoint arcs, then one of these subsystems has at least $|E(G)|/c_k$ members. The corresponding edges form a planar subgraph of G , therefore we would obtain $|E(G)|/c_k \leq 3n - 6$, where $n \geq 3$ denotes the number of vertices of G . This would imply $|E(G)| = O_k(n)$, as required. However, Pawlik, Kozik, Krawczyk, Lasoń, Miczek, Trotter, and Walczak [91] constructed systems of n segments, no 3 of which are pairwise intersecting, such that they cannot be decomposed into fewer than $\log \log n$ subsystems of disjoint segments. Therefore, the answer to Erdős' question is no. It is interesting to observe that Conjecture 17 would also follow from the following weaker statement, which was not refuted by the construction of Pawlik *et al.*

Conjecture 19. *For any positive integer k , there is a constant $\varepsilon_k > 0$ with the property that every system on n continuous arcs (or segments) in the plane, no k of which are pairwise intersecting, has at least $\varepsilon_k n$ disjoint members.*

As the number of edges of a topological graph G with n vertices substantially exceeds the critical threshold $3n - 6$, more complicated crossing configurations appear. A $k \times l$ grid in G is a pair of disjoint subsets $E_1, E_2 \subset E(G)$ with $|E_1| = k$ and $|E_2| = l$ such that every edge in E_1 crosses all edges in E_2 . It was proved in [81] that for any integer $k > 0$, there is a constant C_k such that every topological graph with n vertices and more than $C_k n$ edges has a $k \times k$ grid. See [46], for a different proof. The strongest result in this direction was proved by Tardos and Tóth [104]: There is a constant C_k such that in every topological graph with n vertices and more than $C_k n$ edges one can find 3 disjoint k -element sets of edges such that two of the subsets consist of edges incident to a vertex and every pair of edges from different subsets cross.

At first glance, one might believe that it is much easier to guarantee the existence of a $k \times k$ grid in “general position” in the sense that no pair of its edges share an endpoint. However, in this case the proof breaks down and we can only prove that every topological graph with n vertices and at least $C_k n \log^* n$ edges contains such a grid, where \log^* denotes the iterated logarithm function [2].

Conjecture 20 (Ackerman, Fox, Pach, Suk [2]). *For any integers $k, l \geq 1$, there is a constant $C_{k,l}$ such that every topological graph with n vertices which contains no $k \times l$ grid with distinct vertices has at most $C_{k,l} n$ edges.*

This conjecture is known to be true for $l = 1$.

In lack of nontrivial examples (or counterexamples), one can formulate an even bolder conjecture. We call a $k \times l$ grid *natural* if it consists of a set of k disjoint (noncrossing) edges and a set of l disjoint edges with all $2(k + l)$ endpoints distinct, such that every edge in the first subset crosses every edge in the second. There are complete topological graphs in which every pair of edges cross, so they contain no natural 2×1 grid. Hence, to strengthen Conjecture 20, we have to make an additional distinction. For instance, we may restrict our attention to simple topological graphs or to geometric graphs.

Conjecture 21 [2]. *For any integers $k, l \geq 1$, there is a constant $C_{k,l}$ such that the number of edges of any simple topological graph with n vertices which contains no $k \times l$ natural grid is at most $C_{k,l}n$.*

Even for geometric graphs with no natural $k \times k$ grid, the best known upper bound for the number of edges is $O(k^2 n \log^2 n)$. For convex geometric graphs, the validity of the conjecture follows from [62]. In general, the only case in which Conjecture 21 has been verified is $k = 2, l = 1$ (see [2]).

We close this section with another relaxation of planarity, where we do have nontrivial constructions and we know that the number of edges forcing some crossing subconfigurations is superlinear. For any $k \geq 3$, a topological graph G is called *k -locally planar* if G has no selfintersecting path of length at most k . Roughly speaking, this means that the embedding of the graph is planar in a neighborhood of radius $k/2$ around any vertex. It was shown by Pach, Pinchasi, Tardos, and Tóth [82] that there exist 3-locally planar geometric graphs with n vertices and with at least constant times $n \log n$ edges. For larger values of k , Tardos [103] constructed a sequence of k -locally planar geometric graphs with n vertices and a superlinear number of edges (approximately n times the $\lfloor k/2 \rfloor$ times iterated logarithm of n). From the other direction, we only have a much weaker bound.

Theorem 22 [82]. *The number of edges of a 3-locally planar topological graph with n vertices is $O(n^{3/2})$.*

This result is probably far from being optimal. For 3-locally planar geometric graphs (and, more generally, for topological graphs with x -monotone edges) the $\Omega(n \log n)$ bound is known to be tight [82]. Boutin [13] showed that the number of edges of 3-locally planar convex geometric graph with n vertices is $O(n)$.

REFERENCES

- [1] E. Ackerman: On the maximum number of edges in topological graphs with no four pairwise crossing edges, *Discrete Comput. Geom.* **41** (2009), 365–375.
- [2] E. Ackerman, J. Fox, J. Pach, and A. Suk: On grids in topological graphs, in: *25th ACM Symp. on Comput. Geom. (SoCG)*, ACM Press, New York, 2009, 403–412.
- [3] E. Ackerman and G. Tardos: On the maximum number of edges in quasi-planar graphs, *J. Combin. Theory, Ser. A* **114** (2007), 563–571.
- [4] P. K. Agarwal, B. Aronov, J. Pach, R. Pollack, and M. Sharir: Quasi-planar graphs have a linear number of edges, *Combinatorica* **17** (1997), 1–9.
- [5] M. Ajtai, V. Chvátal, M. Newborn, and E. Szemerédi: Crossing free graphs, *Ann. Discrete Math.* **12** (1982), 9–12.
- [6] N. Alon and P. Erdős: Disjoint edges in geometric graphs, *Discrete Comput. Geom.* **4** (1989), 287–290.
- [7] D. Avis, P. Erdős, and J. Pach: Repeated distances in space, *Graphs Combin.* **4** (1988), 207–217.
- [8] S. Avital and H. Hanani: Graphs, continuation, *Gilyonot Le'matematika* **3**, issue 2 (1966), 2–8.
- [9] H. Baron: Lösung der Aufgabe 167, *Jahresbericht Deutsch. Math.-Verein.* **45** (1935), 112.
- [10] B. Bollobás and A. Thomason: Proof of a conjecture of Mader, Erdős and Hajnal on topological complete subgraphs, *European J. Combin.* **19** (1998), 883–887.
- [11] A. V. Bondarenko: On Borsuk's conjecture for two-distance sets, arXiv 1305.2584, 8 pp.
- [12] K. Borsuk: Drei Sätze über die n -dimensionale euklidische Sphäre, *Fund. Math.* **20** (1933), 177–190.
- [13] D. Boutin: Convex geometric graphs with no short self-intersecting paths, *Congr. Numer.* **160** (2003), 205–214.
- [14] P. Brass: On the maximum number of unit distances among n points in dimension four, in: *Intuitive Geometry (I. Bárány et al., eds.)*, Bolyai Soc. Math. Studies **4**, Springer, Berlin, 1997, 277–290.
- [15] P. Brass, W. Moser, and J. Pach: *Research Problems in Discrete Geometry*, Springer, New York, 2005.
- [16] G. Cairns and Y. Nikolayevsky: Bounds for generalized thrackles, *Discrete Comput. Geom.* **23** (2000), 191–206.
- [17] G. Cairns and Y. Nikolayevsky: Outerplanar thrackles, *Graphs Combin.* **28** (2012), 85–96.
- [18] P. A. Catlin: Hajós' graph-coloring conjecture: variations and counterexamples, *J. Combin. Theory, Ser. B* **26** (1979), 268–274.
- [19] J. Černý: Geometric graphs with no three disjoint edges, *Discrete Comput. Geom.* **34** (2005), 679–695.
- [20] F. R. K. Chung: On the number of different distances determined by n points in the plane, *J. Combin. Theory, Ser. A* **36** (1984), 342–354.

- [21] F. R. K. Chung, E. Szemerédi, and W. T. Trotter: The number of different distances determined by a set of points in the Euclidean plane, *Discrete Comput. Geom.* **7** (1992), 1–11.
- [22] K. Clarkson, H. Edelsbrunner, L. Guibas, M. Sharir, and E. Welzl: Combinatorial complexity bounds for arrangements of curves and spheres, *Discrete Comput. Geom.* **5** (1990), 99–160. See also: H. Kaplan, J. Matoušek, Z. Safernová, and M. Sharir: Unit distances in three dimensions, *Combin. Probab. Comput.* **21** (2012), 597–610; and J. Zahl: An improved bound on the number of point-surface incidences in three dimensions, *Contrib. Discrete Math.*, to appear.
- [23] G. A. Dirac: Homomorphism theorems for graphs, *Math. Ann.* **153** (1964), 69–80.
- [24] G. A. Dirac: Chromatic number and topological complete subgraphs, *Canad. Math. Bull.* **8** (1965), 711–715.
- [25] V. L. Dolnikov: Some properties of graphs of diameters, *Discrete Comput. Geom.* **24** (2000), 293–299.
- [26] H. G. Eggleston: Covering a three-dimensional set with sets of smaller diameter, *J. London Math. Soc.* **30** (1955), 11–24.
- [27] G. Elekes: On the number of sums and products, *Acta Arith.* **81** (1997), 365–367.
- [28] G. Elekes and M. Sharir: Incidences in three dimensions and distinct distances in the plane, *Combin. Probab. Comput.* **20** (2011), 571–608.
- [29] P. Erdős: On sets of distances of n points, *Amer. Math. Monthly* **53** (1946), 248–250.
- [30] P. Erdős: Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53** (1947), 292–294.
- [31] P. Erdős: Néhány geometriai problémáról, *Mat. Lapok* **8** (1957), 86–92.
- [32] P. Erdős: On sets of distances of n points in Euclidean space, *Magyar Tudom. Akad. Matem. Kut. Int. Közl. (Publ. Math. Inst. Hung. Acad. Sci.)* **5** (1960), 165–169.
- [33] P. Erdős: On some problems of elementary and combinatorial geometry, *Ann. Mat. Pura Appl. (4)* **103** (1975), 99–108.
- [34] P. Erdős: Extremal problems in number theory, combinatorics and geometry, in: *Proceedings of the International Congress of Mathematicians, Vol. 1 (Warsaw, 1983)*, PWN, Warsaw, 1984, 51–70.
- [35] P. Erdős: Problems and results in discrete mathematics, *Discrete Math.* **136** (1994), 53–73.
- [36] P. Erdős and S. Fajtlowicz: On the conjecture of Hajós, *Combinatorica* **1** (1981), 141–143.
- [37] P. Erdős and A. Hajnal: On complete topological subgraphs of certain graphs, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **7** (1964), 143–149.
- [38] P. Erdős, L. Lovász, and K. Vesztegombi: On graphs of large distances, *Discrete Comput. Geom.* **4** (1989), 541–549.
- [39] P. Erdős and J. Pach: Variations on the theme of repeated distances, *Combinatorica* **10** (1990), 261–269.
- [40] P. Erdős and G. Purdy: Extremal problems in combinatorial geometry, in: *Handbook of Combinatorics, Vol. 1*, Elsevier Sci. B. V., Amsterdam, 1995, 809–874.

- [41] P. Erdős and A. H. Stone: On the structure of linear graphs, *Bull. Amer. Math. Soc.* **52** (1946), 1087–1091.
- [42] S. Felsner: *Geometric Graphs and Arrangements*, Vieweg & Sohn, Wiesbaden, 2004.
- [43] W. Fenchel and J. W. Sutherland: Lösung der Aufgabe 167, *Jahresbericht Deutsch. Math.-Verein.* **45** (1935), 33–35.
- [44] J. Fox and J. Pach: Coloring K_k -free intersection graphs of geometric objects in the plane, *European J. Combin.* **33** (2012), 853–866.
- [45] J. Fox, J. Pach, and A. Suk: The number of edges in k -quasi-planar graphs, *SIAM J. Discrete Math.*, **27** (2013), 550–561.
- [46] J. Fox, J. Pach, and C. D. Tóth: A bipartite strengthening of the crossing lemma, *J. Combin. Theory, Ser. B* **100** (2010), 23–35.
- [47] J. Fox and B. Sudakov: Density theorems for bipartite graphs and related Ramsey-type results, *Combinatorica* **29** (2009), 153–196.
- [48] R. Fulek and J. Pach: A computational approach to Conway’s thrackle conjecture, *Comput. Geom.* **44** (2011), 345–355.
- [49] R. Fulek and A. Ruiz-Vargas: Topological graphs: empty triangles and disjoint matchings, *Proc. 29th Symposium on Computational Geometry (SoCG ’13)*, ACM Press, New York, 2013, to appear.
- [50] W. Goddard, M. Katchalski, and D. Kleitman: Forcing disjoint segments in the plane, *European J. Combin.* **17** (1996), 391–395.
- [51] J. E. Goodman and J. O’Rourke, eds.: *Handbook of Discrete and Computational Geometry. 2nd edition*, Chapman & Hall/CRC, Boca Raton, 2004.
- [52] B. Grünbaum: A proof of Vázsonyi’s conjecture, *Bull. Res. Council Israel, Sect. A* **6** (1956), 77–78.
- [53] L. Guth and N. H. Katz: On the Erdős distinct distance problem in the plane, preprint, arXiv:1011.4105.
- [54] A. Heppes: Beweis einer Vermutung von A. Vázsonyi, *Acta Math. Acad. Sci. Hungar.* **7** (1956), 463–466.
- [55] A. Heppes and P. Révész: Zum Borsukschen Zerteilungsproblem, *Acta Math. Acad. Sci. Hungar.* **7** (1956), 159–162.
- [56] A. Hinrichs and Ch. Richter: New sets with large Borsuk numbers, *Discrete Math.* **270** (2003), 137–147.
- [57] H. Hopf and E. Pannwitz: Aufgabe Nr. 167, *Jahresbericht d. Deutsch. Math.-Verein.* **43** (1934), 114.
- [58] H. A. Jung: Anwendung einer Methode von K. Wagner bei Färbungsproblemen für Graphen, *Math. Ann.* **161** (1965), 325–326.
- [59] J. Kahn and G. Kalai: A counterexample to Borsuk’s conjecture, *Bull. Amer. Math. Soc. (N.S.)* **29** (1993), 60–62.
- [60] N. H. Katz: On arithmetic combinatorics and finite groups, *Illinois J. Math.* **49** (2005), 33–43.
- [61] N. H. Katz and G. Tardos: A new entropy inequality for the Erdős distance problem, in: *Towards a Theory of Geometric Graphs, Contemp. Math.* **342**, Amer. Math. Soc., Providence, 2004, 119–126.

- [62] M. Klazar and A. Marcus: Extensions of the linear bound in the Füredi-Hajnal conjecture, *Adv. in Appl. Math.* **38** (2007), 258–266.
- [63] J. Komlós and E. Szemerédi: Topological cliques in graphs. II, *Combin. Probab. Comput.* **5** (1996), 79–90.
- [64] M. van Kreveld and B. Speckmann, eds.: *Graph Drawing*. (Revised selected papers from the 19th International Symposium (GD 2011) held at the Technical University of Eindhoven, Eindhoven.) *Lecture Notes in Computer Science* **7034**, Springer, Heidelberg, 2012.
- [65] Y. S. Kupitz: *Extremal Problems of Combinatorial Geometry*, *Lecture Notes Series* **53**, Aarhus University, Denmark, 1979.
- [66] Y. S. Kupitz, H. Martini, and M. A. Perles: Finite sets in R^d with many diameters—a survey, in: *Proceedings of the International Conference on Mathematics and Applications (ICMA-MU 2005, Bangkok)*, Mahidol University Press, Bangkok, 2005, 91–112. Also in: *East-West J. Math.: Contributions in Mathematics and Applications* (2007), 41–57.
- [67] Y. S. Kupitz, H. Martini, and B. Wegner: Diameter graphs and full equi-intersectors in classical geometries, in: *IV. International Conference in Stoch. Geo., Conv. Bodies, Emp. Meas. & Apps. to Eng. Sci., Vol. II, Rend. Circ. Mat. Palermo (2) Suppl. No. 70, part II* (2002), 65–74.
- [68] T. Leighton; *Complexity Issues in VLSI. Foundations of Computing Series*, MIT Press, Cambridge, MA, 1983.
- [69] L. Lovász, J. Pach, and M. Szegedy: On Conway’s thrackle conjecture, *Discrete Comput. Geom.* **18** (1997), 369–376.
- [70] W. Mader: Homomorphieeigenschaften und mittlere Kantendichte von Graphen, *Math. Ann.* **174** (1967), 265–268.
- [71] W. Mader: $3n - 5$ edges do force a subdivision of K_5 , *Combinatorica* **18** (1998), 569–595.
- [72] B. Mohar and C. Thomassen: *Graphs on Surfaces*, Johns Hopkins University Press, Baltimore, MD, 2001.
- [73] F. Morić and J. Pach: Large simplices determined by finite point sets, *Beiträge zur Algebra und Geometrie*, to appear.
- [74] F. Morić and J. Pach: Remarks on Schur’s conjecture, manuscript.
- [75] L. Moser: On the different distances determined by n points, *Amer. Math. Monthly* **59** (1952), 85–91.
- [76] A. Nilli: On Borsuk’s problem, in: *Jerusalem Combinatorics ’93, Contemporary Math.* **178** (1994), 209–210.
- [77] M. Morse: The International Congress in Oslo, *Bull. Amer. Math. Soc.* **42** (1936), 777–781.
- [78] J. Pach, ed.: *Towards a Theory of Geometric Graphs. Contemp. Math.* **342**, Amer. Math. Soc., Providence, RI, 2004.
- [79] J. Pach, ed.: *Thirty Essays on Geometric Graph Theory*, Springer, New York, 2013.
- [80] J. Pach and P. K. Agarwal: *Combinatorial Geometry*, John Wiley & Sons, New York, 1995.

- [81] J. Pach, R. Pinchasi, M. Sharir, and G. Tóth: Topological graphs with no large grids, *Graphs and Combinatorics* **21** (2005), 355–364.
- [82] J. Pach, R. Pinchasi, T. Tardos, and G. Tóth: Geometric graphs with no self-intersecting path of length three, *European J. Combin.* **25** (2004), 793–811.
- [83] J. Pach, R. Radoičić, and G. Tóth: Relaxing planarity for topological graphs, in: *Discrete and Computational Geometry, Lecture Notes in Comput. Sci.* **2866**, Springer, Berlin, 2003, 221–232.
- [84] J. Pach, R. Radoičić, G. Tardos, and G. Tóth: Improving the crossing lemma by finding more crossings in sparse graphs, *Discrete Comput. Geom.* **36** (2006), 527–552.
- [85] J. Pach, F. Shahrokhi, and M. Szegedy: Applications of the crossing number, *Algorithmica* **16** (1996), 111–117.
- [86] J. Pach, J. Solymosi, and G. Tóth: Unavoidable configurations in complete topological graphs, *Discrete Comput. Geom.* **30** (2003), 311–320.
- [87] J. Pach and E. Sterling: Conway’s conjecture for monotone thrackles, *Amer. Math. Monthly* **118**, 544–548.
- [88] J. Pach and G. Tardos: Forbidden paths and cycles in ordered graphs and matrices, *Israel J. Math.* **155** (2006), 359–380.
- [89] J. Pach and G. Tóth: Disjoint edges in topological graphs, *J. Comb.* **1** (2010), 335–344.
- [90] J. Pach and J. Töröcsik: Some geometric applications of Dilworth’s theorem, *Discrete Comput. Geom.* **12** (1994), 1–7.
- [91] A. Pawlik, J. Kozik, T. Krawczyk, M. Lasoń, P. Miczek, W. Trotter, and B. Walczak: Triangle-free intersection graphs of line segments with large chromatic number, preprint, arXiv:1209.1595.
- [92] A. Perlestein and R. Pinchasi: Generalized thrackles and geometric graphs in \mathbb{R}^3 with no pair of strongly avoiding edges, *Graphs Combin.* **24** (2008), 373–389.
- [93] A. M. Raigorodskii: *Three Lectures on the Borsuk Partition Problem. Surveys in Contemporary Mathematics, London Math. Soc. Lecture Note Ser.* **347**, Cambridge Univ. Press, Cambridge, 2008, 202–247.
- [94] Z. Schur, M. A. Perles, H. Martini, and Y. S. Kupitz: On the number of maximal regular simplices determined by n points in \mathbb{R}^d , in: *Discrete and Computational Geometry, The Goodman-Pollack Festschrift (Aronov et al., eds.)*, *Algorithms Combin.* **25**, Springer, Berlin, 2003, 767–787.
- [95] J. Solymosi and Cs. Tóth: Distinct distances in the plane, *Discrete Comput. Geom.* **25** (2001), 629–634.
- [96] J. Spencer, E. Szemerédi, and W. T. Trotter: Unit distances in the Euclidean plane, in: *Graph Theory and Combinatorics* (B. Bollobás, ed.), Academic Press, London, 1984, 293–303.
- [97] S. Straszewicz: Sur un problème géométrique de P. Erdős, *Bull. Acad. Pol. Sci., Cl. III* **5** (1957), 39–40.
- [98] A. Suk: Disjoint edges in complete topological graphs, in: *Proc. 28th Symposium on Computational Geometry (SoCG’12)*, ACM Press, New York, 2012, 383–386.

- [99] K. J. Swanepoel: A new proof of Vázsonyi's conjecture, *J. Combinat. Theory, Ser. A* **115** (2008), 888–892.
- [100] K. J. Swanepoel: Unit distances and diameters in Euclidean spaces, *Discrete Comput. Geom.* **41** (2009), 1–27.
- [101] L. A. Székely: Crossing numbers and hard Erdős problems in discrete geometry, *Combin. Probab. Comput.* **6** (1997), 353–358.
- [102] G. Tardos: On distinct sums and distinct distances, *Adv. Math.* **180** (2003), 275–289.
- [103] G. Tardos: Construction of locally plane graphs with many edges, in: *Thirty Essays on Geometric Graph Theory* (J. Pach, ed.), Springer, New York, 2013, 541–562.
- [104] G. Tardos and G. Tóth: Crossing stars in topological graphs, *SIAM J. Discrete Math.* **21** (2007), 737–749.
- [105] C. Thomassen: Some remarks on Hajós' conjecture, *J. Combin. Theory, Ser. B* **93** (2005), 95–105.
- [106] G. Tóth: Note on geometric graphs, *Journal of Combinatorial Theory, Ser. A* **89** (2000), 126–132.
- [107] G. Tóth and P. Valtr: Geometric graphs with few disjoint edges, *Discrete Comput. Geom.* **22** (1999), 633–642.
- [108] P. Turán: Egy gráfelméleti szélsőértékfeladatról, *Matematikai és Fizikai Lapok* **48** (1941), 436–452.
- [109] P. Valtr: Graph drawing with no k pairwise crossing edges, in: *Graph Drawing, Lecture Notes in Comput. Sci.* **1353**, Springer, Berlin, 1997, 205–218.
- [110] P. Valtr: On geometric graphs with no k pairwise parallel edges, *Discrete Comput. Geom.* **19** (1998), 461–469.
- [111] K. Vesztérgombi: On the distribution of distances in finite sets in the plane, *Discrete Math.* **57** (1985), 129–145.
- [112] S. Vidor: *Síkgráfok és Általánosításai*, Diploma thesis, Eötvös University, Budapest, 2009.
- [113] K. Wagner: Beweis einer Abschwächung der Hadwiger-Vermutung, *Math. Ann.* **153** (1964), 139–141.
- [114] D. R. Woodall: Thrackles and deadlock, in: *Combinatorics, Proc. Conf. Comb. Math.* (D. Welsh, ed.), Academic Press, London, 1971, 335–347.

János Pach

EPFL, Station 8,
CH-1015 Lausanne,
Switzerland

e-mail: pach@cims.nyu.edu

and

Alfréd Rényi Institute of
Mathematics,
Hungarian Academy of Sciences,
Budapest, Reáltanoda u. 13–15,
H-1053, Hungary

PAUL ERDŐS AND THE DIFFERENCE OF PRIMES

JÁNOS PINTZ*

In the present work we discuss several problems concerning the difference of primes, primarily regarding the difference of consecutive primes. Most of them were either initiated by Paul Erdős (sometimes with coauthors), or were raised ahead of Erdős; nevertheless he was among those who reached very important results in them (like the problem of the large and small gaps between consecutive primes).

1. INTRODUCTION

Number theory, especially primes, belonged to one of the most favourite subjects of Paul Erdős. He writes in the obituary of his long-time friend and collaborator Paul Turán [27, 1980]: “We first met at the University of Budapest in September 1930 and immediately discovered our common interest in number theory and prime numbers in particular.” His first result which made him famous was a new simple proof of Chebyshev’s theorem, according to which there is always a prime between n and $2n$ for any natural number n . The elementary proof of the Prime Number Theorem (PNT) by Erdős [20, 1949] and Selberg [86, 1949], asserting

$$(1.1) \quad \pi(x) = \sum_{p \leq x} 1 \sim \frac{x}{\log x} \sim \int_2^x \frac{dt}{\log t},$$

was a great sensation in mathematics. Among his 1595 mathematical works listed in MathSciNet 77 have the word prime in the title and in total 259 contain the word prime in the abstract (although some of them belong to other subjects, for example, to combinatorics).

*Supported by OTKA grants K72731, K100291, NK 104183 and ERC-AdG. 228005.

Apart from the well-known global problems, like the PNT (see (1.1)) he raised and investigated many problems about local questions, like gaps between consecutive primes, which he characterized as unconventional problems in the same obituary [27, 1980].

In view of the enormously rich mathematical activity of Paul Erdős, it would be hopeless to give a full survey of his works concerning primes (especially in a paper of about 20 pages). Another handicap is that although he formulated numerous interesting questions about local distribution of primes, no progress was made in many of them despite the often 5–7 decades which passed since their first appearance in a work of his.

Hence, I choose 9 groups of problems concerning the difference of primes (and, exceptionally, in Section 9 the difference of almost primes, that is, numbers with a bounded number of prime divisors). Most of them (seven out of nine, the exceptions being Sections 6 and 10, which deal with some other important and natural problems concerning differences of primes) were either initiated by Paul Erdős (sometimes with coauthors like in the case of the famous Erdős–Turán problem on arithmetic progressions in the sequence of primes or that of the Erdős–Mirsky conjecture on consecutive equal values of the divisor function), or were raised ahead of Erdős; nevertheless he was among those who reached very important results in them (like the problem of the large and small prime gaps in Sections 2 and 3, respectively).

Finally, I have chosen those problems where I (very often with the coauthors S. W. Graham, D. Goldston and C. Yıldırım) succeeded to reach some progress in the last few years. Some of the works (containing Theorems 11–19 and Theorem 25) are still in preparation.

During our work p , p' , p_i will always denote primes (usually with p_n being the n^{th} prime), \mathcal{P} the set of all primes and

$$(1.2) \quad d_n = p_{n+1} - p_n$$

the n^{th} difference between consecutive primes. It is a trivial consequence of the PNT (see (1.1)) that the average of d_n is $\log n$. The most basic (and hopelessly difficult) question would be the problem of small and large values of d_n where the following classical conjectures are well known:

Twin Prime Conjecture. $\liminf d_n = 2$.

Cramér's Conjecture ([12, 1934], [13, 1936]). $\limsup_{n \rightarrow \infty} \frac{d_n}{\log^2 n} = 1$.

Cramér's conjecture implies the 100-year-old

Landau's Conjecture ([61, 1912]). *There is always a prime between two neighbouring squares.*

Finally, we mention that while the Twin Prime Conjecture and Landau's conjecture are generally believed to be true, there are serious doubts on the validity of Cramér's conjecture (see the works of Granville [44, 1994], [45, 1995] and Hildebrand–Maier [53, 1989]). Nevertheless, it is still believed that the correct order of magnitude of the largest values of d_n is $(\log n)^{2+o(1)}$. Granville conjectures (on the basis of theoretic arguments in connection with Maier's matrix method) that Cramér's conjecture would be true if the constant 1 were substituted by $2e^{-\gamma} = 1.1229\dots$ (See [44, 1994], [45, 1995], [69, 2007].) It is interesting to mention that Cramér based his conjecture on the behaviour of a random model, where each number $n > 2$ is independently “chosen to be prime” with a probability $1/\log n$, corresponding to the density of primes near n .

Finally, we mention Erdős's feeling about the enormous difficulty of Cramér's above mentioned conjecture. He commented on it in [25, 1976]: “This is clearly hopeless with the techniques which are at our disposal at present (and perhaps for the next few hundred or thousand years).”

2. LARGE DIFFERENCES BETWEEN CONSECUTIVE PRIMES: THE ERDŐS–RANKIN PROBLEM

Due to the Prime Number Theorem (1.1) we clearly have

$$(2.1) \quad \lambda := \limsup_{n \rightarrow \infty} \frac{d_n}{\log p_n} \geq 1.$$

This was improved in subsequent works of Backlund [1, 1929] and Brauer–Zeitzi [7, 1930] to $\lambda \geq 2$ and $\lambda \geq 4$, respectively. Already one year later Westzynthius [98, 1931] showed

$$(2.2) \quad \limsup_{n \rightarrow \infty} \frac{d_n \log_4 p_n}{\log p_n \log_3 p_n} \geq 2e^\gamma,$$

where γ is Euler's constant and $\log_\nu n$ denotes the ν -times iterated logarithmic function.

G. Ricci [78, 1934] eliminated the factor $\log_4 p_n$ in (2.2).

Erdős [17, 1935] was very much interested in the problem and he succeeded in showing

$$(2.3) \quad \limsup_{n \rightarrow \infty} \frac{d_n (\log_3 p_n)^2}{\log p_n \log_2 p_n} > 0.$$

Rankin [74, 1938] reached a further improvement of this by a $\log \log \log \log p_n$ factor three years later:

$$(2.4) \quad \limsup_{n \rightarrow \infty} \frac{d_n (\log_3 p_n)^2}{\log p_n \log_2 p_n \log_4 p_n} \geq C_1 = \frac{1}{3}.$$

Erdős commented on possible improvement of the function in (2.4): “It seems very hard to improve it.” [22, 1955].

In the following 40 years only the constant C_1 was improved to $e^{\gamma/2}$ by Schönhage [85, 1963] and in two independent works by Ricci [79, 1952] and Rankin [77, 1962] to e^γ , respectively. The lack of progress inspired Erdős in 1979 to offer USD 10,000 (see [28, 1981], for example), the greatest prize ever offered by him, for a proof that (2.4) holds for every constant C_1 .

Notwithstanding Erdős’s offer, even the further improvements referred only to the value of C_1 . H. Maier and C. Pomerance [65, 1990] used deep methods from analytic number theory beyond the original classical sieve methods to prove a Bombieri–Vinogradov type theorem for generalized twin primes. Afterwards they arrived at a combinatorial problem which they solved by the greedy algorithm and obtained (2.4) with the value

$$(2.5) \quad C_1 = 1.3126 \dots e^\gamma.$$

A few years later I succeeded in improving the combinatorial part by using probabilistic methods. A deep method of combinatorics, the semi-random method of Szemerédi, led to a full solution of the combinatorial problem and yielded a constant 2 in the combinatorial problem and thereby the result [68, 1997].

Theorem 1 ([68, 1997]). (2.4) holds with $C_1 = 2e^\gamma$.

It is interesting to note that the deterministic approach (the greedy algorithm) of Maier and Pomerance yielded the constant 1.3126 in the combinatorial problem, while a pure (and relatively simple) probabilistic method would have yielded a weaker estimate, only $1.04 \dots$, but nevertheless an improvement over the earlier best results of Ricci and Rankin. Finally, the semirandom method turned to be the optimal one leading to the constant 2. The fact that the given combinatorial result cannot be further improved shows that essential new ideas are necessary to improve (2.4) to any $C_1 > 2e^\gamma$.

Erdős was the first to consider the problem whether neighboring prime gaps can be simultaneously large. He succeeded in showing [21, 1949]

$$(2.6) \quad \limsup_{n \rightarrow \infty} \frac{\min(d_n, d_{n+1})}{\log p_n} = \infty.$$

This was significantly superseded by H. Maier, who proved the analogue of (2.4) for k consecutive differences 32 years later [63, 1981]; namely,

$$(2.7) \quad \limsup_{n \rightarrow \infty} \frac{\min(d_{n+1}, \dots, d_{n+k})}{\log n \log_2 n \log_4 n / \log_3^2 n} > 0$$

for any natural number k (we remark that by $\log_\nu n \sim \log_\nu p_n$ we can clearly substitute p_n by n in all formulae).

Concerning upper bounds for d_n , we will be brief, since Erdős himself did not work on such problems. Exactly 100 years ago Landau [61, 1912] formulated the conjecture that there is always a prime between two neighbouring squares. (Some other similar conjectures were known already earlier.) The starting point was 18 years later, when G. Hoheisel [54, 1930] showed the existence of primes in intervals of type

$$(2.8) \quad [x, x + x^C] \quad C = 1 - \frac{1}{33\,000}.$$

This was improved nearly 20 times during the next seven decades (for a history see [67, 2000]) until in a joint work with R. Baker and G. Harman we reached the present record.

Theorem 2 ([2, 2001]). (2.8) holds with $C = \frac{1}{2} + \frac{1}{40} = 0.525$.

3. SMALL DIFFERENCES BETWEEN CONSECUTIVE PRIMES

As it has been already mentioned in the Introduction, according to the twin prime conjecture the smallest possible prime gap occurring infinitely often between consecutive primes should be two. The weaker relation

$$(3.1) \quad \Delta_1 := \liminf_{n \rightarrow \infty} \frac{d_n}{\log p_n} \leq 1$$

follows immediately from the Prime Number Theorem (1.1).

The first, although conditional improvement over the “trivial” estimate (3.1) was achieved by Hardy and Littlewood in 1926 (unpublished, see [75, 1940])

$$(3.2) \quad \Delta_1 \leq 2/3$$

under the assumption of the Generalized Riemann Hypothesis (GRH). This happened three years before the first non-trivial result $\lambda \geq 2$ of Backlund

(cf. (2.1)) was reached concerning large differences. In the case of large differences, the result $\lambda = \infty$ of Weszynthius (cf. (2.2)) was achieved two years later, in 1931; while the presently known largest order of magnitude of differences, the estimate (2.4) of Rankin, was proved in 1938. The progress in the last 75 years affected only the value of the constant C_1 in the estimate (2.4).

In contrast to this, the first non-trivial unconditional estimate for small differences, the relation

$$(3.3) \quad \Delta_1 < 1,$$

was shown by Paul Erdős in 1940 [18, 1940], two years after Rankin's above mentioned result. Erdős used Brun's sieve to prove (3.3).

The progress was afterwards much slower than in the case of large differences. During a period of 25 years, the estimate of Erdős was reduced finally to $\Delta_1 < 29/32$ in three subsequent works by Rankin [76, 1947], Ricci [80, 1954] and later by Wang Yuan, Xie Sheng-gang and Yu Kun-rui [97, 1965].

In 1966 Bombieri and Davenport [4, 1966] refined and made the method of Hardy and Littlewood unconditional by substituting the Bombieri–Vinogradov theorem for the GRH, and thus obtained $\Delta_1 \leq 1/2$. They also combined their method with that of Erdős, which led to

$$(3.4) \quad \Delta_1 \leq \frac{2 + \sqrt{3}}{8} = 0.4665 \dots$$

In the following two decades this bound was diminished in five works by Piltjai, Huxley and at last by Fouvry and Grupp in 1986 to $\Delta_1 \leq 0.4342$.

In 1988 H. Maier [64, 1988] succeeded in significantly improving the result $0.4425 \dots$ of Huxley [56, 1977] to

$$(3.5) \quad \Delta_1 \leq e^{-\gamma} \cdot 0.4425 \dots = 0.2484 \dots < 1/4.$$

He combined the methods of Erdős and of Bombieri–Davenport (including the refinements of Huxley) with his newly invented matrix method which he had used a few years earlier to show his result (2.7) on successive large differences.

Finally, in 2005, in a joint work with D. A. Goldston and C. Yıldırım [38, 2009] we showed by a combination of Selberg's sieve and the large sieve:

Theorem 3 ([38, 2009]). $\Delta_1 = 0$, that is, $\liminf_{n \rightarrow \infty} \frac{d_n}{\log n} = 0$.

It turned out during our investigations that the existence of short gaps between consecutive primes strongly depends on the uniform distribution of primes in arithmetic progressions. A good measure to this phenomenon is the *level of distribution of primes*. We say that the primes have level ϑ of distribution, if the relation

$$(3.6) \quad \sum_{q \leq X^{\vartheta - \varepsilon}} \max_{\substack{a \\ (a,q)=1}} \left| \sum_{\substack{p \equiv a \pmod{q} \\ p \leq X}} \log p - \frac{X}{\varphi(q)} \right| \ll_{A,\varepsilon} \frac{X}{(\log X)^A}$$

holds for any $A, \varepsilon > 0$.

The best known value $\vartheta = 1/2$ was shown independently by E. Bombieri [3, 1965] and A. I. Vinogradov [94, 1965]: this is the celebrated Bombieri–Vinogradov’s theorem.

Soon after this result, P. D. T. A. Elliott and H. Halberstam [16, 1970] expressed their conjecture that (3.6) holds with $\vartheta = 1$, namely, the largest possible value. This is generally referred to as the Elliott–Halberstam conjecture (EH).

The original qualitative result of Theorem 3 could be significantly sharpened, especially under the assumption that the primes have a level of distribution $\vartheta > 1/2$, even if ϑ is arbitrarily close to $1/2$. In case of $\vartheta = 1$, that is, when we suppose the Elliott–Halberstam conjecture, we could even show that $d_n \leq 16$ for infinitely many n .

Theorem 4 ([38, 2009]). *If the primes have a level of distribution $\vartheta > 1/2$, then with a suitable explicit constant $C(\vartheta)$, depending on ϑ we have*

$$(3.7) \quad \liminf_{n \rightarrow \infty} d_n \leq C(\vartheta).$$

Here we have $C(0.971) = 16$, in particular EH implies

$$(3.8) \quad \liminf d_n \leq 16.$$

Theorem 5 ([39, 2010]). *We have unconditionally*

$$(3.9) \quad \liminf_{n \rightarrow \infty} \frac{d_n}{\sqrt{\log p_n} (\log \log p_n)^2} < \infty.$$

Finally, I succeeded in improving this to

Theorem 6. *With a suitable $c > 0$ we have unconditionally*

$$(3.10) \quad \liminf_{n \rightarrow \infty} \frac{d_n}{(\log p_n)^{3/7} (\log \log p_n)^c} < \infty.$$

It seems that the above bound is the limit of the method, so the improvement of the exponent $3/7$ needs new methods.

Similarly to the existence of simultaneous consecutive large differences, we can ask for good bounds for the quantity (ν is fixed)

$$(3.11) \quad \Delta_\nu = \liminf_{n \rightarrow \infty} \frac{d_{n+\nu} - d_n}{\log p_n}$$

beyond the trivial consequence $\Delta_\nu \leq \nu$ of the Prime Number Theorem (1.1). Erdős introduced the problem to show that $\Delta_2 < 1$ (see [24, 1973], for example). In their earlier mentioned work Bombieri and Davenport [4, 1966] showed $\Delta_\nu \leq \nu - 1/2$, which was later improved by Huxley [55, 1969], [56, 1977] to $\Delta_\nu \leq \nu - 5/8 + o(1/\nu)$ and even later by Goldston and Yıldırım [36, 2007] to $\Delta_\nu \leq (\sqrt{\nu} - 1/2)^2$. The estimate of Goldston and Yıldırım gives relatively strong results for small values of ν . For large values of ν the previously mentioned matrix method of Maier was the most successful [64, 1988]. He proved that

$$(3.12) \quad \Delta_\nu < e^{-\gamma} \left(\nu - \frac{5}{8} + o\left(\frac{1}{\nu}\right) \right).$$

Furthermore, answering positively Erdős’s problem he showed in the same work

$$(3.13) \quad \Delta_2 < 0.79.$$

In our joint works [37, 2006], [38, 2009] yielding $\Delta_1 = 0$ we proved the following

Theorem 7 ([37, 2006], [38, 2009]). *Suppose that the primes have level of distribution ϑ . Then, for $\nu \geq 2$ we have*

$$(3.14) \quad \Delta_\nu \leq \left(\sqrt{\nu} - \sqrt{2\vartheta} \right)^2.$$

Moreover, we have unconditionally

$$(3.15) \quad \Delta_\nu \leq e^{-\gamma} (\sqrt{\nu} - 1)^2.$$

Remark 1. The unconditional result (3.15) yields beyond the estimate $\Delta_2 < 0.79$ of Maier

$$(3.16) \quad \Delta_2 \leq \left(\sqrt{2} - 1 \right)^2 e^{-\gamma} = 0.56146\dots, \quad \Delta_j < 1 \text{ for } 3 \leq j \leq 5.$$

Remark 2. (3.14) shows that EH implies that $\Delta_2 = 0$.

It is interesting to note that in order to show the existence of infinitely many bounded differences between consecutive primes, it is sufficient to assume that primes have a level of distribution $\vartheta > 1/2$. Even $\vartheta > 0.971$ already implies $\liminf_{n \rightarrow \infty} d_n \leq 16$. On the other hand, assuming the same with any fixed $\vartheta < 1$ we are still unable to show the seemingly more harmless consequence $\Delta_2 = 0$, that is, the existence of at least three primes in intervals of length $o(\log x)$.

The crucial result behind Theorem 4 was that the assumption $\vartheta > 1/2$ implies a weak form of the conjecture of Dickson [15, 1904] about k -tuples of primes.

Dickson conjectured that if $L_i(x) = a_i x + b_i$ are linear forms with integer coefficients ($i = 1, 2, \dots, k$) and $\prod_{i=1}^k L_i(x)$ has no fixed prime divisor (that is, we have no prime p_0 with the property that $p_0 \mid \prod_{i=1}^k L_i(x)$ should hold for every integer x), then there are infinitely many different integers n such that $L_i(n)$ are simultaneously primes for all $i = 1, 2, \dots, k$.

Hardy and Littlewood [48, 1923], probably unaware of Dickson’s conjecture, expressed a quantitative form of it, according to which – in the special case of $a_i = 1$ ($i = 1, 2, \dots, k$) – one has

$$(3.17) \quad \pi_{\mathcal{H}}(x) = \sum_{\substack{n \leq x \\ n+h_i \in \mathcal{P} \ (1 \leq i \leq k)}} 1 = (\mathfrak{S}(\mathcal{H}) + o(1)) \frac{x}{\log^{k+1} x},$$

for any $\mathcal{H} = \{h_i\}_{i=1}^k$ ($h_i < h_{i+1}$), where $\mathfrak{S}(\mathcal{H}) > 0$ if and only if the system $\{n + h_i\}_{i=1}^k$ has no fixed prime divisor.

In [38, 2009] we showed a weaker conditional form of this conjecture as

Theorem 8 ([38, 2009]). *If the primes have a level $\vartheta > 1/2$ of distribution, then for any admissible k -tuple \mathcal{H} (that is, if $\{n + h_i\}_{i=1}^k$ has no fixed prime divisor) there are at least two primes among $\{n + h_i\}_{i=1}^k$ for infinitely many values of n , if $k > C_0(\vartheta)$, an explicit constant depending on ϑ .*

4. ARITHMETIC PROGRESSIONS IN PRIMES AND PRIMES IN ARITHMETIC PROGRESSIONS

Perhaps the most famous conjecture of Erdős is the Erdős–Turán conjecture on the existence of arbitrarily long (finite) arithmetic progressions (AP) in

sequences having positive (upper) density within the set of natural numbers [31, 1936]. The problem was solved for $k = 3$ (three terms AP's) by K. F. Roth [81, 1952], [82, 1953], later for $k = 4$ by E. Szemerédi [89, 1970], [90, 1969]. Soon afterwards it was proved for arbitrary values of k , again by E. Szemerédi [91, 1975]. This problem had the highest prize Paul Erdős offered for any problem (USD 1000) among the ones which were solved until now. Progress and developments later in the conjecture (to be explained below) played a decisive role in awarding of the Wolf prize once (H. Furstenberg, 2006/2007), twice in the Fields Medals (T. Gowers, 1998 and T. Tao, 2006), and once in the Abel prize (E. Szemerédi, 2012). Additionally, as described in K. F. Roth's Fields Medal laudation, his partial solution (the mentioned case $k = 3$) was characterized as his second most important work.

After Szemerédi's result, it turned out that his combinatorial solution led to the discovery of his celebrated Regularity Lemma [92, 1978]. This expresses that the vertex set of every graph \mathcal{G} with a positive density of edges can be partitioned into a bounded number of k vertex sets U_i with an equal number of vertices (up to an error of one vertex) in such a way that for most pairs U_i, U_j and any pairs $V_i \subset U_i, V_j \subset U_j$ the density of edges between V_i and V_j is approximately the same as the density of edges between U_i and U_j . The Regularity Lemma became one of the most powerful tools in graph theory. Furstenberg [35, 1977] found another proof of the Erdős–Turán conjecture based on ergodic theory. Gowers' approach [42, 1998] was somewhat similar to that of Roth's and he was the first to reach a quantitative result for $k = 4$, similar to that of Roth for $k = 3$. He showed in the case of $k = 4$ that instead of positive upper density, it is sufficient to suppose that the sequence contains at least $N(\log \log N)^{-C}$ elements until N . He extended [43, 2001] his method for a full proof of Szemerédi's theorem later (with a much stronger estimate than Szemerédi's original one, similar to the above mentioned one with a $C = C(k)$ depending only on k).

This was a novel feature, since Szemerédi's original proof gave very weak bounds, while Furstenberg's ergodic approach was completely ineffective.

As mentioned in the work of Green and Tao [47, 2008], the problem of finding long AP's in the sequence of primes was already considered more than 200 years ago by Waring and Lagrange. It is unclear who was the first to mention the following conjecture in a written form. However, it certainly appears in the above mentioned work of Erdős and Turán.

Conjecture 1 [31, 1936]. *The primes contain infinitely many k -term AP's for every k .*

This would follow from the following more general conjecture which is definitely due to Erdős.

Conjecture 2 [24, 1973]. *If the sum of reciprocals of the members of a set \mathcal{A} of positive integers diverges, then \mathcal{A} contains arbitrarily long (finite) arithmetic progressions.*

In the above work Erdős emphasizes that Conjecture 2 implies Conjecture 1 and in a later work [26, 1977] he offered USD 3000 for a solution of Conjecture 2.

The case of $k = 3$ of Conjecture 1 was solved already in 1939 by Van der Corput [11, 1939]. However, it is a simple consequence of the deep theorem that the size $E(X)$ of the exceptional set in Goldbach's problem satisfies the estimate

$$(4.1) \quad E(X) = \#\{n \leq X; 2 \mid n, n \neq p + p', p, p' \in \mathcal{P}\} \ll_A \frac{X}{(\log X)^A}$$

for any $A > 0$. This was proved simultaneously and independently by Van der Corput [10, 1937], Estermann [33, 1938] and Cudakov [14, 1938]. They all used Vinogradov's method [95, 1937], [96, 1976] which essentially solved the ternary Goldbach problem by proving that every sufficiently large odd integer can be written as the sum of three primes. The case of $k = 4$ was left open for nearly 70 years, although Heath-Brown [50, 1981] succeeded to show that there are infinitely many 4-term AP's containing three primes and a fourth term which has at most two prime factors (in other words the 4th term is a P_2 number).

Finally, in 2004 Ben Green and T. Tao proved Conjecture 1 for any k [47, 2008].

Conjecture 2 is still open for any $k \geq 3$. However, important progress was done for $k = 3$. After the initial work of Roth [82, 1953], Szemerédi [93, 1990], Heath-Brown [52, 1987], Bourgain [5, 1999], [6, 2008] and Sanders [83, 2010] proved with different values of $c < 1$ that if $\mathcal{A} \subset [1, N]$ does not contain any three-term AP, then

$$(4.2) \quad |\mathcal{A}| \ll \frac{N}{(\log N)^c}.$$

Finally, recently T. Sanders [84, 2011] has shown that in this case we have

$$(4.3) \quad |\mathcal{A}| \ll \frac{N(\log \log N)^5}{\log N}.$$

This is very close to Conjecture 2 for the case $k = 3$, since an improvement of (4.3) by a factor $(\log \log N)^{c'}$ with any $c' > 6$ would already imply Conjecture 2 for $k = 3$.

I recently investigated whether a common generalization of Theorem 1 and the Green–Tao theorem is possible. It turned out that an improvement of Theorem 8 played a crucial role in this. The following Theorem 9 is also crucial in several other results announced in the present work. In the following let us denote the smallest prime factor of n by $P(n)$.

Theorem 9 ([70, 2010]). *Suppose that primes have a level of distribution $\vartheta = \frac{1}{2} + \delta$, $\delta > 0$. Let $\mathcal{H} = \{h_i\}_{i=1}^k$ be an admissible k -tuple with $k \geq C_0(\vartheta) = (2\lceil 1/2\delta \rceil + 1)^2$. Then the number of $n \leq N$ for which $\{n + h_i\}_{i=1}^k$ contains at least two consecutive primes and almost primes satisfying $P^-(n + h_\nu) > n^{c_1(k)}$ in each component is at least*

$$(4.4) \quad c_2(\mathcal{H}) \frac{N}{\log^k N}.$$

Using a modified form of Green–Tao’s method the above result helped to show

Theorem 10 ([70, 2010]). *Let us suppose that the level ϑ of distribution of primes exceeds $1/2$. Then there is a positive $d \leq C_3(\vartheta)$ so that there are arbitrarily long AP’s of primes p with the property that $p' = p + d$ is the prime following p for each element of the progression. If $\vartheta > 0.971$, then the above holds for some d with $0 < d \leq 16$.*

A more elaborated condition which implies the existence of arbitrarily long arithmetic progressions in the sequence of twin primes is proved in [72, 2012].

5. THE NORMALIZED VALUE DISTRIBUTION OF d_n

As mentioned in the Introduction, the Prime Number Theorem (1.1) implies that

$$(5.1) \quad \lim_{n \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \frac{d_n}{\log n} = 1.$$

Hence it is natural to investigate the value distribution of $d_n/\log n$. Denoting by J the set of limit points of $d_n/\log n$, Erdős formulated the conjecture

$$(5.2) \quad J = \left\{ \frac{d_n}{\log n} \right\}' = [0, \infty].$$

For example, he writes in [22, 1955]: “It seems certain that $d_n/\log n$ is everywhere dense in $(0, \infty)$ ” (after mentioning the conjecture $\liminf_{n \rightarrow \infty} d_n/\log n = 0$). The fact that $\infty \in J$ was proved already 80 years ago by Weszynthius (cf. (2.2)). However, no finite limit point was known before our result $\Delta_1 = 0$ (cf. Theorem 3), which is equivalent to

$$(5.3) \quad 0 \in J.$$

The lack of knowledge about finite limit points of J is even more surprising in light of the fact that already nearly 60 years ago Erdős [22, 1955], and independently Ricci [80, 1954] proved that the set J has a positive Lebesgue measure.

I could not show anything unconditionally about the set J . However, it turned out that (similarly to Theorem 3 vs Theorem 4) the situation changes drastically if we assume $\vartheta > 1/2$.

In this case I could prove the following:

Theorem 11. *Suppose that primes have a level $\vartheta > 1/2$ of distribution. Then there is a constant $c = c(\vartheta)$ such that*

$$(5.4) \quad [0, c] \subset J.$$

It is interesting to remark that the proof does not supply any explicit value for c even if we know a concrete level $\vartheta > 1/2$ of the distribution of primes. While ineffective results in the distribution of primes are often connected with Siegel zeros, this is not the case here.

After lecturing earlier on (5.4) Kálmán Győry asked me the following question. Can we also say something about the value distribution of $d_n/f(n)$ if $f(n)$ is another function with $f(n) < \log n$. The investigation of

$$(5.5) \quad J_f = \left\{ \frac{d_n}{f(n)} \right\}'$$

would namely tell us more about the occurring small values of d_n .

A trivial reformulation of Theorem 5 naturally says

$$(5.6) \quad 0 \in J_f \text{ if } f(n) \gg (\log n)^{3/7} (\log \log n)^{c'}$$

if $c' > c$, where c is the constant appearing in (3.11).

It turned out that (also assuming a level of distribution $\vartheta > 1/2$ for the primes), one can affirmatively answer Kálmán Győry’s question under the very mild further assumption $f(n) < \log n$, $f(n) \rightarrow \infty$.

Theorem 12. *Let us suppose that primes have a level $\vartheta > 1/2$ of distribution. Let $f(n)$ be any function with $f(n) \rightarrow \infty$, $f(n) < \log n$. Then we have an ineffective $c_f = c_f(\vartheta)$ depending on the function $f(n)$ and ϑ such that*

$$(5.7) \quad [0, c_f] \subset J_f = \left\{ \frac{d_n}{f(n)} \right\}'.$$

6. VALUE DISTRIBUTION OF d_n . DEFINITION OF POLIGNAC NUMBERS. CONDITIONAL RESULTS ABOUT POLIGNAC NUMBERS

The present section can be considered a natural continuation of the previous one, although the problems of this section were not raised by Erdős, though he mentioned for example in [28, 1981] a weaker form of Polignac’s conjecture (see below): “An old (and at present hopeless) conjecture states that d_n assumes all even values.”

It is unclear when any written form of the twin prime conjecture appeared. At any rate, the following generalization of it was formulated by de Polignac in 1849 [73, 1849].

Polignac’s Conjecture. *For every positive even number n there are infinitely many prime gaps of size n .*

Definition. Let us call an even number n a Polignac number, if n can be written in infinitely many ways as the difference of two consecutive primes.

It is trivial that the existence of at least one Polignac number is equivalent with the

Bounded Gap Conjecture. $\liminf_{n \rightarrow \infty} d_n < \infty$.

In fact, the value $C_0 = \liminf_{n \rightarrow \infty} d_n$ is definitely the smallest Polignac number P_0 if the \liminf is finite.

In view of this, a reformulation of Theorem 4 is the following

Theorem 4’ ([38, 2009]). *If the primes have a level $\vartheta > 1/2$ of distribution, then with a suitable constant $C(\vartheta)$ we have at least one Polignac number*

$$(6.1) \quad P_0 \leq C(\vartheta).$$

Furthermore, assuming the Elliott–Halberstam conjecture (or even $\vartheta > 0.971$) we have

$$(6.2) \quad P_0 \leq 16.$$

However, the original work [38, 2009] does not imply the existence of more than one Polignac number. In contrast to this, I recently showed

Theorem 13 ([70, 2010]). *If the primes have a level of distribution $\vartheta > 1/2$, then we have at least*

$$(6.3) \quad c'(\vartheta)N$$

Polignac numbers below N .

This can be further improved, and in fact, I can show that under the same assumption $\vartheta > 1/2$ not only primes have bounded differences infinitely often, but the difference between consecutive Polignac numbers is uniformly bounded.

Theorem 14. *If the primes have a level $\vartheta > 1/2$ of distribution, then there exists a $C'(\vartheta)$ such that every interval of type*

$$(6.4) \quad [M, M + C'(\vartheta)]$$

contains at least one Polignac number. In other words, if $\{P_i\}_{i=1}^\infty$ denotes the sequence of Polignac numbers, then we have for all n

$$(6.5) \quad P_{n+1} - P_n \leq C'(\vartheta).$$

7. COMPARISON OF TWO CONSECUTIVE VALUES OF d_n

Erdős formulated (and proved) many interesting assertions about two or more consecutive values of d_n . In a joint work with Turán they showed (see [32, 1948]) that

$$(7.1) \quad d_{n+1} - d_n$$

changes sign infinitely often. After this Erdős [19, 1948] proved that

$$(7.2) \quad \liminf_{n \rightarrow \infty} \frac{d_{n+1}}{d_n} < 1 < \limsup_{n \rightarrow \infty} \frac{d_{n+1}}{d_n}.$$

He mentioned in [22, 1955]: “One would of course conjecture that

$$(7.3) \quad \liminf \frac{d_{n+1}}{d_n} = 0 \quad \text{and} \quad \limsup \frac{d_{n+1}}{d_n} = \infty$$

but these conjectures seem very difficult to prove.”

Since there were no further developments in the past almost 60 years, the following conditional result might be of some interest.

Theorem 15. *If primes have a level $\vartheta > 1/2$ of distribution, then*

$$(7.4) \quad \liminf_{n \rightarrow \infty} \frac{d_{n+1}/d_n}{(\log n)^{-1}} \leq C_5(\vartheta),$$

and

$$(7.5) \quad \limsup_{n \rightarrow \infty} \frac{d_{n+1}/d_n}{\log n} \geq c_6(\vartheta).$$

As one can see, the above results are much stronger than the original conjecture (7.2); however, we need the deep unproved condition $\vartheta > 1/2$ for the level of distribution of primes.

We remark that beyond (7.2), Erdős [22, 1955] also formulated the conjecture that d_{n+1}/d_n is everywhere dense in $(0, \infty)$. (Cf. Section 5 about these problems.) He mentioned that he and Ricci proved that the set of limit points of d_{n+1}/d_n has a positive measure. This result can be proved by the same method which yielded the result that the set of limit points of $d_n/\log n$ has a positive measure.

We can extend this method (also conditionally) to show the following

Theorem 16. *Let us suppose that primes have a level $\vartheta > 1/2$ of distribution and let $\varepsilon > 0$ be arbitrary. Let I denote the set of limit points of $\left\{ \frac{d_n}{d_{n+1}} \right\}_{n=1}^{\infty}$ while I^* that of $\left\{ \frac{d_{n+1}}{d_n} \right\}_{n=1}^{\infty}$. Then the Lebesgue measures of both sets*

$$(7.6) \quad [0, \varepsilon] \cap I \quad \text{and} \quad [0, \varepsilon] \cap I^*$$

are positive.

8. COMPARISON OF ℓ CONSECUTIVE VALUES OF d_n FOR $\ell > 2$

Extending the problem of sign changes of $d_{n+1} - d_n$ (cf. (7.1)) Erdős and Turán [32, 1948] asked for a necessary and sufficient condition that

$$(8.1) \quad \sum_{i=1}^k a_i p_{n+i}$$

should have infinitely many sign changes as n runs through the sequence of natural numbers. They observed that $\sum_{i=1}^k a_i = 0$ is clearly a necessary condition (by the Prime Number Theorem). Furthermore, Erdős mentions [23, 1972] that Pólya observed that if (8.1) changes sign infinitely often, then the numbers

$$(8.2) \quad \alpha_j = \sum_{i=1}^j a_i$$

cannot all have the same sign. Thus in the following, we will always suppose

$$(8.3) \quad \alpha_k := \sum_{i=1}^k a_i = 0, \quad k = \ell + 1, \quad \ell \geq 2$$

and investigate

$$(8.4) \quad \sum_{i=1}^{\ell} \alpha_i d_{n+i} \quad \left(= - \sum_{i=1}^k a_i p_{n+i} \right).$$

Erdős writes: “It would be reasonable to conjecture that Pólya’s condition is necessary and sufficient for (8.4) to change sign infinitely often. Unfortunately the proof of this is not likely to succeed at the present stage of science.” After this he shows the much weaker result that (8.4) changes sign infinitely often if

$$(8.5) \quad \sum_{i=1}^{\ell} \alpha_i = 0, \quad \alpha_{\ell} \neq 0.$$

It is certainly hopeless to prove the above very deep conjecture of Erdős, which for the case $\ell = 2$ is equivalent with his conjecture (7.3). As an

approximation to it I was able to show a common generalization of Erdős's results (7.3) and (8.5) and to reach a conditional result in the case where primes have a level of distribution greater than $1/2$. Finally, I observed that the full conjecture can be shown supposing the incredibly deep Hardy–Littlewood's prime k -tuple conjecture (cf. (3.17)), which clearly supports the truth of Erdős's conjecture.

Theorem 17. *Let $\ell \geq 2$, $c_0(\ell)$ be a sufficiently small explicitly calculable constant depending on ℓ , and suppose that the real numbers α_i ($i = 1, 2, \dots, \ell$) are not all zero and satisfy*

$$(8.6) \quad \left| \sum_{i=1}^{\ell} \alpha_i \right| \leq c_0(\ell) \sum_{i=1}^{\ell} |\alpha_i|.$$

Then the expression

$$(8.7) \quad \sum_{i=1}^{\ell} \alpha_i d_{n+i}$$

changes sign infinitely often as n runs through all natural numbers.

Theorem 18. *Let $\alpha_1, \dots, \alpha_\ell$ be real numbers with the property that there is one $j \in [1, \ell]$ such that*

$$(8.8) \quad \sum_{i=1, i \neq j}^{\ell} |\alpha_i| < |\alpha_j|, \quad \operatorname{sgn} \alpha_i \neq \operatorname{sgn} \alpha_j, \quad i \in [1, \ell] \setminus \{j\}.$$

If primes have a level $\vartheta > 1/2$ of distribution, then (8.7) changes sign infinitely often as $n = 1, 2, 3, \dots$

Theorem 19. *If the Hardy–Littlewood's prime k -tuple conjecture is true for $k = \ell$, then Erdős's conjecture is true for $k = \ell$ too, that is, (8.7) changes sign infinitely often as $n = 1, 2, 3, \dots$*

We remark here that the original prime k -tuple conjecture (3.17) does not refer to consecutive differences.

In Paul Turán's obituary, Erdős [27, 1980] expresses a more general conjecture: "We never could prove that $d_{n+2} > d_{n+1} > d_n$ has infinitely many solutions, and in fact we could not even prove that at least one of the set of inequalities $d_n > d_{n+1} > d_{n+2}$ or $d_n < d_{n+1} < d_{n+2}$ has infinitely many solutions. In fact, there is no doubt that all the $k!$ orderings between d_n, \dots, d_{n+k-1} will occur, but this is again probably beyond our powers."

In connection with this, I remark that the Hardy–Littlewood prime k -tuple conjecture gives a positive answer for the above general problem and actually for all problems mentioned in the present work, due to the following

Theorem 20. *If the Hardy–Littlewood prime k -tuple conjecture is true for a given k , then it is also true for the same k , if we only count the numbers $n \leq x$ for which $n + h_1, n + h_2, \dots, n + h_k$ are consecutive primes.*

9. THE CONJECTURES OF ERDŐS AND ERDŐS–MIRSKY ON CONSECUTIVE VALUES OF NUMBER THEORETIC FUNCTIONS. THE PARITY PHENOMENON

The problems mentioned here have seemingly nothing to do with the difference of primes. However, both the motivation behind the formulation of the conjectures and their proofs involved problems about the existence of certain configurations of almost primes. These problems can be solved by methods similar to that leading to short gaps between consecutive primes (cf. (3.6) and [38, 2009]). In the following let $d(n)$ denote the number of divisors of n , $\Omega(n)$ the number of prime divisors of n counted with multiplicity and $\omega(n)$ the number of prime divisors of n without multiplicity. 60 years ago Erdős and Mirsky [30, 1952] formulated

Conjecture A1. $d(n) = d(n + 1)$ *infinitely often.*

Erdős also formulated the analogous conjectures for $\Omega(n)$ and $\omega(n)$.

Conjecture A2. $\Omega(n) = \Omega(n + 1)$ *infinitely often.*

Conjecture A3. $\omega(n) = \omega(n + 1)$ *infinitely often.*

These conjectures follow from some analogues of the twin prime conjecture. A well-known conjecture states that there are infinitely many primes p such that $2p + 1$ is also a prime. These primes are called Sophie Germain primes. Jing-Run Chen's celebrated result [8, 1966], [9, 1973] states that there are infinitely many primes p with

$$(9.1) \quad p + 2 \in \mathcal{P} \quad \text{or} \quad p + 2 = p_1 p_2, \quad p_1, p_2 \in \mathcal{P}.$$

Similarly to this, his method shows the infinitude of primes p satisfying

$$(9.2) \quad 2p + 1 \in \mathcal{P} \quad \text{or}$$

$$(9.3) \quad 2p + 1 = p_1 p_2, \quad p_1, p_2 \in \mathcal{P}.$$

It is generally believed that both (9.2) and (9.3) hold for infinitely many primes p . This conjecture is analogous to the twin prime conjecture and

it is generally believed to be true and to have the same level of difficulty. Now, (9.3) immediately implies the truth of A1–A3, since

$$(9.4) \quad d(2p) = d(2p + 1) = 4, \quad \omega(2p) = \Omega(2p) = \omega(2p + 1) = \Omega(2p + 1) = 2.$$

Due to these connections, Conjectures A1–A3 were also considered extremely difficult, if not hopeless.

It was therefore a great surprise for Erdős [29, 1986] when C. Spiro [88, 1981] showed

$$(9.5) \quad d(n) = d(n + 5040) \text{ infinitely often.}$$

Three years later Heath-Brown [51, 1984] succeeded in proving A1, the original Erdős–Mirsky conjecture, by using both Spiro’s approach as well as other new ideas of his own. His method also worked for A2, but not for A3. Conjecture A3 was solved only 20 years later by Schlage-Puchta [87, 2003], using somewhat similar ideas combined with computer calculations.

All these proofs used sieve methods and all the solutions (both n and $n + 1$) of Conjectures A1–A3, produced by the mentioned proofs, were almost primes, that is, numbers with a bounded number of prime factors. A common feature of all proofs was that they could not prove any of the Conjectures A1–A3 with an a priori given value (not even an a priori given parity) of the corresponding functions $d(n)$, $\Omega(n)$ or $\omega(n)$, respectively. The phenomenon which prevents us in showing the twin prime conjecture or (9.3), for example, is called the parity phenomenon and was described first by Selberg (cf. Chapter 4 of [46, 2001]). The parity phenomenon can be formulated with some simplification that sieve methods cannot differentiate between integers with an even and an odd number of prime factors. Despite the ‘parity barrier’, the mentioned approaches were successful, since, as Heath-Brown [51, 1984] commented on Spiro’s proof of (9.5), “Thus one does not know the value of $\Omega(n)$ for the particular n which satisfies $d(n) = d(n + 5040)$. In this way, one sidesteps the ‘parity problem’.”

In a series of joint works with D. Goldston, S. W. Graham and C. Yıldırım, we succeeded in generalizing the original ideas of our method [38, 2009] leading to short prime gaps for gaps between semi-primes, that is, numbers with exactly two different prime divisors. In particular, we showed [40, 2009] that if q_n denotes the n^{th} semiprime, then

$$(9.6) \quad \liminf_{n \rightarrow \infty} (q_{n+1} - q_n) \leq 6,$$

thereby breaking the parity barrier in this special case.

More generally, we showed the following weaker form of Dickson’s conjecture [15, 1904, cf. Section 3] for semiprimes.

Theorem 21 ([40, 2009]). *If $L_i(n) = a_i n + b_i$ ($1 \leq i \leq 3$) is a triple of linear forms such that their product has no fixed prime divisor, then there are infinitely many integers n such that at least two of the values $L_i(n)$ ($1 \leq i \leq 3$) are semiprimes.*

Remark. Taking the triple $\{n, n + 2, n + 6\}$ we obtain (9.6).

With the aid of the above theorem we could find a universal method to show Conjectures A1–A3 in a stronger form where we could almost freely choose the common value of the number-theoretic functions $d(n)$, $\Omega(n)$ and $\omega(n)$ respectively, apart from the case when the prescribed value is too small. For example, we are not able to solve (9.4). Our method is also able to guarantee in many cases the validity of Conjectures A1–A3 simultaneously for the same pair $(n, n + 1)$.

Theorem 22 ([41, 2011]). *Given any integer $B \geq 0$ we have infinitely many integers n such that*

$$(9.7) \quad \omega(n) = \omega(n + 1) = 4 + B, \quad \Omega(n) = \Omega(n + 1) = 5 + B, \\ d(n) = d(n + 1) = 24 \cdot 2^B.$$

We also showed

Theorem 23 ([41, 2011]). *$\omega(n) = \omega(n + 1) = 3$ for infinitely many integers n .*

Theorem 24 ([41, 2011]). *$\Omega(n) = \Omega(n + 1) = 4$ for infinitely many integers n .*

10. THE DISTRIBUTION OF THE DIFFERENCE OF PRIMES

The problem discussed in the present section probably does not appear in the 1600 papers of Erdős. However, it is closely related to the distribution of d_n (often investigated by Erdős, as one can see from the preceding sections). The difference is that now we are generally interested in the distribution of even numbers which can be written as the difference of two primes

$$(10.1) \quad m = p' - p'', \quad p', p'' \in \mathcal{P}.$$

This question is related to at least six conjectures (one of them, Polignac's conjecture was discussed already in Section 6).

Conjecture C1 (de Polignac [73, 1849]). *Every even number can be written in infinitely many ways as the difference of two consecutive primes (cf. Section 6).*

I did not find a reference for the following conjecture, but Erdős [28, 1981] mentions it as follows.

Conjecture C2. *An old (and at present hopeless) conjecture states that d_n assumes all even values.*

Conjecture C3 (Kronecker [60, 1901]). *Every even number can be written in infinitely many ways as the difference of two primes.*

Conjecture C4 (Maillet [66, 1905]). *Every even number is the difference of two primes.*

Conjecture C5 (Bounded Gap Conjecture). *We have infinitely many bounded gaps between consecutive primes, that is,*

$$(10.2) \quad \liminf_{n \rightarrow \infty} d_n < \infty.$$

Finally, we mention Goldbach's conjecture from 1742.

Conjecture C6 (Goldbach (cf. [34, 1965])). *Every even number larger than two can be written as the sum of two primes.*

This is almost completely analogous to Maillet's conjecture C4.

The existence of at least one Polignac number or the existence of at least one Kronecker number is clearly equivalent to the Bounded Gap Conjecture, which we proved [38, 2009] under the assumption that primes have a level of distribution $\vartheta > 1/2$ (cf. Section 3).

Furthermore, in Section 6 under the same assumption we showed that we have bounded gaps between consecutive Polignac numbers.

The present section will therefore be devoted to Conjecture C4, that is, to Maillet numbers which we define as those even integers which can be written as the difference of two primes (cf. (10.1)) analogously to Goldbach numbers. Due to the similarity with Goldbach numbers, one gets the impression that the results which we can show for Goldbach and Maillet numbers are completely analogous. In fact, until very recently, essentially all results for both Goldbach and Maillet numbers could be proved for each other.

In fact, this was the case with the size of exceptional sets for both problems or the case with unconditional results about gaps between consecutive Goldbach or Maillet numbers, respectively.

For example, the result that there are Goldbach numbers in intervals of type

$$(10.3) \quad [x, x + x^{21/800}] \quad \text{for } x > x_0,$$

can be transferred without any problem to Maillet numbers as well. ((10.3) follows from the works of Baker–Harman–Pintz [2, 2001] on the existence of primes in short intervals and from another one of Ch. Jia [57, 1996] on the existence of primes in almost all short intervals.) Together with J. Kaczorowski and A. Perelli, [58, 1995] we showed under the assumption of the Generalized Riemann Hypothesis that almost all even integers are Maillet (or Goldbach) numbers in intervals of type

$$(10.4) \quad [x, x + (\log x)^{C_0}] \quad \text{for } C_0 > 6, \quad x > x_0.$$

On the other hand under the weaker assumption of the classical Riemann Hypothesis it was shown much earlier by Linnik [62, 1952] that Maillet (or Goldbach) numbers appear in every interval of type

$$(10.5) \quad [x, x + (\log x)^{C_1}] \quad \text{for } x > x_0,$$

if $C_1 > 3$. This was shown by Kátai [59, 1967] to be true with $C_1 = 2$ (again on RH). Both works refer to the sum of two primes, but the proofs also work for the difference of two primes.

Very recently I succeeded in showing [71, 2012] unconditionally a bound of type (10.5) with a larger value of C_1 , thereby improving significantly the former best result (10.3).

Theorem 25. *There exists an explicitly calculable C_1 (C_1 is about 40) such that every interval of type (10.5) contains at least one Maillet number, that is, a number expressible as the difference of two primes.*

The paper [71, 2012] does not contain the proof of (10.5), but the following weaker form of it is shown there.

Theorem 26. *Let $\varepsilon > 0$ be arbitrary. If $x > x_0(\varepsilon)$, then the interval*

$$(10.6) \quad [x, x + x^\varepsilon]$$

contains a difference of two primes.

This bound significantly supersedes the former best result (10.3), valid unconditionally both for Goldbach and Maillet numbers. It is interesting to note that the method is not able to show unconditionally the existence of Goldbach numbers in intervals of type (10.5) or even in those of type (10.6).

Added in proof (19.05.2013)

Yitang Zhang (Bounded Gaps Between Primes, manuscript) proved very recently a fantastic new result which shows the existence of infinitely many gaps of size at most $7 \cdot 10^7$ between consecutive primes. He does not prove that primes have a distribution level greater than $1/2$, but nevertheless (cf. our present Theorem 8) he shows that any admissible k -tuple of size at least $3.5 \cdot 10^6$ contains infinitely often at least two primes (which implies the gap size $7 \cdot 10^7$). Using either the above mentioned second statement of Zhang's work about primes in k -tuples or in some cases the method of his proof, the earlier proved Theorems 9–10 of the author [see [70], 2010] and the announced Theorems 11–16 of the present work will be unconditionally valid with absolute constants (instead of constants depending on the hypothetical distribution level of primes). These absolute constants are explicitly calculable in case of Theorems 9–10, 13 and 15–16 and they are ineffective in case of Theorems 11, 12 and 14.

REFERENCES

- [1] R. J. Backlund, *Über die Differenzen zwischen den Zahlen, die zu den ersten n Primzahlen teilerfremd sind. Commentationes in honorem E. L. Lindelöf.* Annales Acad. Sci. Fenn. **32** (1929), Nr. 2, 1–9.
- [2] R. C. Baker, G. Harman, J. Pintz, *The difference between consecutive primes, II.* Proc. London Math. Soc. (3) **83** (2001), no. 3, 532–562.
- [3] E. Bombieri, *On the large sieve.* Mathematika **12** (1965), 201–225.
- [4] E. Bombieri, H. Davenport, *Small differences between prime numbers.* Proc. Roy. Soc. Ser. A **293** (1966), 1–18.
- [5] J. Bourgain, *On triples in arithmetic progressions.* Geom. Funct. Anal. **9** (1999), no. 5, 968–984.
- [6] J. Bourgain, *Roth's theorem on progressions revisited.* J. Anal. Math. **104** (2008), 155–192.
- [7] A. Brauer, H. Zeitz, *Über eine zahlentheoretische Behauptung von Legendre.* Sber. Berliner Math. Ges. **29** (1930), 116–125.
- [8] Chen Jing Run, *On the representation of a large even integer as the sum of a prime and the product of at most two primes.* Kexue Tongbao **17** (1966), 385–386 (Chinese).
- [9] Chen Jing Run, *On the representation of a large even integer as the sum of a prime and the product of at most two primes.* Sci. Sinica **16** (1973), 157–176.
- [10] J. G. van der Corput, *Sur l'hypothèse de Goldbach pour presque tous les nombres pairs.* Acta Arith. **2** (1937), 266–290.
- [11] J. G. van der Corput, *Über Summen von Primzahlen und Primzahlquadraten.* Math. Ann. **116** (1939), 1–50.

- [12] H. Cramér, *Prime numbers and probability*. 8. Skand. Math. Kongr., Stockholm, 1935, 107–115.
- [13] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*. Acta Arith. **2** (1936), 23–46.
- [14] N. G. Čudakov, *On the density of the set of even numbers which are not representable as a sum of two primes*. Izv. Akad. Nauk. SSSR **2** (1938), 25–40.
- [15] L. E. Dickson, *A new extension of Dirichlet's theorem on prime numbers*. Messenger of Math. (2), **33** (1904), 155–161.
- [16] P. D. T. A. Elliott, H. Halberstam, *A conjecture in prime number theory*. Symposia Mathematica Vol. 4 (1970) (INDAM, Rome, 1968/69), 59–72, Academic Press, London.
- [17] P. Erdős, *On the difference of consecutive primes*. Quart. J. Math. Oxford ser. **6** (1935), 124–128.
- [18] P. Erdős, *The difference of consecutive primes*. Duke Math. J. **6** (1940), 438–441.
- [19] P. Erdős, *On the difference of consecutive primes*. Bull. Amer. Math. Soc. (1948), 885–889.
- [20] P. Erdős, *On a new method in elementary number theory which leads to an elementary proof of the prime number theorem*. Proc. Nat. Acad. Sci. U.S.A. **35** (1949), 374–384.
- [21] P. Erdős, *Problems and results on the differences of consecutive primes*. Publ. Math. Debrecen **1** (1949), 33–37.
- [22] P. Erdős, *Some problems on the distribution of prime numbers*. Teoria dei Numeri, Math. Congr. Varenna, 1954, 8 pp., 1955.
- [23] P. Erdős, *Some problems on Consecutive Prime Numbers*. Mathematika **19** (1972), 91–95.
- [24] P. Erdős, *Résultats et problèmes en théorie des nombres*. Séminaire Delongé–Pisot–Poitou (14e année: 1972/73), Théorie des nombres, Fasc. 2. Exp. No. 24, 7 pp. Secrétariat Mathématique, Paris, 1973.
- [25] P. Erdős, *Problems and results on number theoretic properties of consecutive integers and related questions*. Proceedings of the Fifth Manitoba Conference on Numerical Mathematics (Univ. Manitoba, Winnipeg, Man., 1975), Congress Numer. XVI, pp. 25–44, Utilitas Math., Winnipeg, Man., 1976.
- [26] P. Erdős, *Problems in number theory and combinatorics*. Proceedings of the Sixth Manitoba Conference on Numerical Mathematics (Univ. Manitoba, Winnipeg, Man., 1976), Congress. Numer. XVIII, pp. 35–58, Utilitas Math., Winnipeg, Man., 1977.
- [27] P. Erdős, *Some personal reminiscences of the mathematical work of Paul Turán*. Acta Arith. **37** (1980), 3–8.
- [28] P. Erdős, *Many old and on some new problems of mine in number theory*, Proceedings of the Tenth Manitoba Conference on Numerical Mathematics and Computing, Vol. I (Univ. Manitoba, Winnipeg, Man., 1980), Congress. Numer. **30** (1981), 3–27.
- [29] P. Erdős, *Some problems on Number Theory, in: Analytic and elementary number theory (Marseille, 1983)*. Publ. Math. Orsay, 86-1, pp. 53–67, Univ. Paris XI, Orsay, 1986.

- [30] P. Erdős, L. Mirsky, *The distribution of values of the divisor function $d(n)$* . Proc. London Math. Soc. (3) **2** (1952), 257–271.
- [31] P. Erdős, P. Turán, *On some sequences of integers*. J. London Math. Soc. **11** (1936), 261–264.
- [32] P. Erdős, P. Turán, *On some new questions on the distribution of prime numbers*. Bull. Amer. Math. Soc. **54** (1948), 371–378.
- [33] T. Estermann, *On Goldbach's problem: Proof that almost all even positive integers are sums of two primes*. Proc. London Math. Soc. (2) **44** (1938), 307–314.
- [34] L. Euler, C. Goldbach, *Briefwechsel 1729–1764*. Akademie Verlag, Berlin, 1965.
- [35] H. Furstenberg, *Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions*. J. Analyse Math. **31** (1977), 204–256.
- [36] D. A. Goldston, C. Yıldırım, *Higher correlations of divisor sums related to primes. III. Small gaps between primes*. Proc. London Math. Soc. (3) **95** (2007), no. 3, 653–686.
- [37] D. A. Goldston, J. Pintz, C. Yıldırım, *Primes in Tuples III: On the difference $p_{n+\nu} - p_n$* . Funct. Approx. Comment. Math. **35** (2006), 79–89.
- [38] D. A. Goldston, J. Pintz, C. Yıldırım, *Primes in tuples. I*. Ann. of Math. (2) **170** (2009), no. 2, 819–862.
- [39] D. A. Goldston, J. Pintz, C. Yıldırım, *Primes in tuples. II*. Acta Math. **204** (2010), no. 1, 1–47.
- [40] D. A. Goldston, S. W. Graham, J. Pintz, C. Y. Yıldırım, *Small gaps between products of two primes*. Proc. London Math. Soc. (3) **98** (2009), no. 3, 741–774.
- [41] D. A. Goldston, S. W. Graham, J. Pintz, C. Y. Yıldırım, *Small gaps between almost primes, the parity problem, and some conjectures of Erdős on consecutive integers*. Int. Math. Res. Not. IMRN 2011, no. 7, 1439–1450.
- [42] W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*. Geom. Funct. Anal. **8** (1998), no. 3, 529–551.
- [43] W. T. Gowers, *A new proof of Szemerédi's theorem*. Geom. Funct. Anal. **11** (2001), no. 3, 465–588.
- [44] A. Granville, *Unexpected irregularities in the distribution of prime numbers, in: Proceedings of the International Congress of Mathematicians (Zürich, 1994)*. Vol. 1, 2, 388–399, Birkhäuser, Basel, 1995.
- [45] A. Granville, *Harald Cramér and the Distribution of Prime Numbers*. Scand. Actuarial J. No. 1 (1995), 12–28.
- [46] G. Greaves, *Sieves in Number Theory*. Springer, 2001.
- [47] B. Grein, T. Tao, *The primes contain arbitrarily long arithmetic progressions*. Ann. of Math. (2) **167** (2008), no. 2, 481–547.
- [48] G. H. Hardy, J. E. Littlewood, *Some problems of 'Partitio Numerorum', III: On the expression of a number as a sum of primes*. Acta Math. **44** (1923), 1–70.
- [49] G. H. Hardy, J. E. Littlewood, *Some problems of 'Partitio Numerorum', V: A further contribution to the study of Goldbach's problem*. Proc. London Math. Soc. (2) **22** (1924), 46–56.

- [50] D. R. Heath-Brown, *Three primes and an almost prime in arithmetic progression*. J. London Math. Soc. (2) **23** (1981), no. 3, 396–414.
- [51] D. R. Heath-Brown, *The divisor function at consecutive integers*. Mathematika **31** (1984), 141–149.
- [52] D. R. Heath-Brown, *Integer sets containing no arithmetic progressions*. J. London Math. Soc. (2) **35** (1987), no. 3, 385–394.
- [53] A. J. Hildebrand, H. Maier, *Irregularities in the distribution of primes in short intervals*. J. Reine Angew. Math. **397** (1989), 162–193.
- [54] G. Hoheisel, *Primzahlprobleme in der Analysis*. SBer. Preuss. Akad. Wiss., Berlin (1930), 580–588.
- [55] M. N. Huxley, *On the differences of primes in arithmetical progressions*. Acta Arith. **15** (1968/69), 367–392.
- [56] M. N. Huxley, *Small differences between consecutive primes II*. Mathematika **24** (1977), 142–152.
- [57] Chaohua Jia, *Almost all short intervals containing prime numbers*. Acta Arith. **76** (1996), 21–84.
- [58] J. Kaczorowski, A. Perelli, J. Pintz, *A note on the exceptional set for Goldbach's problem in short intervals*. Monatsh. Math. **116**, no. 3–4 (1993), 275–282. Corrigendum: *ibid.* **119** (1995), 215–216.
- [59] I. Kátai, *A remark on a paper of Ju. V. Linnik*. Magyar Tud. Akad. Mat. Fiz. Oszt. Közl. **17** (1967), 99–100.
- [60] L. Kronecker, *Vorlesungen über Zahlentheorie, I*. p. 68, Teubner, Leipzig, 1901.
- [61] E. Landau, *Gelöste und ungelöste Probleme aus der Theorie der Primzahlverteilung und der Riemannschen Zetafunktion*. Jahresber. Deutsche Math. Ver. **21** (1912), 208–228. [Proc. 5th Internat. Congress of Math., **I**, 93–108, Cambridge 1913; Collected Works, **5**, 240–255, Thales Verlag.]
- [62] Yu. V. Linnik, *Some conditional theorems concerning the binary Goldbach problem*. Izv. Akad. Nauk. SSSR **16** (1952), 503–520 (Russian).
- [63] H. Maier, *Chains of large gaps between consecutive primes*. Adv. in Math. **39** (1981), no. 3, 257–269.
- [64] H. Maier, *Small differences between prime numbers*. Michigan Math. J. **35** (1988), 323–344.
- [65] H. Maier, C. Pomerance, *Unusually large gaps between consecutive primes*. Trans. Amer. Math. Soc. **322** (1990), 201–237.
- [66] E. Maillet, *L'intermédiaire des math.* **12** (1905), p. 108.
- [67] W. Narkiewicz, *The Development of Prime Number Theory. From Euclid to Hardy and Littlewood*. Springer, 2000.
- [68] J. Pintz, *Very large gaps between consecutive primes*. J. Number Th. **63** (1997), 286–301.
- [69] J. Pintz, *Cramér vs. Cramér. On Cramér's probabilistic model for primes*. Funct. Approx. Comment. Math. **37** (2007), part 2, 361–376.

- [70] J. Pintz, *Are there arbitrarily long arithmetic progressions in the sequence of twin primes?* An irregular mind, Bolyai Soc. Math. Stud. 21, Springer, 2010, pp. 525–559.
- [71] J. Pintz, *On the difference of primes.* arXiv: 1206.0149 [math.NT]
- [72] J. Pintz, *Are there arbitrarily long arithmetic progressions in the sequence of twin primes? II.* Proceedings of the Steklov Institute **276** (2012), 227–232.
- [73] A. de Polignac, *Six propositions arithmologiques déduites de crible d'Ératosthène.* Nouv. Ann. Math. **8** (1849), 423–429.
- [74] R. A. Rankin, *The difference between consecutive prime numbers.* J. London Math. Soc. **13** (1938), 242–244.
- [75] R. A. Rankin, *The difference between consecutive prime numbers. II.* Proc. Cambridge Philos. Soc. **36** (1940), 255–266.
- [76] R. A. Rankin, *The difference between consecutive primes. III.* J. London Math. Soc. **22** (1947), 226–230.
- [77] R. A. Rankin, *The difference between consecutive prime numbers, V.* Proc. Edinburgh Math. Soc. (2) **13** (1962/63), 331–332.
- [78] G. Ricci, *Ricerche aritmetiche sui polinomi, II. (Intorno a una proposizione non vera di Legendre).* Rend. Palermo **58** (1934), 190–208.
- [79] G. Ricci, *La differenza di numeri primi consecutivi.* Rendiconti Sem. Mat. Univ. e Politecnico Torino **11** (1952), 149–200. Corr. ibidem **12** (1953), p. 315.
- [80] G. Ricci, *Sull'andamento della differenza di numeri primi consecutivi.* Riv. Mat. Univ. Parma **5** (1954), 3–54.
- [81] K. F. Roth, *Sur quelques ensembles d'entiers.* C.R. Acad. Sci. Paris **234** (1952), 388–390.
- [82] K. F. Roth, *On certain sets of integers.* J. London Math. Soc. **28** (1953), 104–109.
- [83] T. Sanders, *On certain sets of integers.* arXiv: 1007.5444 [math.NT]
- [84] T. Sanders, *On Roth's theorem on progressions.* Ann. of Math. (2) **174** (2011), no. 1, 619–636.
- [85] A. Schönhage, *Eine Bemerkung zur Konstruktion grosser Primzahllücken.* Arch. Math. Basel **14** (1963), 29–30.
- [86] A. Selberg, *An elementary proof of the prime-number theorem.* Ann. of Math. (2) **50** (1949), 305–313.
- [87] J.-C. Schlage-Puchta, *The equation $\omega(n) = \omega(n + 1)$.* Mathematika **50** (2003), no. 1–2, 99–101 (2005).
- [88] C. Spiro, *Thesis.* Urbana, 1981.
- [89] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression.* 1970 Number Theory (Colloq. János Bolyai Math. Soc. Debrecen, 1968), pp. 197–204, North-Holland, Amsterdam.
- [90] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression.* Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104.
- [91] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression.* Acta Arith. **27** (1975), 199–245.

- [92] E. Szemerédi, *Regular partitions of graphs*. Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976) Colloq. Internat. CNRS 260, CNRS, Paris, 1978, pp. 399–401.
- [93] E. Szemerédi, *Integer sets containing no arithmetic progressions*. Acta Math. Hungar. **56** (1990), no. 1–2, 155–158.
- [94] A. I. Vinogradov, *The density hypothesis for Dirichlet L-series*. Izv. Akad. Nauk. SSSR **29** (1965), 903–934 (Russian). Corr.: ibidem, **30** (1966), 719–720.
- [95] I. M. Vinogradov, *Representation of an odd number as a sum of three prime numbers*. Doklady Akad. Nauk. SSSR **15** (1937), 291–294 (Russian).
- [96] I. M. Vinogradov, *Special Variants of the Method of Trigonometric Sums*. Nauka, Moskva, 1976 (Russian).
- [97] Wang Yuan, Xie Sheng-gang, Yu Kun-rui, *Remarks on the difference of consecutive primes*. Sci. Sinica **14** (1965), 786–788.
- [98] E. Westzynthius, *Über die Verteilung der Zahlen, die zu der n ersten Primzahlen teilerfremd sind*. Comm. Phys. Math. Helsingfors (5) **25** (1931), 1–37.

János Pintz

*Alfréd Rényi Institute of Mathematics,
Hungarian Academy of Sciences,
Budapest, Reáltanoda u. 13–15, H-1053,
Hungary*

e-mail: `pintz.janos@renyi.mta.hu`

PAUL ERDŐS AND THE RISE OF STATISTICAL THINKING IN ELEMENTARY NUMBER THEORY

PAUL POLLACK and CARL POMERANCE*

1. INTRODUCTION

It might be argued that elementary number theory began with Pythagoras who noted two-and-a-half millennia ago that 220 and 284 form an amicable pair. That is, if $s(n)$ denotes the sum of the proper divisors of n (“proper divisor” means $d \mid n$ and $1 \leq d < n$), then

$$s(220) = 284 \quad \text{and} \quad s(284) = 220.$$

When faced with remarkable examples such as this it is natural to wonder how special they are. Through the centuries mathematicians tried to find other examples of amicable pairs, and they did indeed succeed. But is there a formula? Are there infinitely many? In the first millennium of the common era, Thâbit ibn Qurra came close with a formula for a subfamily of amicable pairs, but it is far from clear that his formula gives infinitely many examples and probably it does not.

A special case of an amicable pair m, n is when $m = n$. That is, $s(n) = n$. These numbers are called perfect, and Euclid came up with a formula for some of them (and perhaps all of them) that probably inspired that of Thâbit for amicable pairs. Euler showed that Euclid’s formula covers all even perfect numbers, but we still don’t know if Euclid’s formula gives infinitely many examples and our knowledge about odd perfects, even whether any exist, remains rudimentary.

These are colorful and attractive problems from antiquity, but what is a modern mathematician to make of them? Are they just curiosities? After

*The authors would like to thank Enrique Treviño and the anonymous referee for their helpful suggestions. The second author was supported in part by NSF grant DMS-1001180.

all, not all problems are good. Ancient people wondered why and how the planets wandered through the stellar constellations, and such musings became the foundation of astronomy, trigonometry, and modern physics. They also wondered why the sun and moon are the same apparent size, with such musings leading nowhere!

Euclid also studied another special subset of the natural numbers: the primes. Already he had a proof of their infinitude. Euler was able to quantify the reciprocal sum for primes in $[1, x]$ as $x \rightarrow \infty$, and so we had the birth of a statistical viewpoint in number theory. This led to the prime-counting conjectures of Gauss and Lagrange, the estimates of Chebyshev, the provocative outline of Riemann, and the proofs of Hadamard, de la Vallée Poussin, Erdős, and Selberg. There is a great story here which we feel sure will be told in another essay.

So we have a prime number theorem, but is there a perfect number theorem, an amicable number theorem, and others of this sort? By asking such questions about the statistical distribution of special sets of numbers one opens the door to a host of interesting problems in which modern mathematicians can participate in this millennia-old quest. And leading the way was Paul Erdős.

2. DISTRIBUTION

The function s defined in the Introduction partitions the positive integers into 3 sets: those n with $s(n) < n$, those with $s(n) = n$, and those with $s(n) > n$. Perhaps, it is not so natural to consider such a partition, but it is historically correct, going back thousands of years. Numbers with $s(n) < n$ are called *deficient* and those with $s(n) > n$ are called *abundant*, with the case of equality already met as the perfect numbers. Putting these concepts into modern garb, we have the immediate question of asymptotic density. It is clear at least that the lower density of the abundant numbers is positive, since any multiple of 6 that is larger than 6 is abundant. But it is not so clear that the abundant numbers possess an asymptotic density.¹

In 1933, Davenport [5] resolved the problem by proving that the sets of abundant numbers and deficient numbers each possesses a positive asymptotic density, while the set of perfect numbers has asymptotic density 0. In fact, Davenport proved a much more general theorem. Let σ denote the sum-of-divisors function, so that $\sigma(n) = s(n) + n$. And let $h(n) := \frac{\sigma(n)}{n}$.

¹It is also clear that the deficient numbers have positive lower density since it is easy to see that $s(n)/n$ has mean value $\pi^2/6 - 1$, which is smaller than 1.

So, for example, n is perfect when $h(n) = 2$ and abundant when $h(n) > 2$. Davenport's result is the following:

Theorem 1. For each real number u , let $\mathcal{D}(u) = \{n \in \mathbf{N} : h(n) \leq u\}$. The set $\mathcal{D}(u)$ always possesses an asymptotic density. Denoting this density by $D(u)$, the function $D(u)$ is continuous and strictly increasing for $u \geq 1$. Moreover, $D(1) = 0$ and $\lim_{u \rightarrow \infty} D(u) = 1$.²

Since $D(u)$ is continuous, it follows immediately that the perfect numbers have density zero. We subsequently deduce that the deficient numbers have density $D(2)$, where $0 < D(2) < 1$, and that the abundant numbers comprise a set of density $1 - D(2)$. The numerical values of these densities were investigated by Behrend [2, 3], who succeeded in showing that the density of the abundant numbers lies between 0.241 and 0.314. Later authors (Salié [58], Wall [62], and Deléglise [6]) have tightened these bounds; the current state of the art, due to Kobayashi [42], is that the density of the abundants has decimal expansion starting with 0.2476.

Davenport's proof of this result was decidedly analytic, requiring a study of the complex moments of the function $h(n)$. In this respect, he was following a model laid down by Schoenberg [59], who had earlier proved the analogue of Theorem 1 for the closely-related function $n/\varphi(n)$, where φ is Euler's function. The non-elementary nature of Davenport's argument would surely have irked Erdős, and in the mid-1930s, Erdős took it upon himself to give a purely arithmetic proof of Theorem 1. This resulted in a series of three papers [9, 11, 12], culminating in what we now know as the sufficiency half of the Erdős–Wintner Theorem (see [30]), one of the foundational results in the field known as *probabilistic number theory*. Studying distribution functions eventually led to the landmark collaboration of Erdős and Kac and their celebrated theorem on the normal distribution of the number of prime factors of an integer. As these subjects are discussed elsewhere in this volume, we do not dwell on them here, but rather return to the theme of elementary number theory.

3. AMICABLES

Recall from the Introduction that a pair n, m of positive integers is said to be amicable if $s(n) = m$ and $s(m) = n$, with the perfect numbers corresponding to the degenerate case of $n = m$. We have seen that the perfect numbers have asymptotic density 0, but do the amicable?

²Davenport did not prove that $D(u)$ is strictly increasing; this was established a few years later by Schoenberg [60].

A first approach to counting amicable numbers is suggested by the following simple observation: Suppose that n and m form an amicable pair, with $n < m$. Then $s(n) = m > n$, so that n is abundant. Thus, the upper density of the amicable numbers is at most twice the density of the abundant numbers, and so from [6] or [42], the upper density of the amicables is smaller than $\frac{1}{2}$.

When one considers that essentially none of the theory of amicable pairs was used in this argument, this result seems quite respectable!

In fact, all we used above was that the smaller member of a non-perfect amicable pair is abundant. An equally simpleminded observation, dual to the first, is that the larger member is deficient. Putting these together, we see that if n is the smaller member of a non-perfect amicable pair, then n is an abundant number for which $s(n)$ is deficient. Erdős had the great insight that this two-fold condition on n should be quite restrictive. His argument in [15] that the amicable numbers have asymptotic density zero is actually a proof of the following theorem:

Theorem 2. *The set of abundant natural numbers n for which $s(n)$ is deficient has asymptotic density zero.*

Erdős's proof of Theorem 2 is naturally split into three identifiable components. The first of these is an immediate consequence of the continuity of the function $D(u)$ appearing in Davenport's Theorem 1.

Lemma 3. *Let $\epsilon > 0$ be arbitrary. For a certain $\delta > 0$, depending on ϵ , the set of solutions n to*

$$2 < h(n) < 2 + \delta$$

has asymptotic density less than ϵ .

For every positive integer n , the bijection between divisors d of n and their co-divisors n/d permits us to write $h(n) = \frac{1}{n} \sum_{d|n} d = \sum_{d|n} \frac{1}{d}$. This expression for $h(n)$ suggests that the small divisors of n play the largest role in determining the size of $h(n)$. To make this precise, we let $y > 0$, and we define the truncated function

$$h_y(n) := \sum_{\substack{d|n \\ d \leq y}} \frac{1}{d}.$$

The second leg on which Erdős's argument rests is the following lemma.

Lemma 4. *Let $x > 0$ and let y be a positive integer. For each $\delta > 0$ the number of $n \leq x$ for which $h(n) - h_y(n) \geq \delta$ does not exceed $\delta^{-1}x/y$.*

Proof. A simple interchange of the order of summation shows that

$$\sum_{n \leq x} (h(n) - h_y(n)) = \sum_{d > y} \frac{1}{d} \sum_{\substack{n \leq x \\ d|n}} 1.$$

The inner sum here is at most $\frac{x}{d}$, from which it is easy to see that the entire sum is at most $\frac{x}{y}$. The claim follows immediately. ■

It seems likely that Erdős would have considered the key innovation in the proof of Theorem 2 to be its third component, which we formulate as follows.

Lemma 5. *Fix $y > 0$. For all natural numbers n outside of a set of asymptotic density zero, n and $s(n)$ share the same set of divisors in $[1, y]$.*

Proof. Let M be the least common multiple of the natural numbers not exceeding y . It suffices to show that $\sigma(n)$ is a multiple of M unless n belongs to a set of density zero. Indeed, if $M \mid \sigma(n)$, then the relation $s(n) = \sigma(n) - n$ implies that

$$s(n) \equiv -n \pmod{d}$$

for all $d \leq y$. Thus, $d \mid s(n)$ if and only if $d \mid n$. Now if p is a prime that exactly divides n , then $p + 1$ divides $\sigma(n)$. Thus, M divides $\sigma(n)$ whenever there is a prime $p \equiv -1 \pmod{M}$ for which $p \parallel n$. For any particular prime $p \equiv -1 \pmod{M}$, we see that $p \parallel n$ precisely when n falls into one of the $(p - 1)$ residue classes $p, 2p, 3p, \dots, (p - 1)p \pmod{p^2}$. So if the relation $p \parallel n$ fails for all $p \leq z$ from the residue class $-1 \pmod{M}$, then n avoids $p - 1$ residue classes modulo p^2 for every such p . By the Chinese remainder theorem, this restricts n to a set of asymptotic density

$$\prod_{\substack{p \equiv -1 \pmod{M} \\ p \leq z}} \left(1 - \frac{1}{p} + \frac{1}{p^2} \right).$$

This product can be made arbitrarily small by taking z sufficiently large, since by Dirichlet, the sum of the reciprocals of the primes $p \equiv -1 \pmod{M}$ diverges. The lemma follows. ■

Remark. The proof of Lemma 5 shows that for a fixed M , the number $\sigma(n)$ is almost always divisible by M . When M is a power of 2, this was previously observed by Kanold [40], who used this to prove that the amicable numbers have upper density less than 0.204.

It is now a simple matter to assemble Lemmas 3–5 to prove Theorem 2.

Proof of Theorem 2. Let n denote a generic abundant natural number for which $s(n)$ is deficient. We will show that for each fixed $\epsilon > 0$, the set of all such n has upper density smaller than 2ϵ . By Lemma 3, we may fix $\delta > 0$ small enough so that the set of solutions to $2 < h(n) < 2 + \delta$ has density less than ϵ . Thus, discarding a set of density less than ϵ , we can assume that

$$h(n) \geq 2 + \delta.$$

We now apply Lemma 4 with

$$y := \left\lceil \frac{1}{\delta\epsilon} \right\rceil$$

and find that discarding a set of upper density bounded by ϵ , we can assume that

$$h_y(n) > h(n) - \delta \geq 2.$$

Discarding a further set of density zero, we can assume (by Lemma 5) that n and $m = s(n)$ have the same set of divisors up to y . But then

$$h(m) \geq h_y(m) = h_y(n) > 2,$$

contradicting that m is deficient. So n must have belonged to one of the exceptional sets described above, which have combined upper density smaller than 2ϵ . ■

In the introduction to [15], Erdős asserted that his method, suitably refined, would show that the count $A(x)$ of amicable numbers in $[1, x]$ satisfies

$$(1) \quad A(x) \ll \frac{x}{\log \log \log x}.$$

Details appeared twenty years later in joint work with Rieger [28] (cf. Rieger's weaker solo result [57]). The Erdős–Rieger upper bound was soon improved by Pomerance [53], who established that

$$(2) \quad A(x) \leq x / \exp(c(\log_3 x \log_4 x)^{1/2})$$

for a certain constant $c > 0$ and all large x (note the subscripts indicate iterated logs). In both cases, what is actually estimated is the count of abundant $n \leq x$ for which $s(n)$ is deficient. (The key innovation in [53] is

the use of Erdős's theory of *primitive abundant numbers*; see [8].) A few years later, and by different methods, Pomerance [54] established the bound

$$A(x) \leq x / \exp(c(\log x \log_2 x)^{1/3}),$$

for some positive constant c and all large x . This bound has not yet been improved, nor do we know that there are infinitely many amicable numbers. Erdős has a heuristic argument suggesting that $A(x) > x^{1-o(1)}$ as $x \rightarrow \infty$.

Fix $\epsilon > 0$. Arguing as in the proof of Theorem 2, one finds that for almost all natural numbers n , we have $h(s(n)) > h(n) - \epsilon$. In the concluding remarks to [15], Erdős claimed that the complementary inequality $h(s(n)) < h(n) + \epsilon$ also holds for almost all n . A proof of this last result eventually appeared in joint work with Granville, Pomerance, and Spiro (see [22, Theorem 5.1]). Hence, $h(s(n)) = h(n) + o(1)$, as $n \rightarrow \infty$ in a set of asymptotic density 1. For another application of their method of proof, see [51].

4. SOCIABLES

One can revisit the definition of an amicable pair from the viewpoint of function iteration. Let $s_k(n)$ denote the k th iterate of $s(n)$. Then n is amicable precisely when $s_2(n) = n$. Generalizing, we say that n is *k-sociable* if $s_k(n) = n$ but $s_j(n) \neq n$ for $1 \leq j < k$, and we call the set $\{n, s(n), \dots, s_{k-1}(n)\}$ an *aliquot k-cycle*. Note that the 1-sociable numbers are exactly the perfect numbers, whose distribution is discussed in detail in the next section.

Questions about the iterates of $s(n)$ began to be asked at the end of the 19th century. For a natural number n , the *aliquot sequence at n* is the sequence $n, s(n), s_2(n), \dots$, where we stop if we reach 0. For instance, the aliquot sequence at 24 is 24, 36, 55, 17, 1, 0, while the aliquot sequence at 25 is 25, 6, 6, 6, \dots . In 1888, Catalan [4] proposed the empirical theorem that these two examples exhaust the possible behaviors of an aliquot sequence; more precisely, every aliquot sequence either terminates or hits a perfect number.

'Empirical theorems', like champion athletes, are always in danger of losing their title. Soon after Catalan's conjecture appeared, Perrott [47] pointed out that the aliquot sequence at 220 was a counterexample. This led Dickson [7] to propose a somewhat tamer, modified conjecture – commonly known today as the *Catalan–Dickson conjecture* – that all aliquot sequences terminate or are eventually periodic. This has been verified numerically for

$n < 276$. However, when $n = 276$, more than 1700 terms of the sequence have been computed [64], with no end in sight.

When Dickson put forward his modified conjecture in 1913, no aliquot cycles of length > 2 were known. The first examples, of lengths 5 and 28, were given by Poulet in 1918. Currently there are 217 such cycles known [45], all but 11 of which have length 4.

What can we prove about the distribution of these cycles? The first asymptotic result on this problem is due to Erdős [21]. Note that the case $k = 2$ is contained in Erdős's earlier work on amicable pairs.

Theorem 6. *Fix $\epsilon > 0$ and fix an integer $k \geq 2$. Then for all n outside of a set of asymptotic density zero, we have*

$$(3) \quad h(s_j(n)) > h(n) - \epsilon \quad \text{for all } 0 < j < k.$$

One consequence of Theorem 6 is that for each fixed k , almost all abundant numbers are k -times abundant: $n < s(n) < s_2(n) < \cdots < s_k(n)$. Suppose now that n is the smallest member of a sociable k -cycle, where $k > 1$. Then n is abundant, but not k -times abundant (since $s_k(n) = n$), and so n belongs to a set of density zero. As a corollary, the set of k -sociable numbers has asymptotic density zero for each fixed k . For quantitative results of this kind, see [43] and [49].

The proof of Theorem 6 employs the same reasoning seen in the previous section, but with Lemma 5 replaced by the following generalization.

Lemma 7. *Fix $y > 0$, and fix $k \geq 2$. For all natural numbers n outside of a set of asymptotic density zero, all of $n, s(n), \dots, s_{k-1}(n)$ share the same set of divisors in $[1, y]$.*

One can ask whether Theorem 6 remains true with (3) replaced by the complementary inequality $h(s_j(n)) < h(n) + \epsilon$. As mentioned above, this is known to be so when $k = 2$, by later work of Erdős et al. [22]. For larger values of k , this constitutes an attractive open problem. Note that the claim of a general proof, made in [21], is retracted in [22].

For more recent developments on sociable numbers, see [43]. For example, it is shown there that if one lumps together all sociable numbers (i.e., one takes the union of the k -sociables over all k), then after discarding a certain set of asymptotic density zero, the remaining elements are all both odd and abundant.

5. PERFECTS

From Euclid and Euler, we know that an even number is perfect precisely when it can be written as $2^{p-1}(2^p - 1)$, where $2^p - 1$ is prime. Thus, the distribution of the even perfect numbers is inextricably linked with the distribution of primes of the form $2^p - 1$, known as *Mersenne primes*. While almost nothing is known rigorously about the distribution of Mersenne primes, Lenstra, Pomerance, and Wagstaff have (independently) given heuristic arguments suggesting that probably

$$\#\{p \leq x : 2^p - 1 \text{ prime}\} \sim \frac{e^\gamma}{\log 2} \log x, \quad \text{as } x \rightarrow \infty.$$

Here γ is the familiar Euler–Mascheroni constant. (See, for example, [61].) The validity of this conjecture would imply that the count of even perfect numbers up to x is asymptotic to $\frac{e^\gamma}{\log 2} \log \log x$.

What about odd perfect numbers? We have already noted that from Davenport’s Theorem 1, these numbers have asymptotic density zero. But this is a rather weak result. There is a short and pretty argument of Hornfeck [38] showing that in fact, the count $P(x)$ of odd perfects in $[1, x]$ is smaller than $x^{1/2}$, for every $x > 1$. We cannot resist reproducing it here. By a classical result of Euler, we can write an odd perfect n as $n = p^e m^2$ where p is a prime not dividing m and $p \equiv e \equiv 1 \pmod{4}$. (This uses only that n is odd and $\sigma(n) \equiv 2 \pmod{4}$.) Since n is perfect,

$$2p^e m^2 = \sigma(p^e)\sigma(m^2), \quad \text{so that} \quad \frac{2m^2}{\sigma(m^2)} = \frac{\sigma(p^e)}{p^e}.$$

But the fraction $\sigma(p^e)/p^e$ is already in lowest terms, since the numerator $\sigma(p^e) = 1 + p + \dots + p^e$ is not divisible by p . Hence, the prime power p^e is uniquely determined from m . If we assume that $n \leq x$, then $1 < m \leq \sqrt{x}$, and so Hornfeck’s bound follows.

The problem of obtaining improved bounds for $P(x)$ attracted some attention in the late 1950s, with several number theorists throwing their hats into the ring. It was Erdős [16] who gave the first significant improvement over Hornfeck’s bound, getting $P(x) \leq x^{1/2-c}$ for a certain $c > 0$ and all large x . His idea is both ingenious and, at least in hindsight, quite natural. We sketch an improvement that obtains the estimate $P(x) \leq x^{1/4+o(1)}$. (A result of this same quality was obtained by Kanold [41] shortly after Erdős’s paper appeared.)

Erdős's starting point is the following 'greedy' algorithm for extracting from an integer M a divisor D of M with D coprime to both M/D and $\sigma(D)$:

Algorithm:

```

Factor  $M = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ , where  $p_1 > p_2 > \cdots > p_k$ .
 $D \leftarrow 1$  // Initialize
for  $i = 1$  to  $k$  do // Loop over prime power divisors of  $M$ 
if  $\gcd(\sigma(p_i^{e_i} D), p_i^{e_i} D) = 1$  then
  |  $D \leftarrow p_i^{e_i} D$ 
end
return  $D$ 

```

In certain special cases, Erdős proved that the output D of this algorithm is bounded below by a fixed power of the input M . However, for our present purposes, the argument is clearer (and stronger) if it is instead made to rest upon the following near-injectivity property, whose proof – given in [52] – involves the same circle of ideas as in [16].

Proposition 8. *Let $\epsilon > 0$. For all sufficiently large values of x , depending on the choice of ϵ , at most x^ϵ inputs $M \leq x$ of the Algorithm correspond to the same output D .*

We now show that $P(x) \leq x^{1/4+o(1)}$ as $x \rightarrow \infty$. Write an odd perfect number $n \leq x$ as $p^e m^2$ as above and apply the Algorithm to $M = m^2$. It produces a divisor D of m^2 coprime to m^2/D and to $\sigma(D)$. Thus $D = d^2$ for some $d \mid m$. Letting v^2 be the co-divisor of d^2 in m^2 , we have $n = p^e v^2 d^2$. Since n is perfect, we have

$$2p^e v^2 d^2 = \sigma(n) = \sigma(p^e v^2) \sigma(d^2).$$

Since d^2 is coprime to $\sigma(d^2)$, we have $d^2 \mid \sigma(p^e v^2)$. If $p^e v^2 \leq x^{1/2}$, then $d^2 \leq 2x^{1/2}$ so that $d < 2x^{1/4}$. But if $p^e v^2 > x^{1/2}$, then $d^2 = n/(p^e v^2) < x^{1/2}$, so in either case, $d < 2x^{1/4}$. So, by Proposition 8 there are at most $x^{1/4+\epsilon}$ inputs m^2 to the Algorithm (for each fixed $\epsilon > 0$ and x sufficiently large depending on ϵ). But by the Hornfeck–Euler argument, m^2 determines n , which proves the theorem that $P(x) \leq x^{1/4+o(1)}$ as $x \rightarrow \infty$.

A year after Erdős's article appeared, Hornfeck and Wirsing [39] published a proof that $P(x) \leq x^{o(1)}$ as $x \rightarrow \infty$. Two years later, Wirsing [63] showed that for an absolute constant W , one has $P(x) < x^{W/\log \log x}$ for all $x > e$. In fact, the same is true for the distribution of those n with

$\sigma(n)/n = r$ for any fixed rational number r . Wirsing's upper bound has not been improved in fifty years, but it is still a rather long way from the widespread belief that $P(x)$ is identically zero.

While Erdős's results on $P(x)$ are now primarily of historical interest, his approach to the problem has borne other fruit. For instance, as Erdős noted at the time in [16], one can use these methods to show that n and $\sigma(n)$ rarely have a large common factor. For a detailed discussion of these problems, see [50], which was written in part to correct and substantiate some of the unproved assertions of [16]. See also [52].

6. ITERATION

It was not always the case, but we now view functions as interesting mathematical objects in and of themselves. For example, for a function whose values are contained in its domain, we can view the function as creating a dynamical system. We discussed this above in the context of the function s , the sum-of-proper-divisors function, where we have sociable cycles and the Catalan–Dickson conjecture.

Euler's function φ provides another attractive dynamical system. Given a positive integer n and the sequence $n, \varphi(n), \varphi(\varphi(n)), \dots$, we note that it is strictly decreasing until it reaches 1. Thus, we may define $k(n)$ as the minimal number $k \geq 1$ of iterates necessary for n to reach 1. For example, $k(13) = 4$, since the sequence is 13, 12, 4, 2, 1, 1, \dots . Seemingly a very exotic function, there is some unexpected structure here! Let $k^*(n) = k(n)$ for n even and $k^*(n) = k(n) - 1$ for n odd. It is not hard to see that $k^*(n)$ is completely additive ($k^*(mn) = k^*(m) + k^*(n)$ for all m, n) and it is inductively defined on the primes by $k^*(2) = 1$ and $k^*(p) = k^*(p - 1)$ for $p > 2$. Erdős and his collaborators show in [22] that under the assumption of the Elliott–Halberstam conjecture (a widely believed conjecture on the distribution of primes in residue classes) there is a positive constant α such that $k(n) \sim \alpha \log n$ as $n \rightarrow \infty$ on a set of asymptotic density 1.

Euler chains $n, \varphi(n), \varphi(\varphi(n)), \dots$ arise in other contexts, for example, primality testing and algebraic number theory. See the very recent paper of Ford [32] and the references therein.

7. VALUES

The set of values of an arithmetic function can also give rise to interesting questions. Take the function s . If p, q are different primes, then $s(pq) =$

$p + q + 1$. So a slightly stronger form of Goldbach's conjecture, namely all even numbers at least 8 are a sum of two distinct primes, implies that all odd numbers at least 9 are in the image of s . Since $s(2) = 1$, $s(4) = 3$, and $s(8) = 7$, presumably the only odd number missing from the image of s is 5. From what we know about the possible exceptional set in Goldbach's conjecture, it follows that the set of odds not in the form $s(n)$ has asymptotic density 0. But what of even numbers? Here, Erdős in [20] showed by a clever argument that a positive proportion of even numbers are missing from the image of s . We still don't know if the image of s has a density or if the range of s contains a positive proportion of even numbers. The issue of numbers of the form $n - \varphi(n)$ was also raised in [20], but here even less is known. See [56] for a recent paper in this area with references to other work.

Here is a proof of the result in [20] that a positive proportion of even numbers are missing from the image of s . If $s(n)$ is even and n is odd, then $\sigma(n)$ must be odd too, and so n is a square, say m^2 . If $s(m^2) \leq x$ and q is the least prime factor of m , then $x \geq s(m^2) > m^2/q$. If m is composite, then $q \leq m^{1/2}$, so that $m^{3/2} < x$ and there are at most $x^{2/3}$ possibilities. If $m = q$ is prime, then $q < x$ and there are at most $\pi(x) = O(x/\log x)$ possibilities. Hence the number of even numbers $s(n)$ in $[1, x]$ with n odd is $o(x)$ as $x \rightarrow \infty$. So we may assume that n is even, which in turn implies that $x \geq s(n) \geq n/2$. Hence $n \leq 2x$. Consider values of s in $[1, x]$ that are divisible by 12. By Lemma 5, but for $o(x)$ choices for $n \leq 2x$, we may assume that $12 \mid n$. Thus, $x \geq s(n) \geq \frac{4}{3}n$, so that $n \leq \frac{3}{4}x$. We conclude that the number of values of $s(n) \leq x$ divisible by 12 is at most $\frac{1}{12} \cdot \frac{3}{4}x + o(x) \sim \frac{1}{16}x$, leaving asymptotically at least 25% of the multiples of 12 not in the range of s .

In 1929 S. S. Pillai [48] proved that the image of Euler's function φ has density 0. Here is the idea of the proof. For each fixed positive integer k consider numbers n with at most k distinct prime factors. It is easy to see that the set of these numbers has density 0 as does their image under φ . But if n is not in this set, then $2^k \mid \varphi(n)$, so we see that the image of φ has upper density at most 2^{-k} . Since k is arbitrary, this proves that the image of φ has density 0. Pillai was able to quantify this result by taking k as a function of x and obtaining an estimate of $O(x/(\log x)^{\frac{1}{e} \log 2})$ for the number of values of φ in $[1, x]$. Since φ is 1-1 on the primes, we immediately have a lower bound of magnitude $x/\log x$.

So what is the correct exponent here?

Erdős's answer: "1." This was in [10], a wonderful and seminal paper submitted to the Quarterly Journal of Mathematics when he was 21. That is, the number of values of φ in $[1, x]$ is $x/(\log x)^{1+o(1)}$ as $x \rightarrow \infty$. The idea is to look not only at the number of factors 2 in $\varphi(n)$, but at the total

number of prime factors. If $\Omega(n)$ is the number of prime factors of n counted with multiplicity, Erdős knew after Hardy and Ramanujan that normally $\Omega(n) \sim \log \log n$. Moreover, exceptional numbers with $\Omega(n) < \epsilon \log \log x$ or $\Omega(n) > \frac{1}{\epsilon} \log \log x$ are so sparse that they are negligible. Erdős then showed (in an early and inventive use of Brun's sieve method) an analog of the Hardy–Ramanujan theorem for “shifted primes”, that is, he showed that $\Omega(p-1)$ is normally near $\log \log p$, with exceptional primes p , with $\Omega(p-1)$ far from this normal order, being quite rare. So, but for very few numbers n , they are divisible by a fair number of non-exceptional primes p . Since $\Omega(\varphi(n)) \geq \sum_{p|n} \Omega(\phi(p))$, we find that $\Omega(\varphi(n))$ is much larger than $\log \log n$, meaning that $\phi(n)$ is quite exceptional! This is all worked out in exquisite detail, not only solving Pillai's problem, but introducing extraordinarily useful tools in the statistical study of elementary number theory.

The problem of the distribution of φ values was taken up later by Erdős and Hall [23, 24], Maier and Pomerance [46], and by Ford [31]. However, we still don't have an asymptotic formula nor do we know if a natural one exists.

The same theorems carry over to the range of σ . Erdős also raised the attractive question (for instance, in [18]) of whether the images of φ and σ have an infinite intersection. If p and $p+2$ are both primes, then $\sigma(p) = p+1 = \varphi(p+2)$, so the answer is affirmative if there are infinitely many twin primes. Also if $2^p - 1$ is prime, then $\sigma(2^p - 1) = 2^p = \varphi(2^{p+1})$, so the answer is again ‘yes’ if there are infinitely many Mersenne primes (and so ‘yes’ if there are infinitely many even perfect numbers). In a recent paper, Ford, Luca, and Pomerance [33] showed unconditionally that there are infinitely many pairs of integers m, n with $\sigma(m) = \varphi(n)$, and Ford and Pollack [34, 35] have some finer results in this direction.

8. ORDER

Euler's function $\varphi(n)$ gives the order of the multiplicative group $(\mathbf{Z}/n\mathbf{Z})^*$. A closely related function, $\lambda(n)$ gives the maximal order of an element in this group. When $(\mathbf{Z}/n\mathbf{Z})^*$ is cyclic, we have $\lambda(n) = \varphi(n)$. We always have $\lambda(n) \mid \varphi(n)$, and since $(\mathbf{Z}/n\mathbf{Z})^*$ is abelian, for all integers a coprime to n , $a^{\lambda(n)} \equiv 1 \pmod{n}$. For this reason, $\lambda(n)$ is referred to as the *universal exponent function*.

Carmichael used the notation λ , but the function appears in Gauss a century earlier. It is easy to give a formula for $\lambda(n)$ based on the prime factorization of n : for a prime power p^α , we have $\lambda(p^\alpha) = \varphi(p^\alpha)$ except

if $p = 2$ and $\alpha \geq 3$ in which case, $\lambda(2^\alpha) = \frac{1}{2}\varphi(2^\alpha)$. (Note that $\varphi(p^\alpha) = p^{\alpha-1}(p-1)$.) Further, for all n , $\lambda(n)$ is the lcm of $\lambda(p^\alpha)$ for prime powers $p^\alpha \mid n$.

Being so closely related to φ , one might expect that statistically λ is quite similar. Here is φ 's story: We know (from Schoenberg, or more generally the Erdős–Wintner theorem) that for each real number $u \in (0, 1]$, the set $\{n : \varphi(n) \leq un\}$ has a positive asymptotic density that varies continuously and strictly monotonically with u . Further, from Mertens' theorem in analytic number theory, it follows that $\varphi(n) \geq (e^{-\gamma} + o(1))n/\log \log n$ as $n \rightarrow \infty$. And on average, $\varphi(n)$ behaves like cn , with $c = 6/\pi^2$.

Erdős took up the normal and average orders of $\lambda(n)$ in [17], stating some results without proof. Full proofs of more precise results, including the minimal order of $\lambda(n)$, were worked out in Erdős–Pomerance–Schmutz [27] in 1991. The function is amazingly different from φ . On average it is not like cn , but rather like $n/(\log n)^{1+o(1)}$, where the “ $o(1)$ ” is asymptotically $c/\log \log \log n$, with c explicitly worked out. The normal order is not of the shape $\asymp n$, but rather much smaller at $n/(\log n)^{\log \log \log n + c + o(1)}$ for a different explicit c . And the minimal order, instead of the large function $n/\log \log n$, is instead the tiny function $(\log n)^{c \log \log \log n}$ (here the precise value of c is still not known), a result that has found application in the analysis of some primality tests. These results have not been improved over the past 2 decades, and there is indeed room for improvement. For example, does $\lambda(n)$ have a “nice” distribution function? That is, for $\varphi(n)$ we compare it with n ; what should $\lambda(n)$ be compared with?

The image of λ is also different than the image of φ . In [27] it is shown that there is some $c > 0$ such that the number of λ -values in $[1, x]$ is $O(x/(\log x)^c)$, a result which strongly uses an earlier result of Erdős and Wagstaff in [29]. It has been announced by Luca and Pomerance that there is some $c' > 0$ such that the count is at least $x/(\log x)^{1-c'}$ for all large x . Probably the truth is $x/(\log x)^{\alpha+o(1)}$ as $x \rightarrow \infty$, where $\alpha = 1 - (1 + \log \log 2)/\log 2 = 0.086\dots$, the Erdős–Tenenbaum–Ford constant, and maybe this is provable.

The iteration of λ also has its surprises, see Harland [37] for some recent work.

From its definition, we see that λ is related to the order-of-an-element function. For n a positive integer and $\gcd(a, n) = 1$, we follow Erdős in using the notation $\ell_a(n)$ for the order of a in $(\mathbf{Z}/n\mathbf{Z})^*$. Thus, $\ell_a(n) \mid \lambda(n)$, and for some number a we have $\ell_a(n) = \lambda(n)$. In a surprisingly difficult paper, Erdős in [19] (he spoke on this at the International Conference of Mathematicians in Nice in 1970), began the statistical study of $\ell_a(n)$. Further developments can be tracked in [25] and in [44].

A pseudoprime to the base a is a composite integer n for which $a^{n-1} \equiv 1 \pmod{n}$. Note that the congruence holds if and only if $\ell_a(n) \mid n-1$. Pseudoprimes are a useful concept since all primes n not dividing a satisfy the congruence and the congruence itself is easily checkable numerically. Thus, pseudoprimes stand as an obstruction against using the congruence as a primality test. Known from experience that pseudoprimes are rare compared with primes, it took some time for this to be proved. Erdős was the first to do so in [14] (announced earlier in [13]). Currently the best upper bound known for their distribution is in Pomerance [55], and a number of other statistical results are discussed in Erdős–Pomerance [26].

Some composites n have the property that $a^{n-1} \equiv 1 \pmod{n}$ for all integers a coprime to n . From what we have said above, this congruence is equivalent to $\lambda(n) \mid n-1$. It is easy to see that this then forces n to be squarefree. In 1899, Korselt essentially gave this criterion for a number n to satisfy $a^{n-1} \equiv 1 \pmod{n}$ for all a coprime to n , but did not give any composite examples. In 1910 and apparently unaware of Korselt's criterion, Carmichael did give some examples, such as 561, 1105, and 1729. Now known as Carmichael numbers, Erdős was the first to prove a result about their distribution, in [17]. He showed that the number of Carmichael numbers in $[1, x]$ is at most $x^{1-c \log \log \log x / \log \log x}$ for some fixed $c > 0$. And he gave a heuristic argument that the count exceeds $x^{1-\epsilon}$ for each fixed $\epsilon > 0$ and all sufficiently large x depending on ϵ .

This was all the more remarkable in that at that time we did not have a proof that there are infinitely many Carmichael numbers and the numerical evidence seemed to indicate a much slower growth rate for the counting function. Shanks was notably vocal in challenging Erdős on this point. It is now known that there are infinitely many Carmichael numbers, Alford–Granville–Pomerance [1]. The proof largely follows the Erdős heuristic in [17], which in turn is based on a proof in [10] that there are numbers $v \leq x$ such that $\varphi(n) = v$ has more than x^c solutions n . In Granville–Pomerance [36] the two incompatible viewpoints of Erdős and Shanks were shown to both have elements of truth, though there is still much to be learned here.

9. CONCLUSION

We have touched on a few of our favorite problems and results of Erdős in elementary number theory, particularly those involving the elementary number theoretic functions. We have not attempted to be encyclopedic, and for a more thorough and complete treatment, we recommend the article of

Schinzel in this volume, as well as the original papers of Erdős, most of which are freely available online.

The point we have tried to make is that viewing classical problems in elementary number theory through a statistical lens allows the tools of modern mathematics to prove interesting and sometimes beautiful theorems. It is through this lens that the mathematics of the ancients lives on. Paul Erdős was an early and consistent exponent of this point of view, changing for the better the entire landscape of elementary number theory.

REFERENCES

- [1] W. R. Alford, A. Granville, and C. Pomerance, *There are infinitely many Carmichael numbers*, Ann. of Math. (2) **139** (1994), 703–722.
- [2] F. Behrend, *Über numeri abundantes. I*, Sitzgsber. Akad. Berlin (1932), 322–328.
- [3] ———, *Über numeri abundantes. II*, Sitzgsber. Akad. Berlin (1933), 289–293.
- [4] E. Catalan, *Propositions et questions diverses*, Bull. Soc. Math. France **16** (1888), 128–129.
- [5] H. Davenport, *Über numeri abundantes*, Sitzgsber. Akad. Berlin (1933), 830–837.
- [6] M. Deléglise, *Bounds for the density of abundant integers*, Experiment. Math. **7** (1998), 137–143.
- [7] L. E. Dickson, *Theorems and tables on the sum of the divisors of a number*, Q.J. Math. **44** (1913), 264–296.
- [8] P. Erdős, *On primitive abundant numbers*, J. London Math. Soc. **10** (1935), 49–58.
- [9] ———, *On the density of some sequences of numbers*, J. London Math. Soc. **10** (1935), 120–125.
- [10] ———, *On the normal number of prime factors of $p - 1$ and some related problems concerning Euler's φ -function*, Quart. J. Math. Oxford Ser. **6** (1935), 205–213.
- [11] ———, *On the density of some sequences of numbers, II*, J. London Math. Soc. **12** (1937), 7–11.
- [12] ———, *On the density of some sequences of numbers, III*, J. London Math. Soc. **13** (1938), 119–127.
- [13] ———, *On the converse of Fermat's theorem*, Amer. Math. Monthly **56** (1949), 623–624.
- [14] ———, *On almost primes*, Amer. Math. Monthly **57** (1950), 404–407.
- [15] ———, *On amicable numbers*, Publ. Math. Debrecen **4** (1955), 108–111.
- [16] ———, *On perfect and multiply perfect numbers*, Ann. Mat. Pura Appl. (4) **42** (1956), 253–258.
- [17] ———, *On pseudoprimes and Carmichael numbers*, Publ. Math. Debrecen **4** (1956), 201–206.

- [18] ———, *Remarks on number theory, II. Some problems on the σ function*, Acta Arith. **5** (1959), 171–177.
- [19] ———, *On the sum $\sum_{d|2^n-1} d^{-1}$* , Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 3, and Israel J. Math. **9** (1971), 43–48.
- [20] ———, *Über die Zahlen der form $\sigma(n) - n$ und $n - \varphi(n)$* , Elem. Math. **28** (1973), 83–86.
- [21] ———, *On asymptotic properties of aliquot sequences*, Math. Comp. **30** (1976), no. 135, 641–645.
- [22] P. Erdős, A. Granville, C. Pomerance, and C. Spiro, *On the normal behavior of the iterates of some arithmetic functions*, Analytic number theory (Allerton Park, IL, 1989), Progr. Math., vol. 85, Birkhäuser Boston, Boston, MA, 1990, pp. 165–204.
- [23] P. Erdős and R. R. Hall, *On the values of Euler's φ -function*, Acta Arith. **22** (1973), 201–206.
- [24] ———, *Distinct values of Euler's φ -function*, Mathematika **23** (1976), 1–3.
- [25] P. Erdős, P. Kiss, and C. Pomerance, *On the prime divisors of Mersenne numbers*, Acta Arith. **57** (1991), 267–281.
- [26] P. Erdős and C. Pomerance, *On the number of false witnesses for a composite number*, Math. Comp. **46** (1986), 259–279.
- [27] P. Erdős, C. Pomerance, and E. Schmutz, *Carmichael's lambda function*, Acta Arith. **58** (1991), 363–385.
- [28] P. Erdős and G. J. Rieger, *Ein Nachtrag über befreundete Zahlen*, J. Reine Angew. Math. **273** (1975), 220.
- [29] P. Erdős and S. S. Wagstaff, Jr., *The fractional parts of the Bernoulli numbers*, Illinois J. Math. **24** (1980), 104–112.
- [30] P. Erdős and A. Wintner, *Additive arithmetical functions and statistical independence*, Amer. J. Math. **61** (1939), 713–721.
- [31] K. Ford, *The distribution of totients*, Ramanujan J. **2** (1998), 67–151. (Updated version on the author's web page.)
- [32] K. Ford, *Sieving by very thin sets of primes and Pratt trees with missing primes*, preprint, 2012, [arXiv:1212.3498](https://arxiv.org/abs/1212.3498) [math.NT], IMRN, to appear.
- [33] K. Ford, F. Luca, and C. Pomerance, *Common values of the arithmetic functions ϕ and σ* , Bull. Lond. Math. Soc. **42** (2010), 478–488.
- [34] K. Ford and P. Pollack, *On common values of $\varphi(n)$ and $\sigma(m)$, I*, Acta Math. Hungarica **133** (2011), 251–271.
- [35] ———, *On common values of $\varphi(n)$ and $\sigma(m)$, II*, Algebra Number Theory **6** (2012), 1669–1696.
- [36] A. Granville and C. Pomerance, *Two contradictory conjectures concerning Carmichael numbers*, Math. Comp. **71** (2001), 883–908.
- [37] N. Harland, *The number of iterates of the Carmichael lambda function required to reach 1*, preprint, 2012, [arXiv:1203.4791](https://arxiv.org/abs/1203.4791) [math.NT].
- [38] B. Hornfeck, *Zur Dichte der Menge der vollkommenen Zahlen*, Arch. Math. (Basel) **6** (1955), 442–443.

- [39] B. Hornfeck and E. Wirsing, *Über die Häufigkeit vollkommener Zahlen*, Math. Ann. **133** (1957), 431–438.
- [40] H. J. Kanold, *Über die Dichten der Mengen der vollkommenen und der befreundeten Zahlen*, Math. Z. **61** (1954), 180–185.
- [41] ———, *Über die Verteilung der vollkommenen Zahlen und allgemeinerer Zahlenmengen*, Math. Ann. **132** (1957), 442–450.
- [42] M. Kobayashi, *On the density of abundant numbers*, Ph.D. thesis, Dartmouth College, 2010.
- [43] M. Kobayashi, P. Pollack, and C. Pomerance, *On the distribution of sociable numbers*, J. Number Theory **129** (2009), 1990–2009.
- [44] P. Kurlberg and C. Pomerance, *On a problem of Arnold: the average multiplicative order of a given integer*, Algebra and Number Theory, to appear.
- [45] D. Moews, *A list of aliquot cycles of length greater than 2*, internet resource, <http://djm.cc/sociable.txt>.
- [46] H. Maier and C. Pomerance, *On the number of distinct values of Euler's ϕ -function*, Acta Arith. **49** (1988), 263–275.
- [47] J. Perrott, *Sur une proposition empirique énoncée au Bulletin*, Bull. Soc. Math. France **17** (1889), 155–156.
- [48] S. S. Pillai, *On some functions connected with $\varphi(n)$* , Bull. Amer. Math. Soc. **35** (1929), 832–836.
- [49] P. Pollack, *A remark on sociable numbers of odd order*, J. Number Theory **130** (2010), no. 8, 1732–1736.
- [50] ———, *On the greatest common divisor of a number and its sum of divisors*, Michigan Math. J. **60** (2011), no. 1, 199–214.
- [51] ———, *Quasi-amicable numbers are rare*, J. Integer Seq. **14** (2011), no. 5, Article 11.5.2, 13 pages.
- [52] P. Pollack and C. Pomerance, *Prime-perfect numbers*, Integers **12A** (2012), article A14, 19 pages.
- [53] C. Pomerance, *On the distribution of amicable numbers*, J. Reine Angew. Math. **293/294** (1977), 217–222.
- [54] ———, *On the distribution of amicable numbers. II*, J. Reine Angew. Math. **325** (1981), 183–188.
- [55] ———, *On the distribution of pseudoprimes*, Math. Comp. **37** (1981), 587–593.
- [56] C. Pomerance and H.-S. Yang, *Variant of a theorem of Erdős on the sum-of-proper-divisors function*, Math. Comp., to appear.
- [57] G. J. Rieger, *Bemerkung zu einem Ergebnis von Erdős über befreundete Zahlen*, J. Reine Angew. Math. **261** (1973), 157–163.
- [58] H. Salié, *Über die Dichte abundanter Zahlen*, Math. Nachr. **14** (1955), 39–46.
- [59] I. J. Schoenberg, *Über die asymptotische Verteilung reeller Zahlen mod 1*, Math. Z. **28** (1928), 171–199.
- [60] ———, *On asymptotic distributions of arithmetical functions*, Trans. Amer. Math. Soc. **39** (1936), 315–330.

- [61] S. S. Wagstaff, Jr., *Divisors of Mersenne numbers*, Math. Comp. **83** (1983), 385–397.
- [62] C. R. Wall, *Density bounds for the sum of divisors function*, The theory of arithmetic functions (Proc. Conf., Western Michigan Univ., Kalamazoo, Mich., 1971), Lecture Notes in Math., vol. 251, Springer, Berlin, 1972, pp. 283–287.
- [63] E. Wirsing, *Bemerkung zu der Arbeit über vollkommene Zahlen*, Math. Ann. **137** (1959), 316–318.
- [64] P. Zimmerman, *Aliquot sequences*, internet resource,
<http://www.loria.fr/~zimmerma/records/aliquot.html>.

Paul Pollack
University of Georgia,
Department of Mathematics,
Athens, GA 30602,
USA
e-mail: pollack@uga.edu

Carl Pomerance
Dartmouth College,
Department of Mathematics,
Hanover, NH 03755,
USA
e-mail: carlp@math.dartmouth.edu

EXTREMAL RESULTS IN RANDOM GRAPHS

VOJTĚCH RÖDL* and MATHIAS SCHACHT†

Dedicated to the memory of Paul Erdős on the occasion of his 100th birthday

According to Paul Erdős [*Some notes on Turán's mathematical work*, J. Approx. Theory **29** (1980), page 4] it was Paul Turán who “created the area of extremal problems in graph theory”. However, without a doubt, Paul Erdős popularized *extremal combinatorics*, by his many contributions to the field, his numerous questions and conjectures, and his influence on discrete mathematicians in Hungary and all over the world. In fact, most of the early contributions in this field can be traced back to Paul Erdős, Paul Turán, as well as their collaborators and students. Paul Erdős also established the *probabilistic method* in discrete mathematics, and in collaboration with Alfréd Rényi, he started the systematic study of *random graphs*. We shall survey recent developments at the interface of extremal combinatorics and random graph theory.

1. EXTREMAL GRAPH THEORY

1.1. Introduction

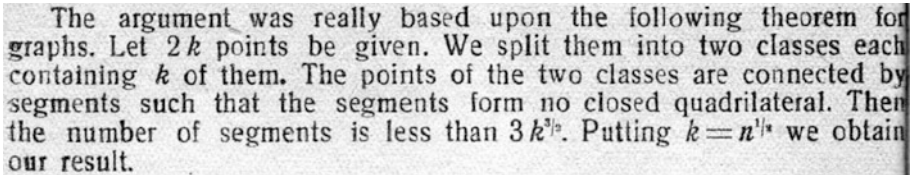
We first discuss a few classical results in extremal graph theory. Since by no means we can give a full account here, we restrict ourselves to some well known results in the area and highlight some of the pivotal questions. For a thorough introduction to the area we refer to the standard textbook of Bollobás [12].

A large part of extremal graph theory concerns the study of graphs G which do not contain a given subgraph F . The first classical problem is to maximize the number of edges of such a graph G with n vertices. An

*First author was supported by NSF grant DMS 0800070.

†Second author was supported through the Heisenberg-Programme of the Deutsche Forschungsgemeinschaft (DFG Grant SCHA 1263/4-1).

instance of this question was addressed already in 1938 by Erdős. In [28] he proved bounds for an extremal problem in combinatorial number theory, and in his proof he asserts a lemma that every n -vertex graph without a cycle of length four can have at most $cn^{3/2}$ edges (see Figure 1 below).



The argument was really based upon the following theorem for graphs. Let $2k$ points be given. We split them into two classes each containing k of them. The points of the two classes are connected by segments such that the segments form no closed quadrilateral. Then the number of segments is less than $3k^{3/2}$. Putting $k = n^{1/2}$ we obtain our result.

Fig. 1. Quote from [28, page 78]

Turán initiated the systematic study of such questions, and in Section 1.3 we give a short account of Turán's theorem [114] in graph theory and some important results related to it. In fact, we will restrict ourselves only to extremal questions in graph theory here. However, even within extremal graph theory we can only discuss a few selected results and are bound to neglect not only many important topics, but also many beautiful generalizations and improvements of those classical results. Our certainly biased selection of results presented here is guided by the recent generalizations, which were obtained for subgraphs of random graphs. First we introduce the necessary notation.

1.2. Notation

Below we recall some notation from graph theory, which will be used here. For notation not defined here we refer to the standard text books [13, 16, 27].

All graphs considered here are finite, simple and have no loops. For a graph $G = (V, E)$ we denote by $V(G) = V$ and $E(G) = E$ its *vertex set* and its *edge set*, respectively. We denote by $e(G) = |E(G)|$ the number of edges of G and by $d(G) = e(G)/\binom{|V(G)|}{2}$ its *edge density*. Moreover, for a subset $U \subseteq V$ let $e_G(U)$ be the number of edges of G contained in U . By $\omega(G)$, $\alpha(G)$, and $\chi(G)$ we denote the standard graph parameters known as *clique number*, *independence number*, and *chromatic number* of G , respectively. We say that a graph G contains a *copy* of a graph F if there is an injective map $\varphi : V(F) \rightarrow V(G)$ such that $\{\varphi(u), \varphi(v)\} \in E(G)$, whenever $\{u, v\} \in E(F)$. If G contains no such copy, then we say G is *F-free*. Also, G and F are *isomorphic* if there exists a bijection $\varphi : G \rightarrow F$ such that $\{\varphi(u), \varphi(v)\} \in E(G)$ if, and only if $\{u, v\} \in E(F)$. In this case we often write $G = F$. A graph H is a *subgraph* of $G = (V, E)$, if $V(H) \subseteq V$ and $E(H) \subseteq E$, which we denote by $H \subseteq G$.

The complete graph on t vertices with $\binom{t}{2}$ edges is denoted by K_t , and a *clique* is some complete graph. A graph G is t -partite or t -colorable, if there is a partition of its vertex set into t classes (some of them might be empty) such that every edge of G has its vertices in two different partition classes. We denote by $\text{Col}_n(t)$ the set of all t -colorable graphs on n vertices, i.e.,

$$\text{Col}_n(t) = \{H \subseteq K_n : \chi(H) \leq t\}.$$

A t -partite graph $G = (V, E)$ with vertex classes $V_1 \cup \dots \cup V_t = V$ is *complete* if for every $1 \leq i < j \leq t$ and every $u \in V_i$ and $v \in V_j$ we have $\{u, v\} \in E$. We denote by $T_{n,t}$ the complete t -partite graph on n vertices with the maximum number of edges. It is easy to show that $T_{n,t}$ is unique up to isomorphism and that it is the complete t -partite graph with every vertex class having cardinality either $\lfloor n/t \rfloor$ or $\lceil n/t \rceil$.

For a graph F with at least one edge and an integer n , we denote by $\text{Forb}_n(F)$ the set of F -free subgraphs of K_n , i.e.,

$$\text{Forb}_n(F) = \{H \subseteq K_n : H \text{ is } F\text{-free}\},$$

and we recall the *extremal function* $\text{ex}_n(F)$ defined by

$$\text{ex}_n(F) = \max\{e(H) : H \in \text{Forb}_n(F)\}.$$

Note that the set $\text{Forb}_n(F)$ is closed under taking subgraphs, i.e., if $H \in \text{Forb}_n(F)$ and $H' \subseteq H$, then $H' \in \text{Forb}_n(F)$. In general such sets of graphs are called *monotone*. In fact, any monotone property \mathcal{P}_n of subgraphs of K_n can be expressed by a family of *forbidden* subgraphs, and many results discussed below allow generalizations in this direction (and even more generally towards hereditary properties). However, we will concentrate on generalizations for subgraphs of random graphs and restrict the discussion to a forbidden set of graphs consisting of only one graph.

1.3. Turán’s Theorem and Related Results

Generalizing a result of Mantel [81] for $F = K_3$, Turán [114] determined $\text{ex}_n(F)$ when F is a complete graph.

Theorem 1.1 (Turán 1941). *For all integers $t \geq 2$ and $n \geq 1$ we have*

$$\text{ex}_n(K_{t+1}) = e(T_{n,t}).$$

Moreover, $T_{n,t}$ is, up to isomorphism, the unique K_{t+1} -free graph on n vertices with $\text{ex}_n(K_{t+1})$ edges.

Theorem 1.1 determines the maximum number of edges of a K_{t+1} -free graph on n vertices. Moreover, it characterizes the *extremal graphs*, i.e., those K_{t+1} -free graphs on n vertices having the maximum number of edges. In fact, these are instances of two very typical questions in extremal combinatorics. The questions below are stated more generally and could be applied in other contexts like hypergraphs, multigraphs, subsets of the integers, etc. However, we shall mostly restrict ourselves to questions in graph theory here.

- (Q1) Given a monotone property of discrete structures, like the monotone set $\text{Forb}_n(K_{t+1})$ of subgraphs of K_n , what maximum density can its members attain?
- (Q2) What are the extremal discrete structures, e.g., like $T_{n,t}$ is the extremal subgraph of K_n for $\text{Forb}_n(K_{t+1})$?

Theorem 1.1 answers (Q1) and (Q2) in a precise way. In fact, it not only determines the maximum density, as required for (Q1), but actually gives a full description of the function $\text{ex}_n(K_{t+1})$. Often only the density question can be addressed.

To this end for a given graph F we recall the definition of the *Turán density* $\pi(F)$, which is given by

$$\pi(F) = \lim_{n \rightarrow \infty} \frac{\text{ex}_n(F)}{\binom{n}{2}}.$$

Note that the limit indeed exists since one can show that $\text{ex}_n(F)/\binom{n}{2}$ is non-increasing in n . Erdős and Stone [40] determined $\pi(F)$ for every graph F .

Theorem 1.2 (Erdős & Stone, 1946). *For every graph F with at least one edge we have*

$$\pi(F) = 1 - \frac{1}{\chi(F) - 1}.$$

In particular, $\pi(F) = 0$ for every bipartite graph F (see also [74] for stronger estimates for this problem). On the other hand, for a graph F of chromatic number at least three the lower bound in Theorem 1.2 is established by the Turán graph $T_{n, \chi(F)-1}$.

Refining Theorem 1.2 by determining $\text{ex}_n(F)$ for arbitrary F is a very hard problem (see, e.g., [39, 105, 106] for some partial results in this direction). Consequently, a precise solution for question (Q2) is still unknown for most graphs F . Owing to the *stability theorem*, which was independently obtained by Erdős [34] and Simonovits [104], we however have an approximate answer for question (Q2). In fact, the stability theorem determines an approximate structure of the extremal, as well as the *almost extremal*, graphs up to $o(n^2)$ edges.

Theorem 1.3 (Erdős 1967, Simonovits 1968). *For every $\varepsilon > 0$ and every graph F with $\chi(F) = t + 1 \geq 3$ there exist $\delta > 0$ and n_0 such that the following holds. If H is an F -free graph on $n \geq n_0$ vertices satisfying*

$$e(H) \geq \text{ex}_n(F) - \delta n^2,$$

then there exists a copy T of $T_{n,t}$ on $V(H)$ such that

$$|E(H) \Delta E(T)| \leq \varepsilon n^2,$$

where Δ denote the symmetric difference of sets.

In other words, H can be obtained from the graph $T_{n,t}$ by adding and deleting up to at most εn^2 edges.

In particular, H can be made t -partite by removing at most εn^2 edges from it.

Note that Theorem 1.3 holds trivially for bipartite graphs F as well, since in this case $\text{ex}_n(F) = o(n^2)$, and $T_{n,1}$ corresponds to an independent set.

Next we state two more commonly asked questions in extremal combinatorics, which we shall discuss in the context of being F -free.

- (Q3) How many discrete structures of given size have the monotone property? E.g., how large is the set $\text{Forb}_n(F)$?
- (Q4) Do the *typical* (drawn uniform at random) discrete structures with this property have some common features? E.g., are there any common features of almost all graphs in $\text{Forb}_n(F)$?

For K_{t+1} -free graphs both of these questions were addressed in the work of Erdős, Kleitman, and Rothschild [37] and Kolaitis, Prömel, and Rothschild [70, 71]. In particular, it was shown that *almost all* K_{t+1} -free graph on n vertices are t -colorable subgraphs of K_n .

Theorem 1.4 (Kolaitis, Prömel & Rothschild, 1985). *For every integer $t \geq 2$ the limit $\lim_{n \rightarrow \infty} |\text{Forb}_n(K_{t+1})|/|\text{Col}_n(t)|$ exists and*

$$\lim_{n \rightarrow \infty} \frac{|\text{Forb}_n(K_{t+1})|}{|\text{Col}_n(t)|} = 1.$$

Similarly to the extension of Turán’s theorem in [105], Theorem 1.4 was extended by Prömel and Steger [87] from cliques K_{t+1} to graphs *containing a color-critical edge*, i.e., $(t + 1)$ -chromatic graphs F with the property that $\chi(F - f) = t$ for some edge $f \in E(F)$ (see also [6, 7] for more recent extensions of Theorem 1.4).

Regarding question (Q3), for arbitrary graphs F , the size of $\text{Forb}_n(F)$ was studied by Erdős, Frankl, and Rödl [36], and those authors arrived at the following estimate (see also [5] for a more recent improvement).

Theorem 1.5 (Erdős, Frankl & Rödl, 1986). *For every $\varepsilon > 0$ and every graph F there exists n_0 such that for every $n \geq n_0$ we have*

$$|\text{Forb}_n(F)| \leq 2^{\text{ex}_n(F) + \varepsilon n^2}.$$

Note that $|\text{Forb}_n(F)| \geq 2^{\text{ex}_n(F)}$ holds trivially, since every subgraph of an extremal graph on n vertices is F -free. Therefore, Theorem 1.5 implies for every graph F that

$$\lim_{n \rightarrow \infty} \frac{\log_2 |\text{Forb}_n(F)|}{\binom{n}{2}} = \pi(F).$$

The extremal results stated above were motivated by Turán's theorem, and the problems addressed by those results allow natural generalizations for subgraphs of random graphs. In the next section we consider such extensions, where the complete graph K_n (in the definition of $\text{ex}_n(F)$ and $\text{Forb}_n(F)$) is replaced by a *random graph* with vanishing edge density. We will discuss some further extremal results, including the *removal lemma* and the *clique density theorem* in Section 4.

2. EXTREMAL PROBLEMS FOR RANDOM GRAPHS

Motivated by questions in Ramsey theory (also known as *Folkman-type problems*), in 1983, at the first *Random Structures and Algorithms* conference in Poznań, Erdős and Nešetřil (see [35]) posed the following extremal problem: Is it true that for every $\varepsilon > 0$ there exists a K_4 -free graph G such that any subgraph $H \subseteq G$ containing at least $(1/2 + \varepsilon)e(G)$ edges must contain a triangle? In other words, Erdős and Nešetřil asked whether for $F = K_3$ one may replace K_n in the Erdős–Stone theorem by a graph which contains no larger cliques than the triangle itself. This question was answered positively by Frankl and Rödl [43] by a random construction. Those authors considered the *binomial random graph* $G(n, p)$ with vertex set $[n] = \{1, \dots, n\}$, in which the edges are chosen independently, each with, probability p (see, e.g., [14, 58] for standard textbooks on the topic). More precisely, it was shown that for $p = n^{-1/2 + o(1)}$ a.a.s. one may remove $o(pn^2)$ edges from $G \in G(n, p)$ (one from every copy of K_4 in G) such that the remaining graph has the desired property. In particular, a.a.s. the largest triangle-free subgraph of $G(n, p)$ contains at most $(\pi(K_3) + o(1))p \binom{n}{2}$ edges (see Theorem 2.1 below).

It will be convenient to extend the definitions $\text{Col}_n(t)$, $\text{Forb}_n(F)$ and $\text{ex}_n(F)$ from Section 1.2 to a more general setting. For a graph G and an integer t let

$$\text{Col}_G(t) = \{H \subseteq G : \chi(H) \leq t\}$$

be the set of t -colorable subgraphs of G . Similarly, for a graph F with at least one edge we denote by $\text{Forb}_G(F)$ the set of all subgraphs of G not containing a copy of F , i.e.,

$$\text{Forb}_G(F) = \{H \subseteq G : H \text{ is } F\text{-free}\},$$

and we define the *generalized extremal function* $\text{ex}_G(F)$ as the maximum number of edges of the elements of $\text{Forb}_G(F)$, i.e.,

$$\text{ex}_G(F) = \max\{e(H) : H \in \text{Forb}_G(F)\}.$$

The following was proved by Frankl and Rödl in [43].

Theorem 2.1. *Let $\varepsilon > 0$ and $p \geq n^{-1/2+\xi}$ for some $\xi > 0$. Then a.a.s. for $G \in G(n, p)$ we have $\text{ex}_G(K_3) \leq (\pi(K_3) + \varepsilon)e(G)$.*

In view of Theorem 2.1 several questions arise (see below). The systematic study of these questions was initiated by the work of Kohayakawa and his collaborators in [56, 57, 63, 65, 78]. In particular, Kohayakawa, Łuczak, and Rödl formulated conjectures in [65], which led to the subsequent work discussed here.

- (R1) What is the smallest p such that Theorem 2.1 holds?
- (R2) For which p can Theorem 2.1 be extended to other graphs F instead of the triangle?
- (R3) Are there stability versions for those results?
- (R4) Is there a strengthening of Theorem 2.1 which, similarly as Mantel’s theorem (Theorem 1.1 for $t = 2$), establishes the equality between the maximum size of a bipartite subgraph and that of a triangle-free subgraph? More precisely, for which p a.a.s. $G \in G(n, p)$ has the following property: every $H \in \text{Forb}_G(K_3)$ with $e(H) = \text{ex}_G(K_3)$ is bipartite?
- (R5) What can be said about extensions of Theorems 1.4 and 1.5, where instead of K_{t+1} -free subgraphs of K_n , one studies K_{t+1} -free subgraphs of $G(n, p)$ for appropriate p ? Are almost all of those t -partite or “close” to being t -partite?

We will address questions (R1)–(R3) in the next section, Section 2.1. In Section 2.2 we address a generalization of the question of Erdős–Nešetřil that led to Theorem 2.1. Results addressing question (R4) will be discussed in Section 2.3 and then we turn to question (R5) in Section 2.4.

2.1. Threshold for the Erdős–Stone Theorem

A common theme in the theory of random graphs is the *threshold phenomenon*. For example, it was already observed by Erdős and Whitney (unpublished) and Erdős and Rényi [38] that within a “small range” of p (around $\ln n/n$) the random graph $G(n, p)$ quickly changes its behavior from being a.a.s. disconnected to being a.a.s. connected. In other words, $\hat{p} = \ln n/n$ is the *threshold* for $G(n, p)$ being connected. In more generality, for a *graph property* \mathcal{P} , i.e. a set of graphs closed under isomorphism, we say $0 \leq \hat{p} = \hat{p}(n) \leq 1$ is a *threshold function* for \mathcal{P} , if

$$(1) \quad \lim_{n \rightarrow \infty} \mathbb{P}(G(n, p) \in \mathcal{P}) = \begin{cases} 0, & \text{if } p \ll \hat{p}, \\ 1, & \text{if } p \gg \hat{p}. \end{cases}$$

We refer to the two statements involved in this definition as the *0-statement* and the *1-statement* of the threshold. It is well known (see, e.g., [15]) that every *monotone property* \mathcal{P} has a threshold.

In Theorem 2.1 the following property is studied for $F = K_3$. For given $\varepsilon > 0$, a graph F with at least one edge, and an integer n , consider

$$\mathcal{G}_n(F, \varepsilon) = \{G = (V, E) : V = [n] \text{ and } \text{ex}_G(F) \leq (\pi(F) + \varepsilon)e(G)\}.$$

We note that $\mathcal{G}_n(F, \varepsilon)$ is not monotone. Consider, for example, the case when $F = K_3$, and let $G \subseteq G'$ be graphs with vertex set $[n]$, where G consists of a clique on $n^{1/3}$ vertices all other vertices isolated and G' consists of the union of G and a perfect matching.

Since $\mathcal{G}_n(F, \varepsilon)$ is not monotone, the threshold is not guaranteed to exist by the aforementioned result from [15]. On the other hand, $\mathcal{G}_n(F, \varepsilon)$ is “probabilistically monotone” (see, e.g., [58, Proposition 8.6]), and from this it follows that indeed it has a threshold for all non-trivial F and $\varepsilon > 0$. In view of this, questions (R1) and (R2) ask to determine the threshold for $\mathcal{G}_n(K_3, \varepsilon)$ and, more generally, for $\mathcal{G}_n(F, \varepsilon)$ for general F .

Concerning the threshold for $\mathcal{G}_n(K_3, \varepsilon)$, it follows from Theorem 2.1 that for every $\varepsilon > 0$ this threshold is at most $\hat{p} < n^{-1/2+\varepsilon}$. However, a more careful analysis of the proof presented in [43] yields $O(n^{-1/2})$ as an upper bound (see, e.g., [58, Section 8.2]). For the lower bound on the threshold, we note that the expected number of triangles in $G(n, p)$ for $p = o(n^{-1/2})$ is $o(pn^2)$. Hence by removing from $G(n, p)$ one edge from every triangle, we expect to be left with a triangle-free subgraph of $G(n, p)$ containing $1 - o(1)$ proportion of the edges of $G(n, p)$. In fact, this argument can be made precise, and it follows that $\hat{p} = n^{-1/2}$ is a threshold for $G(n, p) \in$

$\mathcal{G}_n(K_3, \varepsilon)$ for every $\varepsilon > 0$, which answers question (R1). (We remark that, in particular, the threshold function \hat{p} is independent of ε).

Regarding question (R2), we note that the lower bound for the threshold discussed above can be extended to arbitrary graphs and leads to the definition of the 2-density $m_2(F)$ of a graph F with at least one edge given by

$$(2) \quad m_2(F) = \max\{d_2(F') : F' \subseteq F \text{ with } e(F') \geq 1\},$$

where

$$d_2(F') = \begin{cases} \frac{e(F') - 1}{|V(F')| - 2}, & \text{if } |V(F')| > 2, \\ 1/2, & \text{if } F' = K_2. \end{cases}$$

We say a graph F is 2-balanced if $d_2(F) = m_2(F)$ and it is strictly 2-balanced if $d_2(F') < d_2(F) = m_2(F)$ for all subgraphs $F' \subsetneq F$ with at least one edge.

It follows from the definition of the 2-density that $p = \Omega(n^{-1/m_2(F)})$ if, and only if the expected numbers of copies of F or any of its subgraphs in $G(n, p)$ is at least of order $\Omega(pn^2)$ – the order of the expected number of edges in $G(n, p)$. Similarly as above, one can deduce that for every $\varepsilon > 0$ and every graph F with at least one edge, $n^{-1/m_2(F)}$ is a lower bound for the threshold for $\mathcal{G}_n(F, \varepsilon)$. Moreover, Kohayakawa, Łuczak, and Rödl [65, Conjecture 1(i)] conjectured that this heuristic gives the right bound, and that a matching upper bound for the threshold can be proved. Until recently this conjecture was only proved for cliques of size at most six [65, 52, 48] and for cycles [56, 57]. In 2009 the conjecture was confirmed independently by Conlon and Gowers [22] for strictly 2-balanced graphs F and by Schacht [102] for all graphs F . This work yields the following probabilistic version of the Erdős–Stone theorem for the random graph $G(n, p)$.

Theorem 2.2. *For every graph F with $\delta(F) \geq 2$ and every $\varepsilon > 0$ the function $\hat{p} = n^{-1/m_2(F)}$ is a threshold for $\mathcal{G}_n(F, \varepsilon)$.*

Next we discuss research addressing question (R3). Recall that every graph G contains a t -partite subgraph with at least $(1 - 1/t)e(G)$ edges, which is clearly F -free for every F with chromatic number $t + 1$. On the other hand, the 1-statement (see (1)) of Theorem 2.2 implies that a.a.s. the F -free subgraph of $G \in G(n, p)$ with the maximum number of edges has at most $(1 - 1/t + o(1))e(G)$ edges. The question that arises is whether those two subgraphs of $G(n, p)$, the maximum t -partite subgraph and the maximum F -free subgraph, have similar structure. It was conjectured by

Kohayakawa, Luczak, and Rödl [65, Conjecture 1(ii)] that such a statement is true as long as p is of the order of magnitude given in the 1-statement of the threshold in Theorem 2.2. Conlon and Gowers [22] verified this conjecture for strictly 2-balanced graphs F , and Samotij [100] adapted and simplified the approach of Schacht [102] to obtain such a result for all graphs F . This led to the following probabilistic version of the Erdős–Simonovits stability theorem.

Theorem 2.3. *For every $\varepsilon > 0$ and every graph F with $\chi(F) = t + 1 \geq 3$ there exist constants C and $\delta > 0$ such that for $p > Cn^{-1/m_2(F)}$ the following holds a.a.s. for $G \in G(n, p)$. If H is an F -free subgraph of G satisfying*

$$e(H) \geq \text{ex}_G(F) - \delta pn^2,$$

then H can be made t -partite by removing at most εpn^2 edges from it.

We recall that Theorems 2.2 and 2.3 were conjectured (together with Conjecture 3.6 stated in Section 3.2) in [65]. These conjectures played a central rôle in the area. In particular, partial results towards these conjectures were made by the authors of the conjecture and their collaborators [49, 66, 67, 68], by Gerke and Steger and their collaborators [48, 50, 51, 52, 54] (see also the survey [53]), and by Szabó and Vu [109].

2.2. General Erdős–Nešetřil Problem

Before we continue with the discussion of extremal results for sparse random graphs, we generalize the problem of Erdős and Nešetřil. Based on Theorem 2.2, one now can prove the following generalization of the Erdős–Nešetřil problem, which extends the results of [43] from forbidding triangles to forbidding cliques of arbitrary fixed size.

Corollary 2.4. *For every integer $k \geq 3$ and $\varepsilon \in (0, 1 - \pi(K_k))$ the following holds:*

(i) *there exists a K_{k+1} -free graph G such that*

$$\text{ex}_G(K_k) \leq (\pi(K_k) + \varepsilon) e(G);$$

(ii) *for every fixed $d > 0$ there exists an n_0 such that there is no graph G on $n \geq n_0$ vertices with $e(G) = d \binom{n}{2}$ having the properties from part (i).*

While the first statement of Corollary 2.4 asserts the existence of a K_{k+1} -free graph with the property that every $(\pi(K_k) + \varepsilon)$ proportion of its edges

contains a K_k , the second statement asserts that such a graph must have vanishing density.

In the proof of part (i) we consider $G(n, p)$ for $p = Cn^{-1/m_2(K_k)}$. Owing to Theorem 2.2 we know that a.a.s. $G \in G(n, p)$ satisfies $\text{ex}_G(K_k) \leq (\pi(K_k) + o(1))e(G)$. On the other hand, since $n^{-1/m_2(K_k)} \ll n^{-1/m_2(K_{k+1})}$ for this choice of p , the number of copies of K_{k+1} in G will be of order $o(pn^2)$. Consequently, we may remove $o(pn^2)$ edges from G and the resulting graph is K_{k+1} -free and satisfies the properties of part (i) of Corollary 2.4. In fact, one may check that the same proof works for all values of p with $Cn^{-1/m_2(K_k)} \leq p \leq cn^{-1/m_2(K_{k+1})}$ for appropriate constants C and $c > 0$. We give the details of this proof after the following remark.

Remark 2.5. One can show that statement (ii) is best possible. Indeed, given $d = d(n) = o(1)$, let $(G_m)_{m \in \mathbb{N}}$ be a sequence of m vertex graphs with the properties of part (i) and with density $\varrho = \varrho(m) \ll d(m)$. Since $d = o(1)$, we can find infinitely many values for which $d(n) \sim \varrho(m)$. For such an m we “blow-up” G_m by replacing each vertex by an independent set of size n/m and every edge by a complete bipartite graph with vertex classes of size n/m . The resulting graph G has n vertices, density approximately $d(n)$, and it “inherits” the properties of G_m with respect to statement (i).

Finally, we remark that in Section 4.5 we will generalize part (ii) and show that no relatively dense subgraph of $G(n, p)$ for $p \gg n^{-1/m_2(K_{k+1})}$ satisfies the properties of part (ii) (see Theorem 4.10).

Proof of Corollary 2.4 part (i). Part (i) follows directly from Theorem 2.2 combined with an alteration argument (similar to the one carried out by Erdős in [31], see also [3, Section 3]).

Let $\varepsilon > 0$ and $k \geq 3$ be given. Applying Theorem 2.2 for $\varepsilon/2$ and $F = K_k$ implies that there exists a constant $C > 0$ such that for $p = p(n) = Cn^{-1/m_2(K_k)}$ a.a.s. for $G \in G(n, p)$ we have

$$(3) \quad \text{ex}_G(K_k) \leq \left(\pi(K_k) + \frac{\varepsilon}{2} \right) e(G).$$

Since $k \geq 3$, we have

$$m_2(K_k) = \frac{\binom{k}{2} - 1}{k - 2} = \frac{k + 1}{2} \geq 2,$$

and thus also

$$p = Cn^{-\frac{2}{k+1}} \geq \frac{C}{\sqrt{n}}.$$

It follows that $pn^2 \geq Cn^{3/2}$. Chebyshev's inequality easily yields that a.a.s.

$$(4) \quad e(G) \geq \frac{1}{2}p \binom{n}{2}.$$

Finally, we note that the expected number of copies of K_{k+1} in G is at most

$$p \binom{k+1}{2} n^{k+1} = C \binom{k+1}{2} n \leq \frac{\varepsilon}{4} p \binom{n}{2},$$

for sufficiently large n . Hence it follows from Markov's inequality that, with probability at least $1/2$, the graph G contains at most $(\varepsilon/2)p \binom{n}{2}$ copies of K_{k+1} . Consequently, for sufficiently large n there exists a graph G containing at most $(\varepsilon/2)p \binom{n}{2}$ copies of K_{k+1} and for which (3) and (4) also hold. Let G' be the graph obtained from G by removing one edge from every copy of K_{k+1} in G . Obviously, the graph G' is K_{k+1} -free,

$$e(G') \geq e(G) - \frac{\varepsilon}{4} p \binom{n}{2},$$

and owing to

$$\begin{aligned} (\pi(K_k) + \varepsilon)e(G') &> (\pi(K_k) + \varepsilon)e(G) - \frac{\varepsilon}{4} p \binom{n}{2} \\ &= \left(\pi(K_k) + \frac{\varepsilon}{2}\right) e(G) + \frac{\varepsilon}{2} e(G) - \frac{\varepsilon}{4} p \binom{n}{2} \\ &\stackrel{(4)}{\geq} \left(\pi(K_k) + \frac{\varepsilon}{2}\right) e(G), \end{aligned}$$

it follows from (3) that $\text{ex}_{G'}(K_k) \leq (\pi(K_k) + \varepsilon)e(G')$, which concludes the proof of assertion (i) in Corollary 2.4. ■

Next we prove assertion (ii). The proof follows the main ideas of [43, Theorem 4].

Definition 2.6. For a graph $G = (V, E)$ we call a partition $V_1 \cup \dots \cup V_t = V$ a *t-cut*. We denote by $E_G(V_1, \dots, V_t)$ the *edges of the t-cut*, i.e., those edges of G with its vertices in two different sets of the partition and we denote by $e_G(V_1, \dots, V_t) = |E_G(V_1, \dots, V_t)|$ the *size of the t-cut*. Moreover, we say a *t-cut* is *balanced*, if $|V_1| \leq \dots \leq |V_t| \leq |V_1| + 1$.

A simple averaging argument shows that there always exists a balanced t -cut of G of size at least $(1 - 1/t)e(G)$. The following lemma, which implies assertion (ii) of Corollary 2.4, shows that if on the other hand all balanced t -cuts have size at most $(1 - 1/t + o(1))e(G)$, then G contains cliques of arbitrary size.

Lemma 2.7. *For all integers $s, t \geq 2$ and every $d > 0$ there exist $\varepsilon > 0$ and n_0 such that the following holds. Let $G = (V, E)$ be a graph on $|V| = n \geq n_0$ vertices, with $|E| = d\binom{n}{2}$ edges, and with the property that every balanced t -cut has size at most $(1 - 1/t + \varepsilon)d\binom{n}{2}$. Then G contains a copy of K_s .*

Before we prove Lemma 2.7, we deduce assertion (ii) of Corollary 2.4 from it.

Proof of Corollary 2.4 part (ii). Suppose that part (ii) of Corollary 2.4 fails to be true. We assume that there is a K_{k+1} -free graph G on n vertices with $\text{ex}_n(K_k) \leq (\pi(K_k) + \varepsilon)e(G)$ and with $d\binom{n}{2}$ edges for some constant $d > 0$. We apply Lemma 2.7 with $s = k + 1$ and $t = k - 1$. Since the edges of every $(k - 1)$ -cut span no copy of K_k the assumption of Corollary 2.4 part (ii) guarantees that the size of every $(k - 1)$ -cut in G is bounded from above by

$$(\pi(K_k) + \varepsilon)e(G) = \left(1 - \frac{1}{k - 1} + \varepsilon\right) d\binom{n}{2},$$

and it follows from Lemma 2.7 that G contains a K_{k+1} , which contradicts the assumption on G . ■

The proof of Lemma 2.7 draws on some ideas from the theory of quasi-random graphs [20]. In particular, it is based on the following well known fact (see, e.g., [94, Theorem 2]).

Lemma 2.8. *For all integers $s, t \geq 2$ and every $d > 0$ there exist $\delta > 0$ and n_0 such that the following holds. Let $G = (V, E)$ be a graph on $|V| = n \geq n_0$ vertices such that $e_G(U) = (d \pm \delta)\binom{|n|}{2}$ for every $U \subseteq V$ with $|U| = \lfloor n/t \rfloor$. Then G contains a copy of K_s .*

Proof of Lemma 2.7. Let integers s and $t \geq 2$ be fixed. Suppose for a contradiction that the lemma fails to be true with this choice of s and t . This means that there is a density $d > 0$ for which the statement fails, so we fix “the largest such d ”. More precisely, let $d > 0$ be chosen in such a way that the statement fails for s, t , and d , but it holds for s, t and any $d' > d$ provided $\varepsilon' > 0$ is sufficiently small and n is sufficiently large. We

remark that such a choice is possible, since for fixed s and t the validity of the statement for d implies it for every $d' \geq d$.

Our choice of δ' will be given by Lemma 2.8. First let $\delta > 0$ be the constant guaranteed by Lemma 2.8 for the already fixed s, t , and d and set

$$(5) \quad \delta' = \frac{\delta}{2(t-1)} \quad \text{and} \quad \varepsilon = \min \left\{ \frac{\delta}{4t^2}, \frac{\varepsilon'(d + \delta')}{4t^2} \right\},$$

where $\varepsilon' > 0$ is given by Lemma 2.7 applied with $d' \geq d + \delta'$ (which holds by our assumption). Finally, let n_0 be sufficiently large (for example, so that we can appeal to Lemma 2.8 with s, t, d , and δ and to the validity of Lemma 2.7 for $d' \geq d + \delta'$ and $\varepsilon' > 0$). Let $G = (V, E)$ with $|V| = n \geq n_0$ be a counterexample for those choices. Without loss of generality we assume that t^2 divides n .

Since G contains no copy of K_s , Lemma 2.8 implies that there exists a set V_1 of size n/t such that either $e_G(V_1) < (d - \delta) \binom{n/t}{2}$ or $e_G(V_1) > (d + \delta) \binom{n/t}{2}$. Fix some balanced t -cut $V_1 \cup \dots \cup V_t = V$ which contains V_1 . We will infer that G induces a denser graph on one of the sets of the partition. This is obvious if $e_G(V_1) > (d + \delta) \binom{n/t}{2}$. However, if $e_G(V_1) < (d - \delta) \binom{n/t}{2}$, then we will show that there also is a partition class that induces a denser graph. In fact, using the assumption on G for the sizes of the balanced t -cuts, an averaging argument shows that there exists some $i = 2, \dots, t$ such that

$$\begin{aligned} e_G(V_i) &\geq \frac{e(G) - e_G(V_1, \dots, V_t) - e_G(V_1)}{t-1} \\ &= \frac{d \binom{n}{2} - (1 - \frac{1}{t} + \varepsilon) d \binom{n}{2} - (d - \delta) \binom{n/t}{2}}{t-1} \\ &\geq \frac{(1/t - \varepsilon) d \binom{n}{2} - d \binom{n/t}{2} + \delta \binom{n/t}{2}}{t-1} \\ &\geq \frac{(t-1 - 2\varepsilon t^2) d \binom{n/t}{2} + \delta \binom{n/t}{2}}{t-1} \\ &\stackrel{(5)}{\geq} (d + \delta') \binom{n/t}{2}. \end{aligned}$$

Summarizing, we can fix some $i \in [t]$ such that for $W = V_i$ we have $e_G(W) = d' \binom{n/t}{2}$ for some $d' \geq d + \delta'$.

Since G (and hence also the induced subgraph $G[W]$) contains no copy of K_s , by our assumptions $G[W]$ fails to satisfy the assumptions of Lemma 2.7. Consequently, there exists a balanced t -cut $W_1 \cup \dots \cup W_t = W$ with

$$(6) \quad e_{G[W]}(W_1, \dots, W_t) > \left(1 - \frac{1}{t} + \varepsilon'\right) d' \binom{n/t}{2} = \left(1 - \frac{1}{t} + \varepsilon'\right) e_G(W).$$

We will extend this balanced t -cut of $G[W]$ to a balanced t -cut of G with size bigger than

$$(7) \quad \left(1 - \frac{1}{t} + \varepsilon\right) d \binom{n}{2},$$

which will then contradict the assumptions on G .

We consider a random balanced t -cut $U_1 \cup \dots \cup U_t$ of $U = V \setminus W$. A standard application of Chernoff's inequality for the hypergeometric distribution (see, e.g., [58, Theorem 2.10]) shows that with probability close to one, we have

$$(8) \quad e_G(W_i, U_j) = \left(\frac{1}{t} \pm o(1)\right) e_G(W_i, U) \quad \text{for all } i, j \in [t]$$

and

$$(9) \quad e_G(U_1, \dots, U_t) = \left(1 - \frac{1}{t} \pm o(1)\right) e_G(U).$$

Let such a t -cut be fixed. Since both t -cuts were balanced, the t -cut $V'_1 \cup \dots \cup V'_t$ of V given by $V'_i = W_i \cup U_i$ is also balanced. We estimate the size of this cut as follows:

(10)

$$e_G(V'_1, \dots, V'_t) = e_G(W_1, \dots, W_t) + \sum_{i=1}^t \sum_{j \neq i} e_G(W_i, U_j) + e_G(U_1, \dots, U_t).$$

By (8), we have

$$\begin{aligned} \sum_{i=1}^t \sum_{j \neq i} e_G(W_i, U_j) &= \sum_{i=1}^t \sum_{j \neq i} \left(\frac{1}{t} \pm o(1)\right) e_G(W_i, U) \\ &= \left(\frac{t-1}{t} \pm o(1)\right) \sum_{i=1}^t e_G(W_i, U) \\ &= \left(1 - \frac{1}{t} \pm o(1)\right) e_G(W, U) \end{aligned}$$

and combined with (6) and (9), from (10) we get

$$e_G(V'_1, \dots, V'_t) \geq \left(1 - \frac{1}{t} - o(1)\right) e(G) + \varepsilon' e_G(W).$$

Hence, we obtain (7) from

$$\varepsilon' e_G(W) = \varepsilon' d' \binom{n/t}{2} \geq \frac{\varepsilon' d'}{2t^2} \binom{n}{2} \stackrel{(5)}{\geq} 2\varepsilon \binom{n}{2}. \quad \blacksquare$$

2.3. Turán’s Theorem for Random Graphs

Turán’s theorem not only determines the extremal function $\text{ex}_n(K_{t+1})$ precisely, but also asserts that the complete balanced t -partite graph on n vertices is the unique extremal graph. The extremal results for $G(n, p)$ discussed in Section 2.1 do not fully address this question (see also (R_4)). For example, Theorem 2.2 applied for $F = K_{t+1}$ gives no information about the structure of extremal K_{t+1} -free subgraphs of $G(n, p)$. In this section, we discuss results motivated by this question.

For an integer $t \geq 2$ and a graph G let $\text{col}_G(t)$ be the maximum number of edges of a t -colorable subgraph of G , i.e., the size of the maximum t -cut in G . For simplicity we write $\text{col}_n(t)$ for $\text{col}_{K_n}(t)$. Turán’s theorem establishes

$$\text{ex}_n(K_{t+1}) = \text{col}_n(t).$$

Babai, Simonovits, and Spencer [4] were the first to investigate the extent to which such an identity can be extended to random graphs. In particular, those authors showed that it holds for $G(n, 1/2)$ in the case of triangles ($t = 2$), by showing that a.a.s. $G \in G(n, 1/2)$ satisfies

$$(11) \quad \text{ex}_G(K_3) = \text{col}_G(2).$$

Answering a question from [4], it was shown by Brightwell, Panagiotou, and Steger [17] that $p = 1/2$ can be replaced by $p = n^{-\eta}$ for some $\eta > 0$. Moreover, their proof extends to cliques of arbitrary fixed size and establishes that the identity $\text{ex}_G(K_{t+1}) = \text{col}_G(t)$ holds a.a.s. for $G \in G(n, p)$ as long as $p > n^{-\eta_t}$ for some sufficiently small $\eta_t > 0$. Those authors conjectured that this result can be extended to smaller values of p . Note that (11) holds trivially for very small p , when a.a.s. the random graph itself is bipartite. However, here and below we shall exclude this range of p . It was noted in [17] that (with the exception of small p) in order for (11) to hold,

$p > c(\log n/n)^{1/2}$ is a necessary condition for some sufficiently small $c > 0$. The reason for this is that for $p < c(\log n/n)^{1/2}$, cycles of length five appear in $G(n, p)$ which have the additional property that none of its edges is contained in a triangle. Recently, DeMarco and Kahn [26] obtained a matching upper bound by proving the following probabilistic version of Mantel’s theorem (Theorem 1.1 for $t = 2$).

Theorem 2.9. *There exists a constant $C > 0$ such that for $p > C(\log n/n)^{1/2}$ a.a.s. $G \in G(n, p)$ satisfies $\text{ex}_G(K_3) = \text{col}_G(2)$. Moreover, every triangle-free subgraph of G with the maximum number of edges is bipartite.*

It would be interesting to generalize this results to larger cliques. It seems plausible that a necessary condition on p for such a generalization should come from the requirement that *all* edges of $G(n, p)$ are contained in a cliques of size $t + 1$. In particular, the edges not contained in a copy of K_{t+1} should not form a high chromatic subgraph. For this we require on average $\Omega(\log n)$ such cliques per edge, instead of a constant number of cliques per edge, which gave rise to the 2-density. For K_{t+1} we get

$$p^{\binom{t+1}{2}} n^{t+1} = \Theta(pn^2 \log n).$$

Solving this for p leads to the following conjecture, which was stated by DeMarco and Kahn [26].

Conjecture 2.10. *For every integer $t \geq 2$ there exists a $C > 0$ such that for $p \geq C((\log n)^{\frac{1}{t-1}}/n)^{\frac{2}{t+2}}$ a.a.s. $G \in G(n, p)$ satisfies $\text{ex}_G(K_{t+1}) = \text{col}_G(t)$.*

It would be also of interest to prove similar results for graphs F containing a color-critical edge. Partial results in this direction can be found in [4] (see also [17]).

2.4. Triangle-free Graphs with Given Number of Vertices and Edges

In this section we discuss extensions of Theorems 1.4 and 1.5. Most of the work studied $\text{Forb}_{n,M}(K_3)$, the set of triangle-free graphs with n vertices and M edges. The first result in this direction is due to Prömel and Steger [88], who proved a strengthening of the Erdős–Kleitman–Rothschild theorem (Theorem 1.4 for $t = 2$). It was shown that for $M > Cn^{7/4} \log n$, almost every graph $H \in \text{Forb}_{n,M}(K_3)$ is bipartite. Similarly to the case of Turán’s theorem for random graphs discussed in the last section, such an assertion holds also for very small values of M , but not in the medium

range (see Theorem 2.11 below). It was also noted in [88] that the statement fails to be true if $M = cn^{3/2}$ for some $c > 0$. The gap between $cn^{3/2}$ and $Cn^{7/4} \log n$ was closed by Osthus, Prömel, and Taraz [86] (see also Steger [108] for a bit weaker result). In particular, the following result was shown in [86]. For positive integers n , M , and t , we denote by $\text{Col}_{n,M}(t)$ the set of t -colorable graphs with n vertices and M edges.

Theorem 2.11. *For every $\varepsilon > 0$ the following holds*

$$\lim_{n \rightarrow \infty} \frac{|\text{Forb}_{n,M}(K_3)|}{|\text{Col}_{n,M}(2)|} = \begin{cases} 1, & \text{if } M = M(n) = o(n), \\ 0, & \text{if } n/2 \leq M = M(n) \leq (1 - \varepsilon) \frac{\sqrt{3}}{4} n^{3/2} \sqrt{\ln n}, \\ 1, & \text{if } M = M(n) \geq (1 + \varepsilon) \frac{\sqrt{3}}{4} n^{3/2} \sqrt{\ln n}. \end{cases}$$

Note that similarly to Theorem 2.9 the “critical window” in Theorem 2.11 concerns graphs with $\Theta(n^{3/2} \sqrt{\log n})$ edges. That might not be a coincidence, since having the property that every pair is covered by a path of length two seems to be a necessary condition for both problems. Generalizing this to the property that adding an edge for any pair of vertices would close a copy of K_{t+1} suggests a joint generalization of the Kolaitis–Prömel–Rothschild theorem, Theorem 1.4, and of Theorem 2.11, which was recently obtained by Balogh, Morris, Samotij, and Warnke [9].

A closely related result was proved by Łuczak. In [78] he studied slightly sparser triangle-free graphs and showed that for $M = M(n) \gg n^{3/2}$, almost every graph $H \in \text{Forb}_{n,M}(K_3)$ is “close” to a bipartite graphs, i.e., it can be made bipartite by removing at most $o(M)$ edges. In fact, he also proved that this result generalizes for larger cliques, provided Conjecture 3.6 (stated below), which we discuss in the next section, holds. Recently Balogh, Morris, and Samotij [8] and Saxton and Thomason [101] developed an approach which, among other results, allowed them to prove Conjecture 3.6 and it could be used to verify Łuczak’s statement directly.

Theorem 2.12. *For every $\delta > 0$ and $t \geq 2$ there exists a $C > 0$ and n_0 such that for $M = M(n) \geq Cn^{2-1/m_2(K_{t+1})}$ almost every graph H drawn uniformly at random from $\text{Forb}_{n,M}(K_{t+1})$ can be made t -colorable by removing at most δM edges.*

It is known that, up to the constant C , this result is best possible, and we also remark that the smallest $M = M(n)$ in Theorem 2.12 coincides in order of magnitude with the expected number of edges in $G(n, p)$ around the thresholds from Theorem 2.2.

3. REGULARITY METHOD

One of the most important tools in extremal graph theory is Szemerédi’s regularity lemma [111], and for a thorough discussion of its history and many of its applications we refer to [72, 73]. In fact, there were some applications of this lemma addressing extremal and Ramsey-type questions of random graphs (see, e.g., [4, 97]). However, for the systematic study of extremal problems of $G(n, p)$ for $p = o(1)$, a variant of the lemma discovered independently by Kohayakawa [61] and Rödl (unpublished) seemed to be an appropriate tool. We begin the discussion with Szemerédi’s regularity lemma.

3.1. Szemerédi’s Regularity Lemma

We first introduce the necessary definitions. Let $H = (V, E)$ be a graph, and let $X, Y \subseteq V$ be a pair of non-empty and disjoint subsets of the vertices. We denote by $e_H(X, Y)$ the number of edges in the bipartite subgraph induced by X and Y , i.e.,

$$e_H(X, Y) = |\{ \{x, y\} \in E : x \in X \text{ and } y \in Y \} |.$$

We also define the *density of the pair* (X, Y) by setting

$$d_H(X, Y) = \frac{e_H(X, Y)}{|X||Y|}.$$

Moreover, we say a the pair (X, Y) is ε -regular for some $\varepsilon > 0$, if

$$|d_H(X, Y) - d_H(X', Y')| < \varepsilon$$

for all subsets $X' \subseteq X$ and $Y' \subseteq Y$ with $|X'| \geq \varepsilon|X|$ and $|Y'| \geq \varepsilon|Y|$. With this notation we can formulate Szemerédi’s regularity lemma from [111].

Theorem 3.1 (Regularity lemma). *For every $\varepsilon > 0$ and $t_0 \in \mathbb{N}$ there exist integers T_0 and n_0 such that every graph $H = (V, E)$ with $|V| = n \geq n_0$ vertices admits a partition $V = V_1 \cup \dots \cup V_t$ satisfying*

- (i) $t_0 \leq t \leq T_0$,
- (ii) $|V_1| \leq \dots \leq |V_t| \leq |V_1| + 1$, and
- (iii) all but at most εt^2 pairs (V_i, V_j) with $i \neq j$ are ε -regular.

Note that most applications of Theorem 3.1 involve dense graphs (i.e., n -vertex graphs with $\Omega(n^2)$ edges). For each graph the lemma allows us to decompose the graph into bipartite “blocks,” the majority of which have a

uniform edge distribution. If such a graph has only $o(n^2)$ edges, it may not provide such control, since all edges may be contained in exceptional pairs (see property (iii) in Theorem 3.1). Moreover, even for ε -regular pairs, we do not gain any information if the density of that pair is $o(1)$.

The following well known fact is used in many applications of the regularity lemma (see, e.g., [73, 98]). For future reference, we state both the *embedding lemma* and the *counting lemma*, even though the latter clearly implies the former.

Fact 3.2 (Embedding and counting lemma for dense graphs). *For every graph F with $V(F) = [\ell]$ and every $d > 0$, there exist $\varepsilon > 0$ and m_0 such that the following holds.*

Let $H = (V_1 \cup \dots \cup V_\ell, E_H)$ be an ℓ -partite graph with $|V_1| = \dots = |V_\ell| = m \geq m_0$ and with the property that for every edge $\{i, j\} \in E(F)$ the pair (V_i, V_j) is ε -regular in H with density $d_H(V_i, V_j) \geq d$.

Embedding Lemma: *Then H contains a partite copy of F , i.e., there exists a graph homomorphism $\varphi : F \rightarrow H$ with $\varphi(i) \in V_i$.*

Counting Lemma: *The number of partite copies satisfies*

$$(12) \quad \left| \{ \varphi : F \rightarrow H : \varphi \text{ is a graph homomorphism with } \varphi(i) \in V_i \} \right| \\ = (1 \pm f(\varepsilon)) \prod_{\{i,j\} \in E(F)} d(V_i, V_j) \prod_{i=1}^{\ell} |V_i|,$$

where $f(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

As mentioned, the counting lemma implies the embedding lemma from Fact 3.2. However, for quite a few applications the existence of one copy is sufficient.

3.2. Sparse Regularity Lemma for Subgraphs of Random Graphs

In this section we state a modified version of Szemerédi’s regularity lemma, which allows applications to sparse graphs. Though more general lemmas are known, we restrict ourselves to a version which applies a.a.s. to all subgraphs of a random graph $G \in G(n, p)$. For that we first strengthen the notion of an ε -regular pair.

Definition 3.3 ((ε, p) -regular pair). Let $\varepsilon > 0$, let $p \in (0, 1]$, let $H = (V, E)$ be a graph, and let $X, Y \subseteq V$ be non-empty and disjoint. We say the pair (X, Y) is (ε, p) -regular if

$$|d_H(X, Y) - d_H(X', Y')| < \varepsilon p$$

for all subsets $X' \subseteq X$ and $Y' \subseteq Y$ with $|X'| \geq \varepsilon|X|$ and $|Y'| \geq \varepsilon|Y|$.

Note that ε -regularity coincides with the case $p = 1$ in the definition above. However, for $p = p(n) = o(1)$ and graphs of density $\Omega(p)$ the notion of (ε, p) -regularity gives additional control and addresses the second concern discussed after Theorem 3.1. The sparse regularity lemma for subgraphs of $G(n, p)$ stated below asserts that, for those graphs ε -regularity in Theorem 3.1 can be replaced by (ε, p) -regularity. In fact, besides the restriction to subgraphs of $G(n, p)$, this is the only difference between the following version of the sparse regularity lemma from [61] and Theorem 3.1.

Theorem 3.4 (Sparse regularity lemma for subgraphs of $G(n, p)$). *For every $\varepsilon > 0$, $t_0 \in \mathbb{N}$, and every function $p = p(n) \gg 1/n$ there exist integers T_0 such that a.a.s. $G \in G(n, p)$ has the following property. Every subgraph graph $H = (V, E)$ of G with $|V| = n$ vertices admits a partition $V = V_1 \cup \dots \cup V_t$ satisfying*

- (i) $t_0 \leq t \leq T_0$,
- (ii) $|V_1| \leq \dots \leq |V_t| \leq |V_1| + 1$, and
- (iii) all but at most εt^2 pairs (V_i, V_j) with $i \neq j$ are (ε, p) -regular.

In order to make Theorem 3.4 applicable in a similar way to Szemerédi’s regularity lemma, one needs extensions of Fact 3.2. Theorem 3.4 can be proved like the original regularity lemma with fairly straightforward adjustments. To prove a corresponding form of Fact 3.2 turns out to be a challenging problem, which was resolved only recently in [8, 23, 101]. In particular, in the work of Balogh, Morris, and Samotij [8] and of Saxton and Thomason [101], a conjecture of Kohayakawa, Łuczak, and Rödl [65] was addressed. This conjecture implies a version of the embedding lemma of Fact 3.2 appropriate for applications of Theorem 3.4. In [23] only such a version was derived (see Theorem 3.8 below). For the formulation of the conjecture from [65], we require some more notation.

Definition 3.5. Let $\varepsilon > 0$, $p \in (0, 1]$, $d > 0$ and let ℓ, m, M be integers. Let F be a graph with vertex set $V(F) = [\ell]$. We denote by $\mathcal{G}(F, m, M, \varepsilon, p, d)$ the set of all ℓ -partite graphs $H = (V_1 \cup \dots \cup V_\ell, E_H)$ with

- (i) $|V_1| = \dots = |V_\ell| = m$,
- (ii) $e_H(V_i, V_j) = M \geq dpm^2$ for all $\{i, j\} \in E(F)$, and
- (iii) (V_i, V_j) is (ε, p) -regular for all $\{i, j\} \in E(F)$.

We denote by $\mathcal{B}(F, m, M, \varepsilon, p, d)$ the set of all those graphs from $\mathcal{G}(F, m, M, \varepsilon, p, d)$, which contain no (partite) copy of F , i.e.,

$$\mathcal{B}(F, m, M, \varepsilon, p, d) = \{H \in \mathcal{G}(F, m, M, \varepsilon, p, d) : \text{there is no graph homomorphism } \varphi : F \rightarrow H \text{ with } \varphi(i) \in V_i\}.$$

The first part of Fact 3.2 asserts that for $p = 1$, sufficiently small $\varepsilon = \varepsilon(F, d) > 0$ and sufficiently large $m = m(F, d)$, the set $\mathcal{B}(F, m, M, \varepsilon, p, d)$ is empty. However, if $p = o(1)$ then $\mathcal{B}(F, m, M, \varepsilon, p, d)$ is not empty for graphs F containing a cycle. In other words, if $p = o(1)$, then the regularity condition does not ensure the occurrences of copies of F . This prohibits a straightforward extension of Fact 3.2 for the sparse regularity lemma. For example, as noted earlier for $p \ll n^{-1/m_2(F)}$ a.a.s. the random graph $G(n, p)$ contains only $o(pn^2)$ copies of some subgraph $F' \subseteq F$. Therefore, a.a.s. $G(n, p)$ contains an F' -free subgraph with $(p - o(1))\binom{n}{2}$ edges. This can be used to construct many F' -free graphs $H \in \mathcal{G}(F, m, M, \varepsilon, p, d)$ for any $p = o(1)$ and appropriate choices of m, M , and d . (For details see the discussion below Conjecture 3.6.) On the other hand, for $p \geq Cm^{-1/m_2(F)}$ for sufficiently large $C > 0$, it was conjectured by Kohayakawa, Łuczak, and Rödl in [65] that $\mathcal{B}(F, m, M, \varepsilon, d)$ contains only “very few” graphs.

Conjecture 3.6 (Kohayakawa, Łuczak & Rödl 1997). *For every $\alpha > 0$, $d > 0$, and every graph F with vertex set $V(F) = [\ell]$, there are $\varepsilon > 0$, $C > 0$ and m_0 such that for every $m \geq m_0$, $p \geq Cm^{-1/m_2(F)}$ and $M \geq dpm^2$ we have*

$$(13) \quad |\mathcal{B}(F, m, M, \varepsilon, p, d)| \leq \alpha^M |\mathcal{G}(F, m, M, \varepsilon, p, d)|.$$

Next we show that the lower bound on p in Conjecture 3.6 is necessary. For this, let $p = \delta m^{-1/m_2(F)}$ for some δ tending to 0 with m . We consider the family of graphs $\tilde{\mathcal{G}}(F, m, p, d)$ satisfying only properties (i) and (ii) of Definition 3.5, with $M = dpm^2$. It is not hard to show that for every $\varepsilon > 0$ almost every $H \in \tilde{\mathcal{G}}(F, m, M, p, d)$ is also contained in $\mathcal{G}(F, m, M, \varepsilon, p, d)$, i.e.,

$$(14) \quad |\mathcal{G}(F, m, M, \varepsilon, p, d)| \geq (1 - o(1)) |\tilde{\mathcal{G}}(F, m, p, d)|.$$

Moreover, let $F' \subseteq F$ be the subgraph with $d_2(F') = m_2(F')$ (see (2)), and let e and v denote its number of edges and vertices, respectively. The expected number of partite copies of F' in a graph H chosen uniformly at random from $\tilde{\mathcal{G}}(F, m, p, d)$ is

$$O((dp)^e m^v) = O((\delta d)^e pm^2) = o(pm^2).$$

Hence, all but $o(|\tilde{\mathcal{G}}(F, m, p, d)|)$ graphs $H \in \tilde{\mathcal{G}}(F, m, p, d)$ have the property that, only $o(pm^2)$ edges of H are contained in a copy of F' , and consequently also in a copy of F . Delete from each such $H \in \mathcal{G}(F, m, M, \varepsilon, p, d)$ the edges contained in copies of F and possibly a few more from each pair (V_i, V_j) , so

that the resulting graph has precisely $M' = d'pm^2 = (1 - o(1))M$ edges for each such pair. This way we obtain a graph $H' \in \mathcal{G}(F, m, M', \varepsilon', p, d')$ with $\varepsilon' = \varepsilon + o(1)$, which is F -free, i.e., H' is contained in $\mathcal{B}(F, m, M', \varepsilon', p, d')$. Consequently, we have

$$\begin{aligned} |\mathcal{B}(F, m, M', \varepsilon', p, d')| &\geq (1 - o(1)) \frac{|\tilde{\mathcal{G}}(F, m, p, d)|}{\binom{M}{o(M)}^{\varepsilon(F)}} \\ &= (1 - o(1)) \left(\frac{\binom{m^2}{M}}{\binom{M}{o(M)}} \right)^{\varepsilon(F)} \\ &\geq (1 - o(1))^{M'} \binom{m^2}{M'}^{\varepsilon(F)} \\ &= (1 - o(1))^{M'} |\tilde{\mathcal{G}}(F, m, p, d')| \\ &\geq (1 - o(1))^{M'} |\mathcal{G}(F, m, M', \varepsilon', p, d')|, \end{aligned}$$

which shows that (13) fails for $p = \delta n^{-1/m_2(F)}$ for sufficiently small $\delta > 0$.

3.3. Sparse Embedding and Counting Lemma

Conjecture 3.6 is obvious, if F is a matching. For all other graphs F , we have $m_2(F) \geq 1$, and the conjecture holds trivially for forests. More interestingly, the conjecture was shown for cliques on at most six vertices [51, 52, 64] and (with an additional technical assumption) for cycles [62] (see also [46] for an earlier related results for $F = C_4$).

Recently Conjecture 3.6 was verified by Balogh, Morris, and Samotij [8] for 2-balanced graphs F and by Saxton and Thomason [101] for all graphs F .

Theorem 3.7. *Conjecture 3.6 holds for all graphs F .*

One of the main motivations for the conjectured bound on the cardinality of $\mathcal{B}(F, m, M, \varepsilon, p, d)$ in (13) was that it easily implies that such “bad” graphs do not appear as subgraph of the random graph $G(n, p)$. In particular, we obtain an appropriate generalization of the embedding lemma from Fact 3.2, for subgraphs of $G(n, p)$ (see Theorem 3.8). This result was also shown by Conlon, Gowers, Samotij, and Schacht [23] directly (without proving Conjecture 3.6).

Theorem 3.8 (Embedding lemma for subgraphs of random graphs). *For every graph F with vertex set $V(F) = [\ell]$ and every $d > 0$ there exists $\varepsilon > 0$ such that for every $\eta > 0$ there exists $C > 0$ such that for $p > Cn^{-1/m_2(F)}$ a.a.s. $G \in G(n, p)$ satisfies the following.*

If $H = (V_1 \cup \dots \cup V_\ell, E_H)$ is an ℓ -partite (not necessarily induced) subgraph of G with $|V_1| = \dots = |V_\ell| \geq \eta n$ and with the property that for every edge $\{i, j\} \in E(F)$ the pair (V_i, V_j) in H is (ε, p) -regular and satisfies $d_H(V_i, V_j) \geq dp$, then H contains a partite copy of F , i.e., there exists a graph homomorphism $\varphi : F \rightarrow H$ with $\varphi(i) \in V_i$.

Proof. We deduce Theorem 3.8 from Theorem 3.7. In fact, it will follow by a standard first moment argument. Since the result is trivial for matchings F we may assume that $m_2(F) \geq 1$.

For given F and d we set

$$\alpha = \left(\frac{d}{2e}\right)^{e(F)},$$

where $e = 2.7182\dots$ is the base of the natural logarithm. Let $\varepsilon' > 0$ be given by the statement of Conjecture 3.6 applied with F , d , and α and set $\varepsilon = \varepsilon'/2$. Following the quantification of Theorem 3.8, we are given η . Finally, let $C' > 0$ be given by Conjecture 3.6 and set

$$b = d\eta^2 \quad \text{and} \quad C = \max \left\{ \frac{C'}{\eta^{1/m_2(F)}}, \frac{\ell}{b} \right\}.$$

Consider a graph $H' \subseteq G \in G(n, p)$ satisfying the assumptions of Theorem 3.8. Let $m \geq \eta n$ be the size of the vertex classes, V_1, \dots, V_ℓ , and set $M = dpm^2$. A straightforward application of Chernoff’s inequality asserts that H' contains a spanning subgraph H such that, for every $\{i, j\} \in E(F)$, the pair (V_i, V_j) is $(2\varepsilon, p)$ -regular, and $e_H(V_i, V_j) = M$. In other words, $H \in \mathcal{G}(F, m, M, 2\varepsilon, p, d)$ and it suffices to show that a.a.s. $G \in G(n, p)$ contains no graph H from $\mathcal{B}(F, m, M, 2\varepsilon, p, d)$.

For that we consider the expected number of subgraphs in G , which belong to $\mathcal{B}(F, m, M, 2\varepsilon, p, d)$ for some $m \geq \eta n$. For $m \geq \eta n$ fixed, our choice of constants allows us to appeal to the conclusion of Theorem 3.7, and we obtain the following upper bound for the expected number of such graphs:

$$p^{M\varepsilon(F)} \cdot |\mathcal{B}(F, m, M, 2\varepsilon, p, d)| \cdot \binom{n}{m}^\ell$$

$$\begin{aligned}
 &\leq p^{Me(F)} \cdot \alpha^M |\mathcal{G}(F, m, M, 2\varepsilon, p, d)| \cdot \binom{n}{m}^\ell \\
 &\leq p^{Me(F)} \left(\frac{d}{2e}\right)^{Me(F)} \binom{m^2}{M}^{e(F)} 2^{\ell n} \\
 &\leq \left(p \cdot \frac{d}{2e} \cdot \frac{e}{pd}\right)^{Me(F)} 2^{\ell n} \\
 &= 2^{\ell n - Me(F)} \\
 &\leq 2^{-bpn^2},
 \end{aligned}$$

where we used for the last estimate $M \geq dp(\eta n)^2$, $e(F) \geq 2$, and $b = d\eta^2$ combined with $\ell n \leq bpn^2$ (which follows from $m_2(F) \geq 1$ and $C \geq \ell/b$).

Summing the obtained bound over all possible values of m shows that the expected number of bad graphs in G is at most $n2^{-bpn^2}$, and hence, Markov’s inequality implies that a.a.s. $G \in G(n, p)$ contains no such graph. ■

Also the counting lemma of Fact 3.2 was partly extended to subgraphs in $G(n, p)$ in [23]. We state these results below.

Theorem 3.9 (Counting lemma for subgraphs of random graphs). *For every graph F with vertex set $V(F) = [\ell]$ and every $d > 0$ there exist $\varepsilon > 0$ and $\xi > 0$ such that for every $\eta > 0$ there exists $C > 0$ such that for $p > Cn^{-1/m_2(F)}$ a.a.s. $G \in G(n, p)$ satisfies the following holds.*

Let $H = (V_1 \cup \dots \cup V_\ell, E_H)$ be an ℓ -partite (not necessarily induced) subgraph of G with $|V_1| = \dots = |V_\ell| \geq \eta n$ and with the property that for every edge $\{i, j\} \in E(F)$ the pair (V_i, V_j) in H is (ε, p) -regular with density $d_H(V_i, V_j) \geq dp$.

(i) Then the number of partite copies of F in H is at least

$$(15) \quad \xi p^{e(F)} \prod_{i=1}^{\ell} |V_i|.$$

(ii) If in addition F is strictly 2-balanced, then the number of partite copies of F in H satisfies

$$(16) \quad (1 \pm f(\varepsilon)) p^{e(F)} \prod_{\{i,j\} \in E(F)} d(V_i, V_j) \prod_{i=1}^{\ell} |V_i|,$$

where $f(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

Let us briefly compare Theorems 3.7–3.9. Theorem 3.7, which was proved in [101], gives an affirmative answer to Conjecture 3.6 for all graphs F , and as we showed above, it implies Theorem 3.8. Also part (i) of Theorem 3.9 is a stronger version of Theorem 3.8. While Theorem 3.8 ensures only one copy of the given graph F in an appropriate (ε, p) -regular environment, part (i) of Theorem 3.9 guarantees a constant fraction of the “expected number” of copies of F . For strictly 2-balanced graphs F , part (ii) of Theorem 3.9 guarantees the expected number of copies of F , which can be viewed as the generalization of the counting lemma of Fact 3.2 for such graphs F .

Although Theorem 3.8 is the weakest result in this direction, it turns out to be sufficient for many natural applications of the regularity lemma or subgraphs of sparse random graphs (Theorem 3.4). For example, it allows new and conceptually simple proofs of Theorems 2.2 and 2.3 (see, e.g., Section 4.2 for such a proof of Theorem 2.3).

However, there are a few applications, where the full strength of Theorem 3.7 was needed. For example, following the proof from [62] (see also [82]), one can use the positive resolution of Conjecture 3.6 to prove the 1-statement of the threshold for the asymmetric Ramsey properties of random graphs (see Section 4.1), but Theorems 3.8 and 3.9 seem to be insufficient for this application. In Section 4 we will also mention some applications, which require the quantitative estimates of Theorem 3.9 (see Section 4.3 and 4.4).

Finally, we remark that $G(n, p)$ has the properties of Theorem 3.8 and of part (i) of Theorem 3.9 with probability $1 - 2^{-\Omega(pn^2)}$, while part (ii) of Theorem 3.9 holds with probability at least $1 - n^{-k}$ for any constant k and sufficiently large n (see [23]). Also we note that, due to the upper bound on the number of copies of F given in part (ii) of Theorem 3.9, an error probability of the form $2^{-\Omega(pn^2)}$ can not hold. This is because, for $\alpha(1) = p \gg 1/n$, the upper tail for the number of copies of a graph F (with at least as many edges as vertices) in $G(n, p)$ fails to have such a sharp concentration. In fact, the probability that $G(n, p)$ contains a clique of size $2pn$ is at least $p^{\binom{2pn}{2}} = 2^{-O(p^2 \log(1/p)n^2)} \gg 2^{-\Omega(pn^2)}$, and such a clique gives rise to $(2pn)^{|V(F)|} > 2p^{\varepsilon(F)} n^{|V(F)|}$ copies of F .

4. APPLICATIONS OF THE REGULARITY METHOD FOR RANDOM GRAPHS

In this section we show some examples how the regularity lemma and its counting and embedding lemmas for subgraphs of random graphs can be applied.

In Section 4.1 we briefly review thresholds for asymmetric Ramsey properties of random graphs. In particular, Theorem 3.7 can be used to establish the 1-statement for such properties. We remark that, even though this is a statement about $G(n, p)$, in the proof suggested by Kohayakawa and Kreuter [62] one applies the sparse regularity lemma to an auxiliary subgraph of $G(n, p)$ with density $o(p)$. As a result Theorems 3.8 and 3.9 cannot be applied anymore and an application of Theorem 3.7 is pivotal here.

In Section 4.2, we transfer the Erdős–Simonovits theorem (Theorem 1.3) to subgraphs of random graphs, i.e., we deduce Theorem 2.3. The proof given here is based on the sparse regularity lemma, and Theorem 3.8 suffices for this application. It also utilizes the Erdős–Simonovits stability theorem, which will be applied to the so-called *reduced graph*.

In Section 4.3 we discuss another application and extend the *removal lemma* (see Theorem 4.3 for the special case of triangles). The standard proof of the removal lemma is based on Szemerédi’s regularity lemma and the counting lemma of Fact 3.2. In fact, the embedding lemma seems not be sufficient for such a proof. The probabilistic version of the removal lemma for subgraphs of random graphs, Theorem 4.4, can be obtained by following the lines of the standard proof, where Szemerédi’s regularity lemma and the counting lemma of Fact 3.2 are replaced by the sparse regularity lemma (Theorem 3.4) and part (i) of Theorem 3.9.

In Section 4.4 we state the recent *clique density theorem* of Reiher [93] (see Theorem 4.5 below) and its probabilistic version for random graphs. In the proof of the probabilistic version the “right” counting lemma (part (ii) of Theorem 3.9), giving the expected number of copies of cliques in an appropriate regular environment is an essential tool. Moreover, the clique density theorem itself will be applied to the *weighted reduced graph*.

Finally in Section 4.5 we briefly discuss some connection between the theory of quasi-random graphs and the regularity lemma. In particular, we will mention a generalization of a result of Simonovits and Sós [107] for subgraphs of random graphs and a strengthening of part (ii) of Corollary 2.4.

4.1. Ramsey Properties of Random Graphs

Ramsey theory is another important field in discrete mathematics, which was influenced and shaped by Paul Erdős. His seminal work with Szekeres [41] laid the ground for a lot of the research in Ramsey theory. For example, Graham, Spencer, and Rothschild [55, page 26] stated that, “*It is difficult to overestimate the effect of this paper.*”

For an integer $r \geq 2$ and graphs F_1, \dots, F_r , we denote by $\mathcal{R}_n(F_1, \dots, F_r)$ the set of all n -vertex graphs G with the Ramsey property, i.e., the n -vertex graphs G with the property that for every r -coloring of the edges of G with colors $1, \dots, r$ there exists a color s such that G contains a copy of F_s with all edges colored with color s . Ramsey’s theorem [89] implies that $\mathcal{R}_n(F_1, \dots, F_r)$ is not empty for any r and all graphs F_1, \dots, F_r for sufficiently large n .

While probabilistic techniques in Ramsey theory were introduced by Erdős [29] in 1947, the investigation of Ramsey properties of the random graph $G(n, p)$ was initiated only in early 90’s by Łuczak, Ruciński, and Voigt [80]. In particular, one was interested in the threshold of $\mathcal{R}_n(F_1, \dots, F_r)$ for the symmetric case, i.e., $F_1 = \dots = F_r = F$, for which we use the short hand notation $\mathcal{R}_n(F; r)$. This question was addressed by Rödl and Ruciński [95, 96, 97]. There it was shown that $n^{-1/m_2(F)}$ is the threshold for $\mathcal{R}_n(F; r)$ for all graphs F containing a cycle and all integers $r \geq 2$. Note that the threshold is independent of the number of colors r . The proof of the 1-statement was based on an application of Szemerédi’s regularity lemma (Theorem 3.1) for dense graphs, even though the result appeals to sparse random graphs. Based on the recent embedding lemma for subgraphs of random graphs (Theorem 3.8) and a standard application of the sparse regularity lemma (Theorem 3.4) a conceptually simpler proof is now possible.

Below we discuss the asymmetric Ramsey properties, i.e., the case when not all F_i are the same graph. Here we restrict ourselves to the two-color case. Thresholds for asymmetric Ramsey properties involving cycles were obtained by Kohayakawa and Kreuter [62]. Furthermore, these authors put forward a conjecture for the threshold of $\mathcal{R}_n(F_1, F_2)$ for graphs F_1 and F_2 containing a cycle.

Conjecture 4.1. *Let F_1 and F_2 be graphs containing a cycle and $m_2(F_1) \leq m_2(F_2)$. Then $\hat{p} = n^{-1/m_2(F_1, F_2)}$ is a threshold for $\mathcal{R}_n(F_1, F_2)$, where*

$$m_2(F_1, F_2) = \max \left\{ \frac{e(F')}{|V(F')| - 2 + 1/m_2(F_1)} : F' \subseteq F_2 \text{ and } e(F') \geq 1 \right\}.$$

There is an intuition behind the definition of $m_2(F_1, F_2)$, which has some analogy to the definition of $m_2(F)$ in (2). One can first observe that

$$m_2(F, F) = m_2(F) \quad \text{and} \quad m_2(F_1) \leq m_2(F_1, F_2) \leq m_2(F_2).$$

Moreover, for $p \geq n^{-1/m_2(F_1, F_2)}$ the expected number of copies of F_2 (and all its subgraphs) in $G(n, p)$ is of the same order of magnitude as the expected number of edges $G(n, n^{-1/m_2(F_1)})$. Assuming that there is a two-coloring of $G(n, p)$ with no copy of F_2 with edges in color two, one may hope that picking an edge of color one in every copy of F_2 may result in a graph with “similar properties” as $G(n, n^{-1/m_2(F_1)})$. In particular, those edges should form a copy of F_1 in color one.

In [62] the 1-statement of Conjecture 4.1 for $\mathcal{R}_n(C, F)$ for any cycle C and any 2-balanced graph F with $m_2(C) \geq m_2(F)$ was verified. Moreover, the 0-statement was shown for the case when F_1 and F_2 are cliques [82], and the 1-statement was shown for graphs F_1 and F_2 with $m_2(F_1, F_2) > m_2(F_1, F')$ for every $F' \subsetneq F_2$ with $e(F') \geq 1$ appeared in [69]. In particular, those results yield the threshold for $\mathcal{R}(K_k, K_\ell)$.

It was also known that the resolution of Conjecture 3.6 for the (sparser) graph F_1 allows us to generalize the proof from [62] to verify the 1-statement of Conjecture 4.1 when F_2 is strictly 2-balanced (see, e.g., [82]). Therefore, Theorem 3.7 has the following consequence.

Theorem 4.2. *Let F_1 and F_2 be graphs with $1 \leq m_2(F_1) \leq m_2(F_2)$ and let F_2 be strictly 2-balanced. There exists a constant $C > 0$ such that for $p \geq Cn^{-1/m_2(F_1, F_2)}$ a.a.s. $G \in G(n, p)$ satisfies $G \in \mathcal{R}_n(F_1, F_2)$.*

4.2. Stability Theorem for Subgraphs of Random Graphs

Below we deduce a probabilistic version of the Erdős–Simonovits theorem from the classical stability theorem, based on the regularity method for subgraphs of random graphs.

Proof of Theorem 2.3. Let a graph F with chromatic number $\chi(F) \geq 3$ and $\varepsilon > 0$ be given. In order to deliver the promised constants C and δ , we have to fix some auxiliary constants. First we appeal to the Erdős–Simonovits stability theorem, Theorem 1.3, with F and $\varepsilon/8$ and obtain constants $\delta' > 0$ and n'_0 . Set

$$\delta = \delta'/3.$$

Moreover, set $d = \min\{\delta/4, \varepsilon/4\}$ and set $\varepsilon_{\text{RL}} = \min\{\delta/8, \varepsilon/8, \varepsilon_{\text{EMB}}\}$, where ε_{EMB} is given by Theorem 3.8 applied with F and d . Then apply the sparse

regularity lemma, Theorem 3.4, with ε_{RL} and $t_0 = \max\{n'_0, 4/\delta, 8/\varepsilon\}$ and obtain the constant T_0 . This gives us a lower bound of n/T_0 on the size of the partition classes after an application of Theorem 3.4. To a suitable collection of those classes, we will want to apply Theorem 3.8. Therefore, we set $\eta = 1/T_0$. Due to our choice of $\varepsilon_{\text{RL}} \leq \varepsilon_{\text{EMB}}$ Theorem 3.8 guarantees a constant $C = C(F, d, \varepsilon_{\text{RL}}, \eta)$ and we let $p \geq Cn^{-1/m_2(F)}$.

For later reference we observe that, due this choice of constants above, for every $t \geq t_0$ we have

$$(17) \quad \frac{t}{2} + d \binom{t}{2} + \varepsilon_{\text{RL}} t^2 < \delta \binom{t}{2}$$

and

$$(18) \quad \frac{1}{t} + \frac{d}{2} + \varepsilon_{\text{RL}} + \frac{\varepsilon}{8} \leq \frac{\varepsilon}{2}.$$

We split the argument below into a probabilistic and a deterministic part. First, in the probabilistic part, we single out a few properties (see (a)–(c) below), which the random graph $G \in G(n, p)$ has a.a.s. In the second, deterministic part, we deduce the stability result for all graphs G satisfying those properties.

In the probabilistic part we note that a.a.s. $G \in G(n, p)$ satisfies the following:

- (a) for all sets $X, Y \subseteq V(G)$ we have $e_G(X, Y) \leq (1 + o(1))p|X||Y|$, where the edges contained in $X \cap Y$ are counted twice,
- (b) G satisfies the conclusion of Theorem 3.4 for $\varepsilon_{\text{RL}}, t_0$, and T_0 ,
- (c) G satisfies the conclusion of Theorem 3.8 for $F, d, \varepsilon_{\text{RL}}, \eta$ and C .

Property (a) follows a.a.s. by a standard application of Chernoff’s inequality, and properties (b) and (c) hold a.a.s. due to Theorems 3.4 and 3.8.

In the deterministic part we deduce the conclusion of Theorem 2.3 for all graphs satisfying properties (a)–(c). To this end, let $G = (V, E)$ be a graph with these properties. Consider an F -free subgraph $H \subseteq G$ with

$$e(H) \geq \text{ex}_G(F) - \delta pn^2.$$

We will show that we can remove at most εpn^2 edges from H , so that the remaining graph is $(\chi(F) - 1)$ -colorable.

Since every graph G contains a $(\chi(F) - 1)$ -cut (see Definition 2.6) of size at least

$$\left(1 - \frac{1}{\chi(F) - 1}\right) e(G) = \pi(F)e(G),$$

it follows from property (a) that

$$(19) \quad e(H) \geq \pi(F)p \binom{n}{2} - 2\delta pn^2.$$

We appeal to property (b), which ensures the existence of a partition $V_1 \cup \dots \cup V_t = V$ having properties (i)–(iii) of Theorem 3.4 for ε_{RL} , t_0 , and T_0 . Without loss of generality, we may assume that t divides n since removing at most t vertices from H affects only $O(tn) = o(pn^2)$ edges.

For the given partition, we consider the so-called *reduced graph* $R = R(H, \varepsilon_{RL}, d)$ with vertex set $[t]$. The pair $\{i, j\}$ is an edge in R if, and only if the pair (V_i, V_j) is (ε_{RL}, p) -regular and $d_H(V_i, V_j) \geq dp$. Note that R does not represent the following edges of H :

- (I) edges which are contained in some V_i ,
- (II) edges which are contained in a pair (V_i, V_j) which is not (ε, p) -regular, and
- (III) edges which are contained in a pair (V_i, V_j) with $d_H(V_i, V_j) < dp$.

Owing to property (a) we infer, that there are at most

$$(20) \quad t \cdot (1 + o(1))p \binom{n/t}{2}$$

edges described in (I) and at most

$$(21) \quad \varepsilon_{RL}t^2 \cdot (1 + o(1))p \left(\frac{n}{t}\right)^2$$

edges described in (II). By definition at most

$$(22) \quad \binom{t}{2} \cdot dp \left(\frac{n}{t}\right)^2$$

edges of H are contained in pairs described in (III).

Moreover, since (again because of property (a))

$$e_H(V_i, V_j) \leq e_G(V_i, V_j) \leq (1 + o(1))p \binom{n}{t}$$

it follows from the definition of R , that the number of edges in R satisfies

$$e(R) \geq \frac{e(H) - t(1 + o(1))p \binom{n/t}{2} - \varepsilon_{RL}t^2(1 + o(1))p \left(\frac{n}{t}\right)^2 - \binom{t}{2}dp \left(\frac{n}{t}\right)^2}{(1 + o(1))p(n/t)^2}$$

$$\stackrel{(17),(19)}{\geq} (\pi(F) - 3\delta) \binom{t}{2} = (\pi(F) - \delta') \binom{t}{2}.$$

Moreover, property (c) implies that R is F -free, since otherwise a copy of F in R would lead to a copy of F in H . In particular, R satisfies the assumptions of the classical Erdős–Simonovits stability theorem, Theorem 1.3. Recall, that $\delta' > 0$ was given by an application of Theorem 1.3 applied with F and $\varepsilon/8$. We will only need the weaker assertion of Theorem 1.3, which concerns the deletion of edges rather than the symmetric difference. Consequently, we may remove up to at most $(\varepsilon/8)t^2$ edges from R , so that the resulting graph R' is $(\chi(F) - 1)$ -colorable. Let $f : [t] \rightarrow [\chi(F) - 1]$ be such a coloring of R' and consider the corresponding partition $W_1 \cup \dots \cup W_{\chi(F)-1} = V$ of H given by

$$W_i = \bigcup \{V_j : j \in f^{-1}(i)\}.$$

It is left to show that

$$\sum_{i=1}^{\chi(F)-1} e_H(W_i) \leq \varepsilon pn^2.$$

Note that there besides the three types of edges described in (I)–(III) the following type of edges of H could be contained in $E_H(W_i)$ for some $i \in [\chi(F) - 1]$

- (IV) edges which are contained in a pair (V_i, V_j) for some $\{i, j\} \in E(R) \setminus E(R')$.

Again property (a) combined with $|E(R) \setminus E(R')| \leq (\varepsilon/8)t^2$ implies that there are at most

$$(23) \quad \frac{\varepsilon}{8}t^2 \cdot (1 + o(1))p \left(\frac{n}{t}\right)^2$$

edges described in (IV). Finally, the desired bound follows from (20)–(23)

$$\begin{aligned} \sum_{i=1}^{\chi(F)-1} e_H(W_i) &\leq t \cdot (1 + o(1))p \binom{n/t}{2} \\ &\quad + \varepsilon_{RL}t^2 \cdot (1 + o(1))p \left(\frac{n}{t}\right)^2 \\ &\quad + \binom{t}{2} \cdot dp \left(\frac{n}{t}\right)^2 \\ &\quad + \frac{\varepsilon}{8}t^2 \cdot (1 + o(1))p \left(\frac{n}{t}\right)^2 \end{aligned}$$

$$\begin{aligned} &\leq (1 + o(1))(1/t + d/2 + \varepsilon_{RL} + \varepsilon/8)pn^2 \\ &\stackrel{(18)}{\leq} \varepsilon pn^2. \end{aligned}$$

This concludes the proof of Theorem 2.3. ■

We remark that there are several other classical results involving forbidden subgraphs F , which can be transferred to subgraphs of random graphs, using a very similar approach, i.e., by applying the classical result to a suitably chosen reduced graph R . For example, the 1-statement of Theorem 2.2 or the 1-statement of the Ramsey threshold from [97] can be reproved by such an approach. In the next section, we discuss an example, where one can obtain the probabilistic result by “repeating” the original proof with the sparse regularity lemma and a matching, embedding or counting lemma replacing Szemerédi’s regularity lemma and Fact 3.2.

4.3. Removal Lemma for Subgraphs of Random Graphs

In one of the first applications of an earlier variant of the regularity lemma, Ruzsa and Szemerédi [99] answered a question of Brown, Sós, and Erdős [18] and essentially established the following removal lemma for triangles.

Theorem 4.3 (Ruzsa & Szemerédi, 1978). *For every $\varepsilon > 0$ there exist $\delta > 0$ and n_0 such that every graph $G = (V, E)$ with $|V| = n \geq n_0$ containing at most δn^3 copies of K_3 can be made K_3 -free by omission of at most εn^2 edges.*

In fact, the same statement holds, when K_3 is replaced by any graph F and δn^3 is replaced by $\delta n^{|V(F)|}$, as was shown by Füredi [47] (see also [2] for the case when F is a clique and [36] for related results). This result is now known as the *removal lemma* for graphs (we refer to the recent survey of Conlon and Fox [21] for a thorough discussion of its importance and its generalizations).

The following probabilistic version for subgraphs of random graphs was suggested by Łuczak [79] and first proved for strictly 2-balanced graphs F by Conlon and Gowers [22]. The general statement for all F follows from the work in [23].

Theorem 4.4. *For every graph F with ℓ vertices and $\varepsilon > 0$ there exist $\delta > 0$ and $C > 0$ such that for $p \geq Cn^{-1/m_2(F)}$ a.a.s. for $G \in G(n, p)$ the following holds. If $H \subseteq G$ contains at most $\delta p^{\varepsilon(F)} n^\ell$ copies of F , then H can be made F -free by omission of at most εpn^2 edges.*

Proof. Let a graph F with $V(F) = [\ell]$ vertices and $\varepsilon > 0$ be given. Since the result is trivial for matchings F , we may assume that $m_2(F) \geq 1$. We will apply the counting lemma for subgraphs of $G(n, p)$ given by part (i) of Theorem 3.9. We prepare for such an application with F by setting $d = \varepsilon/6$ and choosing $\varepsilon_{\text{RL}} = \min\{\varepsilon/6, \varepsilon_{\text{CL}}/\ell\}$, where ε_{CL} is given by Theorem 3.9. Moreover, we set $t_0 = 3/\varepsilon$ and let T_0 be given by the sparse regularity lemma, Theorem 3.4, applied with ε_{RL} and t_0 . We then follow the quantification of Theorem 3.9. For that we set $\eta = (T_0\ell)^{-1}$ and let $C > 0$ be given by Theorem 3.9. Finally, we set

$$(24) \quad \delta = \frac{1}{2}\xi p^{e(F)}\eta^\ell.$$

For later reference we observe that due this choice of constants above, for every $t \geq t_0$ and sufficiently large n we have

$$(25) \quad t \binom{n/t}{2} + 2d \binom{t}{2} \left(\frac{n}{t}\right)^2 + \varepsilon_{\text{RL}}n^2 \leq \frac{\varepsilon}{2}n^2.$$

Similarly, as in the proof given in Section 4.2, we split the argument in a probabilistic and a deterministic part. For the probabilistic part we note that a.a.s. $G \in G(n, p)$ satisfies the following:

- (A) for all sets $X, Y \subseteq V(G)$ we have $e_G(X, Y) \leq (1 + o(1))p|X||Y|$, where the edges contained in $X \cap Y$ are counted twice,
- (B) G satisfies the conclusion of Theorem 3.4 for $\varepsilon_{\text{RL}}, t_0$, and T_0 ,
- (C) G satisfies the conclusion of part (i) of Theorem 3.9 for $F, d - \varepsilon_{\text{RL}}, \varepsilon_{\text{RL}}, \xi, \eta$, and C .

Again property (A) follows a.a.s. by a standard application of Chernoff’s inequality and properties (B) and (C) hold a.a.s. due to Theorems 3.4 and 3.9.

It is left to deduce the conclusion of Theorem 4.4 for any graph $G = (V, E)$ satisfying properties (A)–(C) and with sufficiently large $n = |V|$. Let $H \subseteq G$ containing at most $\delta p^{e(F)}n^\ell$ copies of F .

Next we appeal to property (B), which ensures the existence of a partition $V_1 \cup \dots \cup V_t = V$ having properties (i)–(iii) of Theorem 3.4 for $\varepsilon_{\text{RL}}, t_0$, and T_0 . Without loss of generality we may assume that $t\ell$ divides n since removing at most $t\ell$ vertices from H affects only $O(t\ell n) = o(pn^2)$ edges.

We remove the following edges from H :

- edges which are contained in some V_i ,
- edges which are contained in a pair (V_i, V_j) with $d_H(V_i, V_i) < 2dp$, and
- edges which are contained in a pair (V_i, V_j) which is not (ε, p) -regular.

Let H' be the resulting subgraph. Owing to property (A) we obtain

$$\begin{aligned}
 & e(H) \setminus e(H') \\
 & \leq t \cdot (1 + o(1))p \binom{n/t}{2} + \binom{t}{2} \cdot 2dp \left(\frac{n}{t}\right)^2 + \varepsilon_{\text{RL}} t^2 \cdot (1 + o(1))p \left(\frac{n}{t}\right)^2 \\
 & \stackrel{(25)}{\leq} (1 + o(1)) \frac{\varepsilon}{2} p n^2 \leq \varepsilon p n^2.
 \end{aligned}$$

It is left to show that H' is F -free. Suppose for a contradiction that H' contains a copy of F . Let V_{i_1}, \dots, V_{i_k} be the vertex classes, that contain a vertex from this copy.

Note that if $k = \ell$, i.e., each class contains exactly one vertex from F , then the ℓ -partite induced subgraph $H'[V_{i_1}, \dots, V_{i_\ell}]$ meets the assumptions of part (i) of Theorem 3.9 for the constants $2d > d$, $\varepsilon_{\text{RL}} < \varepsilon_{\text{CL}}$, ξ , η , and C fixed above. Consequently, it follows from property (C) that H' , and hence also H , contains at least

$$\xi p^{e(F)} \left(\frac{n}{t}\right)^\ell \geq \xi p^{e(F)} (\eta \ell)^\ell \cdot n^\ell \stackrel{(24)}{>} \delta p^{e(F)} n^\ell$$

copies of F , which contradicts the assumptions on H .

If $k < \ell$, then subdivide every V_{i_j} for $1 \leq j \leq k$ into ℓ disjoint sets of size $|V_{i_j}|/\ell$ in such a way that every subclass created this way contains at most one vertex of the given copy of F in H . Let W_1, \dots, W_ℓ be the classes containing one vertex of the copy of F and we may assume that W_i contains the copy of vertex i of F . Note that for every $i \in V(F)$ the set W_i has size at least $n/(t\ell) \geq \eta n$. Moreover, if $\{i, j\} \in E(F)$, then $H'[W_i, W_j]$ contains at least one edge. In particular, this edge is contained in H' and, hence, it signifies that (W_i, W_j) is contained in some pair $(V_{i'}, V_{j'})$, which has density at least dp and which is $(\varepsilon_{\text{RL}}, p)$ -regular. Moreover, it follows from the definition of $(\varepsilon_{\text{RL}}, p)$ -regularity (see Definition 3.3), that (W_i, W_j) is still $(\ell\varepsilon_{\text{RL}}, p)$ -regular and has density at least $(d - \varepsilon_{\text{RL}})p$. In other words, $H'[W_1, \dots, W_\ell]$ is ready for an application of of Theorem 3.9 for the constants $2d - \varepsilon_{\text{RL}} \geq d$, $\ell\varepsilon_{\text{RL}} \leq \varepsilon_{\text{CL}}$, ξ , η , and C fixed above. Consequently, it follows from property (C) that H' , and hence also H , contains at least

$$\xi p^{e(F)} \left(\frac{n}{t\ell}\right)^\ell \geq \xi p^{e(F)} \eta^\ell \cdot n^\ell \stackrel{(24)}{>} \delta p^{e(F)} n^\ell$$

copies of F , which also in this case contradicts the assumptions on H and concludes the proof of Theorem 4.4. ■

4.4. Clique Density Theorem for Subgraphs of Random Graphs

Turán’s theorem establishes the minimum number $\text{ex}_n(K_k) + 1$ of edges in an n -vertex graph that implies the existence of a copy of K_k . For the triangle case, it was proved by Hans Rademacher (unpublished) that every n -vertex graph with $\text{ex}_n(K_3) + 1$ edges contains not only one, but at least $n/2$ triangles. More generally, Erdős suggested to study the minimum number of triangles in n -vertex graphs with $\text{ex}_n(K_3) + s$ edges [30, 32]. He conjectured that for $s < n/4$ there are at least $s\lfloor n/2 \rfloor$ triangles, which is best possible due to the graph obtained by balanced, complete bipartite graph with s independent edges in the vertex with $\lfloor n/2 \rfloor$ vertices. This conjecture was proved by Lovász and Simonovits [76] (see also [60]). For larger values of s and k this problem was studied by Erdős [33], Moon and Moser [83], Nordhaus and Stewart [85], Bollobás [10, 11], and Khadzhiiyanov and Nikiforov [59].

In particular, in [77] Lovász and Simonovits formulated a conjecture which relates the minimum density of K_k with a given edges density. More precisely, for an integer $k \geq 3$ and a graph H , let $\mathcal{K}_k(H)$ be the number of (unlabeled) copies of K_k in H . We denote by $\mathcal{K}_k(n, M)$ the minimum over all graphs with n vertices and M edges, i.e.,

$$\mathcal{K}_k(n, M) = \min\{\mathcal{K}_k(H) : |V(H)| = n \text{ and } |E(H)| = M\}.$$

In [76] Lovász and Simonovits conjectured that the extremal graph for $\mathcal{K}_k(n, M)$ is obtained from complete t -partite graph (for some appropriate t) by adding a matching to one of the vertex classes. In [77] those authors proposed an approximate version of this conjecture by considering densities of cliques and edges instead relating the number of cliques with the number of edges. For that we define for $\alpha \in [0, 1)$

$$\kappa_k(\alpha) = \liminf_{n \rightarrow \infty} \frac{\mathcal{K}_k(n, \lceil \alpha \binom{n}{2} \rceil)}{\binom{n}{k}},$$

i.e., n -vertex graphs with $\alpha \binom{n}{2}$ edges contain at least $(\kappa_k(\alpha) - o(1)) \binom{n}{k}$ copies of K_k and $\kappa_k(\alpha)$ is the largest clique density which can be guaranteed. Clearly, $\kappa_k(\cdot)$ is non-decreasing and for $\alpha \in [0, \pi(K_k))$ we have $\kappa_k(\alpha) = 0$. Lovász and Simonovits suggested that the graphs described below attain the infimum of $\kappa_k(\alpha)$: For a given $\alpha > 0$, let t be the integer with the property

$$(26) \quad \alpha \in \left(1 - \frac{1}{t}, 1 - \frac{1}{t+1} \right]$$

and set $\varrho \in \mathbb{R}$ to the smaller root of the quadratic equation

$$2\varrho(1 - \varrho) + \left(1 - \frac{1}{t}\right) (1 - \varrho)^2 = \alpha.$$

One can check that (26) implies $0 < \varrho \leq \frac{1}{t+1}$. We then define the graphs $T_{n,\alpha}$ to be the complete $(t + 1)$ -partite graph with vertex classes $V_1 \cup \dots \cup V_{t+1} = V(T_{n,\alpha})$ satisfying

$$|V_{t+1}| = \lceil \varrho n \rceil \quad \text{and} \quad |V_1| \leq \dots \leq |V_t| \leq |V_1| + 1.$$

Maybe, a more intuitive description of these graphs is the following. For edge densities α of the form $1 - 1/t$ the graph $T_{n,\alpha}$ is the Turán graph $T_{n,t}$ with t classes. For $\alpha \in \left(1 - \frac{1}{t}, 1 - \frac{1}{t+1}\right)$ a “small” $(t + 1)$ st class of size ϱn appears and all other classes have size $(1 - \varrho)n/t$. With α tending to $1 - \frac{1}{t+1}$ the difference in size between the $(t + 1)$ st class and the other classes becomes smaller. Finally, for $\alpha = 1 - \frac{1}{t+1}$ we get $\varrho = 1/(t + 1)$ and $T_{n,\alpha}$ becomes the Turán graph with $t + 1$ classes.

Lovász and Simonovits conjectured that for every k and α

$$(27) \quad \kappa_k(\alpha) = \lim_{n \rightarrow \infty} \frac{\mathcal{K}_k(T_{n,\alpha})}{\binom{n}{k}}.$$

We remark that the conjectured extremal graph $T_{n,\alpha}$ is independent of the size of clique K_k . This conjecture was known to be true in the “symmetric case,” i.e., for densities $\alpha \in \{1 - 1/t : t \in \mathbb{N}\}$, due to the work of Moon and Moser [83] (see [85] for the triangle case). Fisher addressed (27) for $k = 3$ and $1/2 \leq \alpha \leq 2/3$.

A few years ago Razborov introduced the so-called *flag algebra method* in extremal combinatorics [91] (see [90] for a survey on the topic) and based on this calculus he solved the triangle case for every $\alpha \in (0, 1)$ in [92]. This work was followed by Nikiforov [84], which led to the solution of the case $k = 4$ and finally Reiher [93] verified the conjecture for every k .

Theorem 4.5 (Clique density theorem). *For every integer $k \geq 3$ and for every $\alpha \in (0, 1)$ we have*

$$\kappa_k(\alpha) = \lim_{n \rightarrow \infty} \frac{\mathcal{K}_k(T_{n,\alpha})}{\binom{n}{k}}.$$

Based on the counting lemma for subgraphs of random graphs, part (ii) of Theorem 3.9, one can use the sparse regularity lemma to transfer this result to subgraphs of random graphs. The following appears in [23].

Theorem 4.6. *For every graph $k \geq 3$ and $\delta > 0$ there exists $C > 0$ such that for $p \geq Cn^{-1/m_2(K_k)}$ the following holds a.a.s. for $G \in G(n, p)$. If $H \subseteq G$ contains at least $(\alpha + \delta)e(G)$ edges, then*

$$\mathcal{K}_k(H) \geq \kappa_k(\alpha)\mathcal{K}_k(G).$$

As mentioned above, the proof of Theorem 4.6 is based on the regularity method for subgraphs of random graphs and relies on the counting lemma giving the “expected number” of copies of K_k in an appropriate (ε, p) -regular environment. Moreover, in the proof a *weighted version* of the clique density theorem, Theorem 4.5 is applied to the weighted reduced graph (see [23] for details).

4.5. Quasi-random Subgraphs of Random Graphs

In this section we discuss scaled versions of the Chung-Graham-Wilson theorem [20] on quasi-random graphs for subgraphs a random graphs. The systematic study of quasi-random graphs was initiated by Thomason [112, 113] and Chung, Graham, and Wilson [20] (see also [1, 44, 94] for partial earlier results and [75] for a recent survey on the topic). In [20] several properties of dense random graphs, i.e., properties a.a.s. satisfied by $G(n, p)$ for $p > 0$ independent of n , were shown to be equivalent in a deterministic sense. This phenomenon fails to be true for $p = o(1)$ (see, e.g., [19, 66]). For relatively dense subgraphs of sparse random graphs however several deterministic equivalences among (appropriately scaled) quasi-random properties remain valid. Below we will discuss one such equivalence (see Theorem 4.9 below), whose analog for dense graphs was obtained by Simonovits and Sós [107].

Before we mention the result of Simonovits and Sós we begin with the following quasi-random properties of graphs, concerning the edge distribution (see DISC below) and the number of copies (or embeddings) of given graph F (see EMB below).

Definition 4.7. Let F be a graph on ℓ vertices and let $d > 0$.

DISC: We say a graph $H = (V, E)$ with $|V| = n$ satisfies $\text{DISC}(d)$, if for every subset $U \subseteq V$ we have

$$e_H(U) = d \binom{|U|}{2} \pm o(n^2).$$

EMB: We say a graph $H = (V, E)$ with $|V| = n$ satisfies $\text{EMB}(F, d)$, if the number $N_F(H)$ of labeled copies of F in H satisfies

$$N_F(H) = d^{e(F)} n^\ell \pm o(n^\ell).$$

It is well known that the property $\text{DISC}(d)$ implies the property $\text{EMB}(F, d)$ for every graph F . By this we mean that for every $\varepsilon > 0$ there exist $\delta > 0$ and n_0 such that every n -vertex ($n \geq n_0$) graph H satisfying $\text{DISC}(d)$ with $o(n^2)$ replaced by δn^2 also satisfies property $\text{EMB}(F, d)$ with $o(n^\ell)$ replaced by εn^ℓ .

The opposite implication is known to be false. For example, for $F = C_\ell$ being a cycle of length ℓ we may consider n -vertex graphs H consisting of a clique of size dn and isolated vertices. Such a graph H satisfies $\text{EMB}(C_\ell, d)$, but fails to have $\text{DISC}(d)$. However, note that such a graph H fails to have density d . If we add this as an additional condition, then for even ℓ one of the main implications of the Chung-Graham-Wilson theorem asserts the implication $\text{EMB}(C_\ell, d)$ implies $\text{DISC}(d)$.

For odd cycles imposing global density d does not suffice, as the following interesting example from [20] shows: Partition the vertex set $V(H) = V_1 \cup V_2 \cup V_3 \cup V_4$ as equal as possible into four sets and add the edges of the complete graph on V_1 and on V_2 , add edges of the complete bipartite graph between V_3 and V_4 , and add edges of a random bipartite graph with edge probability $1/2$ between $V_1 \cup V_2$ and $V_3 \cup V_4$. One may check that a.a.s. such a graph H has density $d(H) \geq 1/2 - o(1)$ and it satisfies $\text{EMB}(K_3, 1/2)$, but clearly it fails to have $\text{DISC}(1/2)$.

Summarizing the discussion above, while $\text{DISC}(d)$ implies $\text{EMB}(F, d)$ is known to be true for every graph F , $\text{EMB}(C_{2\ell+1}, d)$ does not imply $\text{DISC}(d)$. In fact, $\text{EMB}(C_{2\ell+1}, d) \not\Rightarrow \text{DISC}(d)$ even when restricting to graphs H with density d .

Note that the property $\text{DISC}(d)$ is *hereditary* in the sense that for subsets $U \subseteq V$ the induced subgraph $H[U]$ must also satisfy $\text{DISC}(d)$, if H has $\text{DISC}(d)$. As a result the implication $\text{DISC}(d) \Rightarrow \text{EMB}(F, d)$ extends to the following hereditary strengthening of $\text{EMB}(F, d)$.

HEMB: We say a graph $H = (V, E)$ with $|V| = n$ satisfies $\text{HEMB}(F, d)$, if for every $U \subseteq V$ the number $N_F(H[U])$ of labeled copies of F in the induced subgraph $H[U]$ satisfies

$$(28) \quad N_F(H[U]) = d^{e(F)}|U|^\ell \pm o(n^\ell).$$

It was shown by Simonovits and Sós [107] (see also [115] for a recent strengthening) that **HEMB** indeed is a quasi-random property, i.e., those authors showed that for every graph F with at least one edge and for every $d > 0$ the properties $\text{DISC}(d)$ and $\text{HEMB}(F, d)$ are equivalent, i.e., $\text{DISC}(d) \Rightarrow \text{HEMB}(d)$ and $\text{HEMB}(d) \Rightarrow \text{DISC}(d)$.

Based on the sparse regularity lemma (Theorem 3.4) and its appropriate counting lemma (part (ii) of Theorem 3.9) a generalization of the

Simonovits–Sós theorem for subgraphs of random graphs $G \in G(n, p)$ can be derived. First we introduce the appropriate sparse versions of DISC and HEMB for this context.

Definition 4.8. Let $G = (V, E)$ be a graph with $|V| = n$, let F be a graph on ℓ vertices, and let $d > 0$ and $\varepsilon > 0$.

DISC $_G$: We say a subgraph $H \subseteq G$ satisfies DISC $_G(d)$, if for every subset $U \subseteq V$ we have

$$e_H(U) = d|E(G[U])| \pm o(|E|),$$

i.e., the relative density of $H[U]$ with respect to $G[U]$ is close to d for all sets U of linear size. Furthermore, we say H satisfies DISC $_G(d, \varepsilon)$ if $e_H(U) = d|E(G[U])| \pm \varepsilon|E|$.

HEMB $_G$: We say a subgraph $H \subseteq G$ satisfies HEMB $_G(F, d)$, if for every $U \subseteq V$ the number $N_F(H[U])$ of labeled copies of F in the induced subgraph $H[U]$ satisfies

$$N_F(H[U]) = d^{e(F)} N_F(G[U]) \pm o(|N_F(G)|),$$

i.e., approximately a $d^{e(F)}$ proportion of the copies of F in $G[U]$ is contained in $H[U]$ for sets U spanning a constant proportion of copies of F in G .

Furthermore, we say a subgraph $H \subseteq G$ satisfies HEMB $_G(F, d, \varepsilon)$, if for every $U \subseteq V$ we have $N_F(H[U]) = d^{e(F)} N_F(G[U]) \pm \varepsilon|N_F(G)|$.

For those properties one can prove an equivalence in the sense described above, when G is a random graph.

Theorem 4.9. Let F be a strictly 2-balanced graph with at least one edge and let $d > 0$. For every $\varepsilon > 0$ there exist $\delta > 0$ and $C > 0$ such that for $p \geq Cn^{-1/m_2(F)}$, a.a.s. for $G \in G(n, p)$ the following holds.

- (i) If $H \subseteq G$ satisfies DISC $_G(d, \delta)$, then H satisfies HEMB $_G(F, d, \varepsilon)$.
- (ii) If $H \subseteq G$ satisfies HEMB $_G(F, d, \delta)$, then H satisfies DISC $_G(d, \varepsilon)$.

Consequently, for $p \gg n^{-1/m_2(F)}$ a.a.s. $G \in G(n, p)$ has the property that DISC $_G(d)$ and HEMB $_G(F, d)$ are equivalent.

We will briefly sketch some ideas from the proofs of both implications of the Simonovits–Sós theorem and indicate its adjustments for the proof of Theorem 4.9.

The implication DISC $(d) \Rightarrow$ HEMB (F, d) (for dense graphs) easily follows from the counting lemma in Fact 3.2. Indeed, given $U \subseteq V(H)$ for a

graph H satisfying $\text{DISC}(d)$, we consider a partition of U into $\ell = |V(F)|$ classes $U_1 \cup \dots \cup U_\ell$ with sizes as equal as possible. Based on the identity

$$e_H(U_i, U_j) = e_H(U_i \cup U_j) - e_H(U_i) - e_H(U_j)$$

we infer from $\text{DISC}(d)$ that (U_i, U_j) is $g(\varepsilon)$ -regular and $d_H(U_i, U_j) = d \pm o(1)$, where $g(\varepsilon)$ tends to 0 with the error parameter ε from the property $\text{DISC}(d)$. In particular, for sufficiently small $\varepsilon > 0$ the assumptions of the counting lemma of Fact 3.2 are met for F and d . Using the upper and lower bound on the number of partite copies of F in $U_1 \cup \dots \cup U_\ell$ provided by the counting lemma and a simple averaging argument over all possible partitions $U_1 \cup \dots \cup U_\ell$ yields (28). This simple argument with part (ii) of Theorem 3.9 replacing Fact 3.2 can be transferred to $G(n, p)$ without any further adjustments.

The proof of the opposite implication, $\text{HEMB}(F, d) \Rightarrow \text{DISC}(d)$, is more involved. All known proofs for dense graphs are based on Szemerédi’s regularity lemma (Theorem 3.1) and the counting lemma in Fact 3.2. The proof of Simonovits and Sós requires not only an applications of Fact 3.2 for F , but also for a graph obtained from F by taking two copies of F and identifying one of their edges. Note that this “double- F ” is not strictly 2-balanced, since it contains F as a proper subgraph, which has the same 2-density as double- F . Consequently, we run into some difficulties, if we want to extend this proof to subgraphs of random graphs based on part (ii) of Theorem 3.9. In some recent generalizations of the Simonovits–Sós theorem applications of Fact 3.2 for double- F could be avoided (see, e.g., [103, 24, 25]). In particular, the proof presented in [25, pages 174-175] extends to subgraphs of random graphs, by replacing Szemerédi’s regularity lemma and the counting lemma of Fact 3.2 by its counterparts for sparse random graphs.

4.5.1. Problem of Erdős and Nešetřil revisited. We close this section by returning to the question of Erdős and Nešetřil from Section 2.2, which perhaps led to one of the first extremal results for random graphs.

Here we want to focus on generalizations of part (ii) of Corollary 2.4. That statement asserts that any K_{k+1} -free graph H with the additional property

$$(29) \quad \text{ex}_H(K_k) \leq (\pi(K_k) + o(1))e(H)$$

must have vanishing density $d(H) = o(1)$.

Based on Theorem 3.8 the following generalization can be proved. Consider the random graph $G \in G(n, p)$ for $p \gg n^{-1/m_2(K_{k+1})}$. We will show

that a.a.s. G has the property that any K_{k+1} -free graph $H \subseteq G$ satisfying (29) must have vanishing *relative density* (w.r.t. the density of G), i.e., $d(H) = o(p)$.

Theorem 4.10 (Generalization of Corollary 2.4(ii) for subgraphs of $G(n, p)$). *For every integer $k \geq 3$, every $d > 0$, and every $\varepsilon \in (0, 1 - \pi(K_k))$ there exists some $C > 0$ such that for $p > Cn^{-1/m_2(K_{k+1})}$ the following holds a.a.s. for $G \in G(n, p)$. If $H \subseteq G$ satisfies $e(H) = d|E(G)|$ and $\text{ex}_H(K_k) \leq (\pi(K_k) + \varepsilon)e(H)$, then H contains a K_{k+1} .*

The proof of Theorem 4.10 follows the lines of the proof of Corollary 2.4(ii) given in Section 2.2 and we briefly sketch the main adjustments needed. Recall that the proof given in Section 2.2 relied on Lemma 2.8 from [94], which is based on the embedding lemma of Fact 3.2. Replacing the embedding lemma for dense graphs by the appropriate version for subgraphs of random graphs, i.e., by Theorem 3.8, yields the following.

Lemma 4.11. *For all integers $s, t \geq 2$ and every $d > 0$ there exist $\delta > 0$ and $C > 0$ such that for $p > Cn^{-1/m_2(K_s)}$ the following holds a.a.s. for $G \in G(n, p)$.*

If $H \subseteq G$ satisfying $e_H(U) = (d \pm \delta)e_G(U)$ for every $U \subseteq V$ with $|U| = \lfloor n/t \rfloor$. Then H contains a copy of K_s .

Equipped with Lemma 4.11 one can repeat the proof of Lemma 2.7 and the following appropriate version for subgraphs of $G(n, p)$ can be verified.

Lemma 4.12. *For all integers $s, t \geq 2$ and every $d > 0$ there exist $\varepsilon > 0$ and $C > 0$ such that for $p > Cn^{-1/m_2(K_s)}$ the following holds a.a.s. for $G \in G(n, p)$.*

If $H \subseteq G$ with $e(H) = d|E(G)|$ and with the property that every balanced t -cut has size at most $(1 - 1/t + \varepsilon)de(G)$, then H contains a copy of K_s .

Finally, a standard application of Lemma 4.12 with $s = k + 1$ and $t = k - 1$ yields Theorem 4.10. We omit the details here.

5. CONCLUDING REMARKS

We close with a few comments of related results and open problems.

Related Results. We restricted the discussion to extremal question in random graphs. However, the results of Conlon and Gowers [22] and

Schacht [102] and also the subsequent work of Samotij [100], Balogh, Morris, Samotij [8], and Saxton and Thomason [101] applied in a more general context and led to extremal results for random hypergraphs and random subsets of the integers. Here we state a probabilistic version of Szemerédi's theorem on arithmetic progressions [110] (see Theorem 5.1 below).

For integers $k \geq 3$ and $n \in \mathbb{N}$, and a set $A \subseteq \mathbb{Z}/n\mathbb{Z}$, let $r_k(A)$ denote the cardinality of a maximum subset of A , which contains no arithmetic progression of length k , i.e.,

$$r_k(A) = \{ |B| : B \subseteq A \text{ and } B \text{ contains} \\ \text{no arithmetic progression of length } k \}.$$

Answering a well known conjecture of Erdős and Turán [42], Szemerédi's theorem asserts that

$$r_k(\mathbb{Z}/n\mathbb{Z}) = o(n)$$

for every integer $k \geq 3$. The following probabilistic version of Szemerédi's theorem was obtained for $k = 3$ by Kohayakawa, Łuczak, and Rödl [64] and for all k in [22, 102].

Theorem 5.1. *For every integer $k \geq 3$ and every $\varepsilon > 0$ the function $\hat{p} = n^{-1/(k-1)}$ is a threshold for $\mathcal{S}_n(k, \varepsilon) = \{A \subseteq \mathbb{Z}/n\mathbb{Z} : r_k(A) \leq \varepsilon|A|\}$.*

Note that similarly as for the threshold for the Erdős–Stone theorem for random graphs, the threshold for Szemerédi's theorem coincides with that p for which a random subset of $\mathbb{Z}/n\mathbb{Z}$ has in expectation the same number of elements and number of arithmetic progressions of length k .

Let us remark that the methods from [22, 102] also can be used to derive thresholds for Ramsey properties for random hypergraphs and random subsets of the integers (see [22, 45] for details).

Open Problems. Besides these recent advances, several important questions are still unresolved. For example, it would be very interesting if the result of DeMarco and Kahn [26] (Theorem 2.9) could be extended to cliques of arbitrary size (see Conjecture 2.10). Finally, we would like to point out that for some applications (see, e.g., Theorem 4.9) a generalization of part (ii) of Theorem 3.9 for all graphs F would be useful.

REFERENCES

- [1] N. Alon and F. R. K. Chung, *Explicit construction of linear sized tolerant networks*, Discrete Math., **72** (1988), no. 1–3, 15–19.
- [2] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster, *The algorithmic aspects of the regularity lemma*, J. Algorithms, **16** (1994), no. 1, 80–109.
- [3] N. Alon and J. H. Spencer, *The probabilistic method*, third ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., Hoboken, NJ, 2008, With an appendix on the life and work of Paul Erdős.
- [4] L. Babai, M. Simonovits, and J. Spencer, *Extremal subgraphs of random graphs*, J. Graph Theory, **14** (1990), no. 5, 599–622.
- [5] J. Balogh, B. Bollobás, and M. Simonovits, *The number of graphs without forbidden subgraphs*, J. Combin. Theory Ser. B, **91** (2004), no. 1, 1–24.
- [6] ———, *The typical structure of graphs without given excluded subgraphs*, Random Structures Algorithms, **34** (2009), no. 3, 305–318.
- [7] ———, *The fine structure of octahedron-free graphs*, J. Combin. Theory Ser. B, **101** (2011), no. 2, 67–84.
- [8] J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, submitted.
- [9] J. Balogh, R. Morris, W. Samotij, and L. Warnke, *The typical structure of sparse k_{r+1} -free graphs*, in preparation.
- [10] B. Bollobás, *On complete subgraphs of different orders*, Math. Proc. Cambridge Philos. Soc., **79** (1976), no. 1, 19–24.
- [11] ———, *Relations between sets of complete subgraphs*, Proceedings of the Fifth British Combinatorial Conference (Univ. Aberdeen, Aberdeen, 1975), Congressus Numerantium, No. XV, Utilitas Math., Winnipeg, Man., 1976, pp. 79–84.
- [12] ———, *Extremal graph theory*, London Mathematical Society Monographs, vol. 11, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1978.
- [13] ———, *Modern graph theory*, Graduate Texts in Mathematics, vol. 184, Springer-Verlag, New York, 1998.
- [14] ———, *Random graphs*, second ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.
- [15] B. Bollobás and A. Thomason, *Threshold functions*, Combinatorica **7** (1987), no. 1, 35–38.
- [16] J. A. Bondy and U. S. R. Murty, *Graph theory*, Graduate Texts in Mathematics, vol. 244, Springer, New York, 2008.
- [17] G. Brightwell, K. Panagiotou, and A. Steger, *Extremal subgraphs of random graphs*, Random Structures Algorithms, **41** (2012), no. 2, 147–178.
- [18] W. G. Brown, P. Erdős, and V. T. Sós, *Some extremal problems on r -graphs*, New directions in the theory of graphs (Proc. Third Ann Arbor Conf., Univ. Michigan, Ann Arbor, Mich, 1971), Academic Press, New York, 1973, pp. 53–63.
- [19] F. R. K. Chung and R. L. Graham, *Sparse quasi-random graphs*, Combinatorica **22** (2002), no. 2, 217–244, Special issue: Paul Erdős and his mathematics.
- [20] F. R. K. Chung, R. L. Graham, and R. M. Wilson, *Quasi-random graphs*, Combinatorica, **9** (1989), no. 4, 345–362.

- [21] D. Conlon and J. Fox, *Graph removal lemmas*, submitted.
- [22] D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, submitted.
- [23] D. Conlon, W. T. Gowers, W. Samotij, and M. Schacht, *On the KLR conjecture in random graphs*, manuscript.
- [24] D. Conlon, H. Hàn, Y. Person, and M. Schacht, *Weak quasi-randomness for uniform hypergraphs*, *Random Structures Algorithms*, **40** (2012), no. 1, 1–38.
- [25] D. Dellamonica, Jr. and V. Rödl, *Hereditary quasirandom properties of hypergraphs*, *Combinatorica*, **31** (2011), no. 2, 165–182.
- [26] B. DeMarco and J. Kahn, *Mantel’s theorem for random graphs*, submitted.
- [27] R. Diestel, *Graph theory*, fourth ed., Graduate Texts in Mathematics, vol. 173, Springer, Heidelberg, 2010.
- [28] P. Erdős, *On sequences of integers no one of which divides the product of two others and on some related problems*, *Mitt. Forsch.-Inst. Math. und Mech. Univ. Tomsk*, **2** (1938), 74–82.
- [29] ———, *Some remarks on the theory of graphs*, *Bull. Amer. Math. Soc.* **53** (1947), 292–294.
- [30] ———, *Some theorems on graphs*, *Riveon Lematematika*, **9** (1955), 13–17.
- [31] ———, *Graph theory and probability*, *Canad. J. Math.*, **11** (1959), 34–38.
- [32] ———, *On a theorem of Rademacher-Turán*, *Illinois J. Math.* **6** (1962), 122–127.
- [33] ———, *On the number of complete subgraphs contained in certain graphs*, *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **7** (1962), 459–464.
- [34] ———, *Some recent results on extremal problems in graph theory. Results*, *Theory of Graphs (Internat. Sympos., Rome, 1966)*, Gordon and Breach, New York, 1967, pp. 117–123 (English); pp. 124–130 (French).
- [35] ———, *On some of my conjectures in number theory and combinatorics*, *Proceedings of the fourteenth Southeastern conference on combinatorics, graph theory and computing (Boca Raton, Fla., 1983)*, vol. 39, 1983, pp. 3–19.
- [36] P. Erdős, P. Frankl, and V. Rödl, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, *Graphs Combin.*, **2** (1986), no. 2, 113–121.
- [37] P. Erdős, D. J. Kleitman, and B. L. Rothschild, *Asymptotic enumeration of K_n -free graphs*, *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973)*, Tomo II, Accad. Naz. Lincei, Rome, 1976, pp. 19–27. *Atti dei Convegni Lincei*, No. 17.
- [38] P. Erdős and A. Rényi, *On random graphs. I*, *Publ. Math. Debrecen* **6** (1959), 290–297.
- [39] P. Erdős and M. Simonovits, *An extremal graph problem*, *Acta Math. Acad. Sci. Hungar.*, **22** (1971/72), 275–282.
- [40] P. Erdős and A. H. Stone, *On the structure of linear graphs*, *Bull. Amer. Math. Soc.*, **52** (1946), 1087–1091.
- [41] P. Erdős and G. Szekeres, *A combinatorial problem in geometry*, *Compositio Math.*, **2** (1935), 463–470.
- [42] P. Erdős and P. Turán, *On some sequences of integers.*, *J. Lond. Math. Soc.*, **11** (1936), 261–264.

- [43] P. Frankl and V. Rödl, *Large triangle-free subgraphs in graphs without K_4* , Graphs Combin., **2** (1986), no. 2, 135–144.
- [44] P. Frankl, V. Rödl, and R. M. Wilson, *The number of submatrices of a given type in a Hadamard matrix and related results*, J. Combin. Theory Ser. B **44** (1988), no. 3, 317–328.
- [45] E. Friedgut, V. Rödl, and M. Schacht, *Ramsey properties of random discrete structures*, Random Structures Algorithms, **37** (2010), no. 4, 407–436.
- [46] Z. Füredi, *Random Ramsey graphs for the four-cycle*, Discrete Math. **126** (1994), no. 1–3, 407–410.
- [47] ———, *Extremal hypergraphs and combinatorial geometry*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994) (Basel), Birkhäuser, 1995, pp. 1343–1352.
- [48] St. Gerke, *Random graphs with constraints*, Habilitationsschrift, Institut für Informatik, Technische Universität München, 2005.
- [49] St. Gerke, Y. Kohayakawa, V. Rödl, and A. Steger, *Small subsets inherit sparse ϵ -regularity*, J. Combin. Theory Ser. B, **97** (2007), no. 1, 34–56.
- [50] St. Gerke, M. Marciniszyn, and A. Steger, *A probabilistic counting lemma for complete graphs*, Random Structures Algorithms, **31** (2007), no. 4, 517–534.
- [51] St. Gerke, H. J. Prömel, T. Schickinger, A. Steger, and A. Taraz, *K_4 -free subgraphs of random graphs revisited*, Combinatorica **27** (2007), no. 3, 329–365.
- [52] St. Gerke, T. Schickinger, and A. Steger, *K_5 -free subgraphs of random graphs*, Random Structures Algorithms, **24** (2004), no. 2, 194–232.
- [53] St. Gerke and A. Steger, *The sparse regularity lemma and its applications*, Surveys in combinatorics 2005, London Math. Soc. Lecture Note Ser., vol. 327, Cambridge Univ. Press, Cambridge, 2005, pp. 227–258.
- [54] ———, *A characterization for sparse ϵ -regular pairs*, Electron. J. Combin., **14** (2007), no. 1, Research Paper 4, 12 pp. (electronic).
- [55] R. L. Graham, B. L. Rothschild, and J. H. Spencer, *Ramsey theory*, second ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., New York, 1990, A Wiley-Interscience Publication.
- [56] P. E. Haxell, Y. Kohayakawa, and T. Łuczak, *Turán’s extremal problem in random graphs: forbidding even cycles*, J. Combin. Theory Ser. B, **64** (1995), no. 2, 273–287.
- [57] ———, *Turán’s extremal problem in random graphs: forbidding odd cycles*, Combinatorica, **16** (1996), no. 1, 107–122.
- [58] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.
- [59] N. G. Khadzhivanov and V. S. Nikiforov, *The Nordhaus–Stewart–Moon–Moser inequality*, Serdica, **4** (1978), no. 4, 344–350.
- [60] ———, *Solution of the problem of P. Erdős on the number of triangles in graphs with n vertices and $\lfloor n^2/4 \rfloor + l$ edges*, C. R. Acad. Bulgare Sci., **34** (1981), no. 7, 969–970.
- [61] Y. Kohayakawa, *Szemerédi’s regularity lemma for sparse graphs*, Foundations of computational mathematics (Rio de Janeiro, 1997), Springer, Berlin, 1997, pp. 216–230. MR 1661982 (99g:05145)

- [62] Y. Kohayakawa and B. Kreuter, *Threshold functions for asymmetric Ramsey properties involving cycles*, Random Structures Algorithms, **11** (1997), no. 3, 245–276.
- [63] Y. Kohayakawa, B. Kreuter, and A. Steger, *An extremal problem for random graphs and the number of graphs with large even-girth*, Combinatorica **18** (1998), no. 1, 101–120.
- [64] Y. Kohayakawa, T. Łuczak, and V. Rödl, *Arithmetic progressions of length three in subsets of a random set*, Acta Arith., **75** (1996), no. 2, 133–163.
- [65] ———, *On K^4 -free subgraphs of random graphs*, Combinatorica **17** (1997), no. 2, 173–213.
- [66] Y. Kohayakawa and V. Rödl, *Regular pairs in sparse random graphs. I*, Random Structures Algorithms, **22** (2003), no. 4, 359–434. MR 1980964 (2004b:05187)
- [67] ———, *Szemerédi’s regularity lemma and quasi-randomness*, Recent advances in algorithms and combinatorics, CMS Books Math./Ouvrages Math. SMC, vol. 11, Springer, New York, 2003, pp. 289–351.
- [68] Y. Kohayakawa, V. Rödl, and M. Schacht, *The Turán theorem for random graphs*, Combin. Probab. Comput., **13** (2004), no. 1, 61–91.
- [69] Y. Kohayakawa, M. Schacht, and R. Spöhel, *Upper bounds on probability thresholds for asymmetric Ramsey properties*, Random Structures Algorithms, to appear.
- [70] Ph. G. Kolaitis, H. J. Prömel, and B. L. Rothschild, *Asymptotic enumeration and a 0-1 law for m -clique free graphs*, Bull. Amer. Math. Soc. (N.S.), **13** (1985), no. 2, 160–162.
- [71] ———, *K_{i+1} -free graphs: asymptotic structure and a 0-1 law*, Trans. Amer. Math. Soc., **303** (1987), no. 2, 637–671.
- [72] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi, *The regularity lemma and its applications in graph theory*, Theoretical aspects of computer science (Tehran, 2000), Lecture Notes in Comput. Sci., vol. 2292, Springer, Berlin, 2002, pp. 84–112.
- [73] J. Komlós and M. Simonovits, *Szemerédi’s regularity lemma and its applications in graph theory*, Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc. Math. Stud., vol. 2, János Bolyai Math. Soc., Budapest, 1996, pp. 295–352.
- [74] T. Kövari, V. T. Sós, and P. Turán, *On a problem of K. Zarankiewicz*, Colloquium Math., **3** (1954), 50–57.
- [75] M. Krivelevich and B. Sudakov, *Pseudo-random graphs*, More sets, graphs and numbers, Bolyai Soc. Math. Stud., vol. 15, Springer, Berlin, 2006, pp. 199–262.
- [76] L. Lovász and M. Simonovits, *On the number of complete subgraphs of a graph*, Proceedings of the Fifth British Combinatorial Conference (Univ. Aberdeen, Aberdeen, 1975) (Winnipeg, Man.), Utilitas Math., 1976, pp. 431–441. Congressus Numerantium, No. XV.
- [77] ———, *On the number of complete subgraphs of a graph. II*, Studies in pure mathematics, Birkhäuser, Basel, 1983, pp. 459–495.
- [78] T. Łuczak, *On triangle-free random graphs*, Random Structures Algorithms **16** (2000), no. 3, 260–276.
- [79] ———, *Randomness and regularity*, International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, pp. 899–909.

- [80] T. Łuczak, A. Ruciński, and B. Voigt, *Ramsey properties of random graphs*, J. Combin. Theory Ser. B, **56** (1992), no. 1, 55–68.
- [81] W. Mantel, *Vraagstuk XXVIII*, Wiskundige Opgaven, **10** (1907), 60–61.
- [82] M. Marciniszyn, J. Skokan, R. Spöhel, and A. Steger, *Asymmetric Ramsey properties of random graphs involving cliques*, Random Structures Algorithms **34** (2009), no. 4, 419–453.
- [83] J. W. Moon and L. Moser, *On a problem of Turán*, Magyar Tud. Akad. Mat. Kutató Int. Közl., **7** (1962), 283–286.
- [84] V. Nikiforov, *The number of cliques in graphs of given order and size*, Trans. Amer. Math. Soc., **363** (2011), no. 3, 1599–1618.
- [85] E. A. Nordhaus and B. M. Stewart, *Triangles in an ordinary graph*, Canad. J. Math., **15** (1963), 33–41.
- [86] D. Osthus, H. J. Prömel, and A. Taraz, *For which densities are random triangle-free graphs almost surely bipartite?*, Combinatorica, **23** (2003), no. 1, 105–150, Paul Erdős and his mathematics (Budapest, 1999).
- [87] H. J. Prömel and A. Steger, *The asymptotic number of graphs not containing a fixed color-critical subgraph*, Combinatorica, **12** (1992), no. 4, 463–473.
- [88] ———, *On the asymptotic structure of sparse triangle free graphs*, J. Graph Theory, **21** (1996), no. 2, 137–151.
- [89] F. P. Ramsey, *On a problem in formal logic*, Proc. Lond. Math. Soc. (2) **30** (1930), 264–286.
- [90] A. A. Razborov, *Flag algebras: an interim report*, submitted.
- [91] ———, *Flag algebras*, J. Symbolic Logic, **72** (2007), no. 4, 1239–1282.
- [92] ———, *On the minimal density of triangles in graphs*, Combin. Probab. Comput., **17** (2008), no. 4, 603–618.
- [93] Chr. Reiher, *The clique density theorem*, submitted.
- [94] V. Rödl, *On universality of graphs with uniformly distributed edges*, Discrete Math., **59** (1986), no. 1–2, 125–134.
- [95] V. Rödl and A. Ruciński, *Lower bounds on probability thresholds for Ramsey properties*, Combinatorics, Paul Erdős is eighty, Vol. 1, Bolyai Soc. Math. Stud., János Bolyai Math. Soc., Budapest, 1993, pp. 317–346.
- [96] ———, *Random graphs with monochromatic triangles in every edge coloring*, Random Structures Algorithms, **5** (1994), no. 2, 253–270.
- [97] ———, *Threshold functions for Ramsey properties*, J. Amer. Math. Soc., **8** (1995), no. 4, 917–942.
- [98] V. Rödl and M. Schacht, *Regularity lemmas for graphs*, Fete of combinatorics and computer science, Bolyai Soc. Math. Stud., vol. 20, János Bolyai Math. Soc., Budapest, 2010, pp. 287–325.
- [99] I. Z. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, Colloq. Math. Soc. János Bolyai, vol. 18, North-Holland, Amsterdam, 1978, pp. 939–945.
- [100] W. Samotij, *Stability results for random discrete structures*, Random Structures Algorithms, to appear.

- [101] D. Saxton and A. Thomason, *Hypergraph containers*, submitted.
- [102] M. Schacht, *Extremal results for random discrete structures*, submitted.
- [103] A. Shapira, *Quasi-randomness and the distribution of copies of a fixed graph*, *Combinatorica*, **28** (2008), no. 6, 735–745.
- [104] M. Simonovits, *A method for solving extremal problems in graph theory, stability problems*, *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, Academic Press, New York, 1968, pp. 279–319.
- [105] ———, *External graph problems with symmetrical extremal graphs. Additional chromatic conditions*, *Discrete Math.*, **7** (1974), 349–376.
- [106] ———, *The extremal graph problem of the icosahedron*, *J. Combinatorial Theory Ser. B*, **17** (1974), 69–79.
- [107] M. Simonovits and V. T. Sós, *Hereditarily extended properties, quasi-random graphs and not necessarily induced subgraphs*, *Combinatorica* **17** (1997), no. 4, 577–596.
- [108] A. Steger, *On the evolution of triangle-free graphs*, *Combin. Probab. Comput.* **14** (2005), no. 1–2, 211–224.
- [109] T. Szabó and V. H. Vu, *Turán’s theorem in sparse random graphs*, *Random Structures Algorithms*, **23** (2003), no. 3, 225–234.
- [110] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, *Acta Arith.*, **27** (1975), 199–245, Collection of articles in memory of Jurij Vladimirovič Linnik.
- [111] ———, *Regular partitions of graphs*, *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, *Colloq. Internat. CNRS*, vol. 260, CNRS, Paris, 1978, pp. 399–401.
- [112] A. Thomason, *Pseudorandom graphs*, *Random graphs ’85 (Poznań, 1985)*, *North-Holland Math. Stud.*, vol. 144, North-Holland, Amsterdam, 1987, pp. 307–331.
- [113] ———, *Random graphs, strongly regular graphs and pseudorandom graphs*, *Surveys in combinatorics 1987 (New Cross, 1987)*, *London Math. Soc. Lecture Note Ser.*, vol. 123, Cambridge Univ. Press, Cambridge, 1987, pp. 173–195.
- [114] P. Turán, *Eine Extremalaufgabe aus der Graphentheorie*, *Mat. Fiz. Lapok*, **48** (1941), 436–452.
- [115] R. Yuster, *Quasi-randomness is determined by the distribution of copies of a fixed graph in equicardinal large sets*, *Combinatorica*, **30** (2010), no. 2, 239–246.

Vojtěch Rödl

*Department of Mathematics and
Computer Science,
Emory University,
Atlanta,
GA 30322,
USA*

e-mail: rodl@mathcs.emory.edu

Mathias Schacht

*Fachbereich Mathematik,
Universität Hamburg,
Bundesstraße 55,
D-20146 Hamburg,
Germany*

e-mail:

schacht@math.uni-hamburg.de

ERDŐS'S WORK ON THE SUM OF DIVISORS FUNCTION AND ON EULER'S FUNCTION

ANDRZEJ SCHINZEL

The following notation will be used throughout

$$\sigma(n) = \sum_{d|n} d, \quad \varphi(n) = \sum_{\substack{m \leq n \\ (m,n)=1}} 1, \quad d(n) = \sum_{d|n} 1, \quad \nu(n) = \sum_{\substack{p|n \\ p \text{ prime}}} 1 = \omega(n),$$

$\log_r x$ is the r times iterated logarithm of x ,

$$\begin{aligned} \sigma_1(n) &= \sigma(n), & \sigma_k(n) &= \sigma(\sigma_{k-1}(n)), \\ \varphi_1(n) &= \varphi(n), & \varphi_k(n) &= \varphi(\varphi_{k-1}(n)), \\ s_1(n) &= s(n) = \sigma(n) - n, & s_i(n) &= s_1(s_{i-1}(n)); \end{aligned}$$

γ is Euler's constant, $\alpha_0 = \log \prod_{p \text{ prime}} \left(1 - \frac{1}{p}\right)^{-1/p}$. Density is always the asymptotic density.

I. Papers concerning both σ and φ .

- [1] On a problem of Chowla and some related problems, Proc. Cambridge Philos. Soc. 32 (1936), 530–540.
- [2] (with L. Alaoglu), A conjecture in elementary number theory, Bull. Amer. Math. Soc. 50 (1944), 881–882.
- [3] (with L. Alaoglu), On highly composite and similar numbers, Trans. Amer. Math. Soc. 56 (1944), 448–469.
- [4] On perfect and multiply perfect numbers, Ann. Mat. Pura Appl. (4) 42 (1956), 253–258.
- [5] Remarks on two problems in Mat. Lapok (Hungarian), Mat. Lapok 7 (1956), 10–17.
- [6] Solution of two problems of Jankowska, Bull. Acad. Polon. Sci. sér. sci. math. phys. astronom. 6 (1958), 545–547.
- [7] Some remarks on Euler's φ -function, Acta Arith. 4 (1958), 10–19.

- [8] Remarks on number theory II: Some problems on the σ function, *ibid.* 5 (1959), 171–177.
- [9] Remarks on number theory II: Some remarks on Euler's φ -function, (Hungarian), *Mat. Lapok* 12 (1961), 161–169.
- [10] Some remarks on the functions φ and σ , *Bull. Acad. Polon. Sci. sér. sci. math. phys. astronom.* 10 (1962), 617–619.
- [11] Some remarks on the iterates of the φ and σ function, *Colloq. Math.* 17 (1967), 195–202.
- [12] Asymptotische Untersuchungen über die Anzahl der Teiler von n , *Math. Ann.* 169 (1967), 230–238.
- [13] (with M. V. Subbarao), On the iterates of some arithmetic functions. The theory of arithmetic functions, *Lecture Notes in Math.* 251, 119–125, Springer Verlag, Berlin 1972.
- [14] Über die Zahlen der Form $\sigma(n) - n$ and $n - \varphi(n)$, *Elem. Math.* 28 (1973), 83–86.
- [15] Remarks on some problems in number theory, *Math. Balkanica* 4 (1974), 197–202.
- [16] On the distribution of numbers of the form $\sigma(n)/n$ and on some related questions, *Pacific J. Math.* 52 (1974), 59–65.
- [17] (with H. G. Diamond), A measure of nonmonotonicity of the Euler's φ -function, *ibid.* 77 (1978), 83–101.
- [18] (with K. Györy, Z. Papp), On some new properties of functions $\sigma(n)$, $\varphi(n)$, $d(n)$ and $\nu(n)$ (Hungarian), *Mat. Lapok* 28 (1980), 125–131.
- [19] Some applications of the probability theory to number theory, 4th Panonian Symp. of Math. Statistics, Austria 1982 vol. B, 1–18, Reidel, London 1985.
- [20] (with C. Pomerance, A. Sárközy), On locally repeated values of certain arithmetic functions, *Acta Math. Hungar.* 49 (1987), 251–259.
- [21] (with A. Granville, C. Pomerance and C. Spiro), On the normal behavior of the iterates of some arithmetic functions, *Analytic Number Theory, Progr. Math.* 85 (1990), 165–204.

In [1] Erdős proves for $f(n) = \sigma(n)$ or $\varphi(n)$ that

1. density of n such that $f(n) < f(n+1)$ and density of n such that $f(n) > f(n+1)$ are both $1/2$.

In [2] authors study the conjecture of Poulet [25] that for every n the sequence $f_0(n) = n$, $f_{2k+1}(n) = \sigma(f_{2k}(n))$, $f_{2k+2}(n) = \varphi(f_{2k+1}(n))$ is eventually periodic and claim that it holds for all $n \leq 10^4$. They prove that

2. for every $\varepsilon > 0$, $\varphi(\sigma(n)) < \varepsilon n$ except for a set of density 0 and claim that

1*. for every $c > 0$, $\sigma(\varphi(n)) > cn$ except for a set of density 0.

Moreover

2*. except for a set density zero,

$$e^\gamma \varphi(\sigma(n)) \log_3 n \sim \sigma(n) \quad \text{and} \quad e^{-\gamma} \sigma(\varphi(n)) \log_3 n \sim \varphi(n).$$

Luca and Pomerance [16] proved claim 1* and the second part of 2*.

Much later Pomerance [24] (see also Ford [9]) proved that for a certain $c > 0$ there are no exceptions.

[2] ends with the following conjectures.

- C1. Form the sequence $\sigma(n), \sigma(\sigma(n)), \varphi(\sigma(\sigma(n)))$, in which the functions are successively applied in the order $\sigma, \sigma, \varphi, \sigma, \sigma, \varphi, \sigma, \sigma, \dots$. This sequence seems to tend to infinity if n is large enough.
- C2. The sequence $\varphi(n), \varphi(\varphi(n)), \sigma(\varphi(\varphi(n))), \dots$, in which the order is $\varphi, \varphi, \sigma, \varphi, \varphi, \sigma, \varphi, \varphi, \sigma, \dots$ seems to tend to 1.

In **[3]** the following notions are used. A number n is called highly composite, if $d(m) < d(n)$ for all $m < n$, highly abundant if $\sigma(m) < \sigma(n)$ for all $m < n$, superabundant if $\sigma(m)/m < \sigma(n)/n$ for all $m < n$, colossally abundant if, for a certain $\varepsilon > 0$, $\sigma(m)/m^{1+\varepsilon}$ attains its maximum at $m = n$. Twenty theorems are proved, some of them highly technical; less technical results are as follows.

3. If n is superabundant and $n = 2^{k_2} 3^{k_3} \dots p^{k_p}$, then $k_2 \geq k_3 \geq \dots \geq k_p$.
4. If p is the largest prime factor of a superabundant number n , then $k_p = 1$, except when $n = 4, 36$.
5. The quotient of two consecutive superabundant numbers tends to 1.
6. The number of superabundant numbers less than x exceeds $c \log x \log_2 x / (\log_3 x)^2$, where c is a positive constant.
7. The quotient of two consecutive colossally abundant numbers is either a prime or a product of two primes.
8. Only a finite number of highly abundant numbers are highly composite.
9. For large x the number of highly abundant numbers $\leq x$ exceeds

$$(1 - \varepsilon)(\log x)^2 \quad \text{for every } \varepsilon > 0.$$

This result, as well as 6, has been later improved in **[32]**, **[33]**. It is claimed that

- 3*. the number of highly abundant numbers $\leq x$ is less than $(\log x)^{c \log_2 x}$, where c is a constant,
- 4*. the quotient of two consecutive numbers n such that $\varphi(n) < \varphi(m)$ for all $m > n$ tends to 1

and conjectured that

- C3. the number of highly abundant numbers $\leq x$ is less than

$$(\log x)^{c \log_3 x} \quad \text{where } c \text{ is a constant,}$$

- C4. there are infinitely many highly abundant numbers, which are not superabundant,
- C5. if $n > 1$ is a superabundant number, there are two primes p and q such that np and n/q are superabundant.

Tables of highly abundant numbers less than 10^4 and of superabundant numbers less than 10^{18} are included.

In [4] Erdős proves the following theorems.

- 10. The number $P(x)$ of positive integers $n \leq x$ such that $n|\sigma(n)$ satisfies

$$P(x) = O(x^{\frac{3}{4}+\epsilon}) \quad \text{for every } \epsilon > 0.$$

- 11. The number $P_2(x)$ of positive integers $n \leq x$ such that $\sigma(n) = 2n$ satisfies

$$P_2(x) < x^{(1-c_1)/2} \quad \text{for } x > x_0 \text{ and a certain constant } c_1 > 0$$

He claims the following.

- 5*. Let $f(x)$ be an increasing function satisfying $f(x) > (\log x)^{c_4}$ for some $c_4 > 0$. Then the number of positive integers $n < x$ satisfying

$$(\sigma(n), n) > f(x)$$

is less than $c_6x/f(x)^{c_3}$ for some $c_3 > 0$ and $c_6 > 0$. The same result holds if $\sigma(n)$ is replaced by $\varphi(n)$.

- 6*. 5* is best possible in the following sense: let $f(x) = o((\log x)^\epsilon)$ for every $\epsilon > 0$. Then the number of integers $n < x$ satisfying $(\sigma(n), n) > f(x)$ is greater than $x/f(x)^{c_5}$ for every $c_5 > 0$, if x is sufficiently large.

- 7*. The density of integers n satisfying

$$(\sigma(n), n) < (\log_2 n)^\alpha$$

equals $g(\alpha)$, where $g(\alpha)$, $0 \leq \alpha < \infty$ is an increasing function satisfying $g(0) = 0$, $g(\infty) = 1$. The same result holds if $\sigma(n)$ is replaced by $\varphi(n)$.

The estimates given in 10 and 11 have been improved by Wirsing [27] to the form

$$P(x) = O(\exp(c \log x / \log_2 x)), \quad c \text{ a positive constant.}$$

The claim 5* has been substantiated and improved by Pollack [22] in the part concerning $\sigma(n)$. He has replaced the assumption

$$f(x) > (\log x)^{c_4} \text{ by } f(x) > \exp((\log_2 x)^\beta) \text{ for } \beta > 0 \text{ and } x > x_0(\beta).$$

By the same token by disproved 6*.

In [5] the following notation is used: $L_a = \log_3 a / \log_4 a$. Erdős proves that

12. for every $\varepsilon > 0$ and $n > n_0(\varepsilon)$ there exists $a < n$ for which

$$\varphi(a) + \varphi(a + 1) + \dots + \varphi(a + \lfloor L_a \rfloor) < \varepsilon a,$$

13. for every $\eta > 0$

$$\liminf_{a \rightarrow \infty} \frac{\varphi(a) + \varphi(a + 1) + \dots + \varphi(a + \lceil (1 + \eta)L_a \rceil)}{a} = \infty.$$

He claims the following results.

8*.

$$\liminf_{a \rightarrow \infty} \frac{\varphi(a) + \varphi(a + 1) + \dots + \varphi\left(a + \left\lfloor \frac{\log a}{\log_2 a - \log_3 a} + \frac{c \log_3 a}{(\log_3 a)} \right\rfloor\right)}{a} = e^{c - \alpha_0}$$

for every c .

9*. Let $k_n = \left\lfloor \frac{\log_3 n}{\log_4 n} \right\rfloor$ and let i_1, i_2, \dots, i_{k_n} be any permutation of the integers $1, 2, \dots, k_n$. Then for $n > n_0$ there exists $a < n$ such that

$$\varphi(a + i_1) > \varphi(a + i_2) > \dots > \varphi(a + i_{k_n}).$$

10*. For every $\varepsilon > 0$ and $n > n_0(\varepsilon)$

$$\varphi(n) > \varphi(n + 1) > \dots > \varphi(n + \lfloor (1 + \varepsilon)k_n \rfloor)$$

cannot hold.

11*. The above results hold for $\sigma(n)$ instead of $\varphi(n)$.

[6] contains a proof that

14. there exist infinitely many pairs of integers a and b satisfying $(a, b) = 1$,

$$\varphi(a) = \varphi(b), \quad \sigma(a) = \sigma(b), \quad d(a) = d(b),$$

15. for every k there exists a sequence of distinct square-free integers a_1, \dots, a_k satisfying

$$\varphi(a_i) = \varphi(a_j), \quad \sigma(a_i) = \sigma(a_j) \text{ and } d(a_i) = d(a_j) \text{ for all } 1 \leq i \leq j \leq k.$$

The paper ends with the following conjectures.

C6. For every k there exists a sequence $x_i, 1 \leq i \leq k$, of distinct integers satisfying

- (1) $(x_i, x_j) = 1, \quad 1 \leq i \leq j \leq k,$
- (2) $\varphi(x_1) = \dots = \varphi(x_k),$
- (3) $\sigma(x_1) = \dots = \sigma(x_k),$

$$(4) \quad d(x_1) = \dots = d(x_k).$$

C7. The number of solutions of the system of conditions

$$(a, b) = 1, \quad a < b < n, \quad \varphi(a) = \varphi(b)$$

is $> n^{2-\varepsilon}$ for every $\varepsilon > 0$ if $n > n_0(\varepsilon)$.

The conjecture C6 is studied in [10] under the additional condition that x_i are square-free. Erdős proves that

16. C6 holds if the k integers a_i are required to satisfy only (1), (2) and (4) or only (1), (3) and (4).

In [7] Erdős proves the following theorems.

17. Let $f(n)$ tend to infinity so that

$$f(n) \leq \log_3 n / \log_6 n + (\alpha_0 - \gamma + o(1)) \log_3 n / (\log_6 n)^2.$$

Then

$$\liminf_{n \rightarrow \infty} \left(\max_{1 \leq i \leq f(n)} \varphi(n+i) / \min_{1 \leq j \leq f(n)} \varphi(n+j) \right) = 1$$

18. Let $A(n)$ denote the number of solutions of $\varphi(l) = n$. If there exists an integer n with $A(n) = k$, then there exist infinitely many such integers.

He claims the following theorems

12*. Put $f(n) = \log_3 n / \log_6 n + (c + \alpha_0 - \gamma) \log_3 n / (\log_6 n)^2$ ($c > 0$).

Then

$$\liminf_{n \rightarrow \infty} \left(\max_{1 \leq i \leq f(n)} \varphi(n+i) / \min_{1 \leq j \leq f(n)} \varphi(n+j) \right) = e^c.$$

13*. Let $\lim_{n \rightarrow \infty} g(n) / \log_5 n = 0$. Then there exists an infinite sequence n_k such that for all $1 \leq i \leq g(n_k)$

$$1 - \varepsilon_k \leq \frac{\varphi(n_k + i)}{\varphi(n_k + i - 1)} < 1 + \varepsilon_k, \quad \text{where } \varepsilon_k \rightarrow 0, \text{ as } k \rightarrow \infty.$$

14*. $A(n) < n \exp(-c \log n \log_3 n / \log_2 n)$, $c > 0$.

He makes the following conjectures:

C8. for every $\varepsilon > 0$ there exists an increasing sequence n_k such that

$$A(n_k) > n_k^{1-\varepsilon},$$

C9. for every $\varepsilon > 0$ and $n > n_0(\varepsilon)$

$$\sum_{k=1}^n A(k)^2 > n^{2-\varepsilon}.$$

and claims that

15*. all the results stated hold for $\sigma(n)$
and the same unsolved problems remain.

K. Ford [9] has improved the result 18 above showing that for every $k > 1$ there exist infinitely many n such that $A(n) = k$.

In [8] Erdős proves that

19. the number of distinct rationals of the form $\sigma(a)/a$, $1 \leq a \leq x$ equals $c_1x + o(x)$, where $\frac{6}{\pi^2} \leq c_1 < 1$.

He claims

16*. for every a with $1 \leq a < \infty$ the number of integers $n \leq x$ such that $\sigma(n)/n = a$ is less than $c_4x^{\frac{1}{2}-c_5}$, where positive constants c_4 and c_5 are independent of a
and deduces from it

17*. the number of solutions of $\sigma(a)/a = \sigma(b)/b$ satisfying $a < b \leq x$ equals $c_2x + o(x)$ for some constant $c_2 > 0$.

He also claims that

18*. the similar results hold for φ .

[8] contains the following problems.

P1. Is it true that $\sigma(n) = \varphi(n)$ has infinitely many solutions?

P2. Let $1 \leq c \leq \infty$. Does there exist an infinite sequence of integers n_k, m_k , where $n_k \neq m_k$ for which $\sigma(n_k) = \sigma(m_k)$ and $m_k/n_k \rightarrow \infty$?

P3. Does the number $g(x)$ of solutions of $\sigma(a) = \sigma(b)$, $(a, b) = 1$, $a < b \leq x$ satisfy $\lim_{n \rightarrow \infty} \frac{g(x)}{x} = \infty$?

Problem P1 has been solved by Ford, Luca and Pomerance [11] (see also [12]), a partial solution of P3 is given in [15].

In [9] Erdős proves that

20. for $k = 2$ and every $\varepsilon > 0$ and all n except a sequence of density 0

$$(5) \quad \varphi_k(n) \equiv 0 \left(\text{mod } \prod_{\substack{p \leq (\log_2 x)^{k-\varepsilon} \\ p \text{ prime}}} p \right),$$

21. for $k = 2$ except for a sequence of density 0

$$(6) \quad \lim_{n \rightarrow \infty} \frac{\varphi_{k-1}(n)}{\varphi_k(n) \log_3 n} = e^\gamma$$

and claims

19*. (5) and (6) hold for all $k > 2$,

20*. for all $k > 2$

$$\sum_{n=i}^x \varphi_k(n) = (1 + o(1)) \frac{3}{\pi^2} e^{-\gamma(k-1)} x^2 / (\log_3 x)^{k-1},$$

21*. analogous results hold for the σ -function.

The claim 20* has been proved by Pollack [21].

The following problems are proposed.

P4. Is it true that the density of numbers n for where there is a number m such that $\varphi(n) = \varphi(m)$ and $(n, m) = 1$ is 0?

P5. Is it true that the density of numbers of the form $n - \varphi(n)$ is 0?

Erdős with coauthors returned to the subject in [21] and corrected the claim 19* by proving that (6) holds with the right-hand side multiplied by $k - 1$. The numerical evidence quoted in [2] seems to indicate that the answer to P5 is negative.

In [11] the following notation is used. $N_\varphi(k, \alpha, x)$ and $N_\sigma(k, \alpha, x)$ denote the number of $n \leq x$ such that $\varphi_k(x) \geq \alpha x$ and $\sigma_k(x) \leq \alpha x$, respectively.

Erdős proves the following theorem.

22. For every $\alpha < \frac{1}{2}$, arbitrarily small $\varepsilon > 0$ and arbitrarily large t we have for $x > x_0(\alpha, t, \varepsilon)$ the inequality

$$\frac{x}{\log x} (\log_2 x)^t < N_\varphi(2, \alpha, x) < \frac{x}{\log x} (\log x)^\varepsilon.$$

Further, for every $\alpha > 0$ and $\varepsilon > 0$ we have

$$N_\varphi(3, \alpha, x) < \frac{x}{(\log x)^2} (\log x)^\varepsilon.$$

He claims the following.

22*. We have for every t , if $x > x_0(t)$

$$N_\sigma(2, 2, x) > \frac{x}{\log x} (\log_2 x)^t$$

and for every $\alpha > 0$ and $\varepsilon > 0$ if $x > x_0(\alpha, \varepsilon)$

$$N_\sigma(2, \alpha, x) < \frac{x}{\log x} (\log x)^\varepsilon,$$

$$N_\sigma(3, \alpha, x) < \frac{x}{(\log x)^2} (\log_2 x)^\varepsilon.$$

Furthermore, the inequality for $N_\sigma(2, 2, x)$ is best possible in the sense that $\alpha = 2$ cannot be replaced by any smaller number.

23*. $N_\varphi(4, \alpha, x) < \frac{cx}{(\log x)^2}.$

The claim that for all $k \geq 2$ and all n except a sequence of density 0

$$\frac{\sigma_k(n)}{\sigma_{k-1}(n)} = (1 + o(1)) \frac{\varphi_{k-1}(n)}{\varphi_k(n)} = (1 + o(1)) e^\gamma k \log_3 n$$

contradicts the theorem and claim made earlier in [9]. As proved in [21] k on the right-hand side should be replaced by $k - 1$.

The following problem is proposed.

P6. To give an asymptotic formula for $N_\varphi(2, \alpha, x)$ or $N_\sigma(2, \alpha, x)$.

H. Maier [18] proved that $N_\sigma(3, \alpha, x) > x(\log x)^{-2}$ ($\alpha > \alpha_0 > 0$, $x > x_0(\alpha)$) and outlined the proof of $N_\sigma(3, \alpha, x) > x(\log x)^{-2}(\log_2 x)^t$ and similarly for $N_\varphi(3, \alpha, x)$.

In [12] Erdős proves the following.

23*. Let $\varepsilon > 0$ be arbitrary,

$$g(n) = \frac{\varphi(n)}{n} \text{ or } \frac{\sigma(n)}{n}, \quad \alpha = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n g(k), \quad \lim_{n \rightarrow \infty} h(n) = \infty.$$

Then the density of numbers x such that for all $t > h(x)$

$$(1 - \varepsilon)\alpha < \frac{1}{t} \sum_{i=1}^t g(x + i) < (1 + \varepsilon)\alpha$$

is 1.

In [13] the following notation is used. Let $R(n)$ be the least integer k such that $\varphi_k(n) = 1$, $T(n) = \varphi_1(n) + \varphi_2(n) + \dots + \varphi_{R(n)}(n)$, $F(x, c)$ be the number of integers $n \leq x$ for which $T(n) > cn$.

The authors claim

24*. $T(n) = (1 + o(1))\varphi(n)$ on a sequence of density 1,

25*. for every $c \in (1, \frac{3}{2})$, we have for every $t \geq 0$ and every $\varepsilon > 0$ if $x > x_0(c, t, \varepsilon)$

$$\frac{x}{\log x} (\log_2 x)^t < F(x, c) < \frac{x}{(\log x)^{1-\varepsilon}},$$

26*. $F(x, 1) = (c + o(1)) \frac{x}{\log_4 x}$,

27*. $T(n) > \frac{3}{2}n$ for infinitely many n ,

28*. for $c > \frac{3}{2}$ and every $\varepsilon > 0$

$$F(x, c) = o\left(\frac{x}{(\log x)^{2-\varepsilon}}\right).$$

They conjecture:

C10. for $1 < c_1 < c_2 < \frac{3}{2}$

$$\lim_{x \rightarrow \infty} F(x, c_1)/F(x, c_2) = \infty,$$

C11. $F\left(x, \frac{3}{2}\right) = o\left(\frac{x}{\log x}\right)$.

They propose the following problems.

P7. Does $\frac{R(n)}{\log n}$ has a distribution function?

P8. Does $\frac{R(n)}{\log n}$ approaches a limit for almost all n ? If the limit exists, is it equal to $\frac{1}{\log 2}$ or $\frac{1}{\log 3}$?

They repeat the claim made earlier in [11] that for all $k \geq 2$ and all n except a sequence of density 0

$$\frac{\sigma_k(n)}{\sigma_{k-1}(n)} = (1 + o(1))e^\gamma k \log_3 n,$$

where by analogy with the φ -function treated in [21] the factor k on the right-hand side should probably be replaced by $k - 1$. [21] also addresses P8.

The claim 26* has been proved by Loomis and Luca [14], who showed that $c = e^{-\gamma}$.

In [14] Erdős proves the following theorem.

24. The lower density of numbers m , for which $\sigma(n) - n = m$ is insolvable, is positive.

This is deduced from the following stronger, but more complicated theorem.

25. Let P_k denote the product of the first k primes. The number $A(k, x)$ of numbers n such that $\sigma(n) - n \leq x$ and $\sigma(n) - n \equiv 0 \pmod{P_k}$ is less than $\varepsilon x / P_k$ for every $\varepsilon > 0$ and $x > x_0(\varepsilon)$.

The question open here, whether there are infinitely many numbers not of the form $n - \varphi(n)$ has been solved only much later in [2], see also [7].

In [15] Erdős outlines the proof of the following theorem.

26. Let $h(x)$ be the number of solutions of the conditions $\sigma(a) = \sigma(b)$, $(a, b) = 1$, $a < b \leq x$. Then

$$\limsup_{x \rightarrow \infty} \frac{h(x)}{x} = \infty.$$

He claims

$$29^*. \quad \lim_{x \rightarrow \infty} \frac{h(x)}{x} = \infty$$

and

30*. for almost all n

$$\nu(\sigma(n)) = \left(\frac{1}{2} + o(1) \right) (\log_2 n)^2.$$

The last claim corresponds to the result on φ given later in [46].

Erdős mentions several problems.

P9. Estimate the number of integers $n < x$ for which $\varphi(m) = n$ is solvable in integers $m > x$.

P10. Is density of numbers of the form $n + \varphi(n)$ or $n + \sigma(n)$ positive?

P11. Is there a $\beta > 1$ for which

$$|\sigma(n) - \beta n| \rightarrow \infty \quad \text{as } n \rightarrow \infty?$$

In [16] Erdős proves the following theorem.

27. Let $F(x, a, b)$ be the number of integers $n \leq x$ satisfying $\alpha < \frac{\sigma(n)}{n} < b$. There is a constant c_1 such that for $t > 0$, $x > t$

$$F\left(x, a, a + \frac{1}{t}\right) < c_1 \frac{x}{\log t}.$$

Apart from the value of c_1 the inequality is best possible.

Erdős's proof of 27 is reproduced with commentaries in [6].

He claims the following.

31*. For $t > 0, x > t$

$$F\left(x, a, a\left(1 + \frac{1}{t}\right)\right) < c_2 \frac{x}{\log t}.$$

32*. The same result is true for φ .

Prof. I. Z. Ruzsa remarked that the claim 31* easily follows from the result 27. Indeed, if $t \geq a^2$, then $\log t$ and $\log(t/a)$ are of the same order of magnitude. If $t < a^2$, we can estimate the quantity in question by $F(x, a, \infty) = O(x/a)$, by a first-moment estimate.

In [17] the authors introduce the following notation: for each real valued arithmetic function f satisfying $\lim_{n \rightarrow \infty} f(n) = \infty$, let

$$F_f(n) = \#\{j < n : f(j) \geq f(n)\} + \#\{j > n : f(j) \leq f(n)\}.$$

They prove the following.

- 28. $F_\varphi(n)/n = h\left(\frac{\varphi(n)}{n}\right) + O(\exp(-\sqrt{\log n}))$, where h is a certain strictly convex function with the minimum $h_0 = h(u_0)$.
- 29. $F_\varphi(n)/n$ has a continuous distribution function.

30.
$$\max_{n < x} F_\varphi(n) = x - \left(\frac{\zeta(2)\zeta(3)}{\zeta(6)} e^{-\gamma} + o(1)\right) x / \log_2 x.$$

- 31. If $n_1 = 1$ and n_{k+1} is the least number $> n_k$ such that $F_\varphi(x) < F_\varphi(n_{k+1})$ for all $x < n_{k+1}$, then $n_{k+1} - n_k > n_k^{1-\varepsilon}$ for every $\varepsilon > 0$ and $k \rightarrow \infty$.
- 32. $n_0 > 0.472$ and $h_0 < 0.324$.
- 33. $n_0 < 0.475$ and $h_0 > 0.321$.
- 34. $\min_{n > x} F(x) \sim h_0 x$.

A result concerning general functions f immediately implies

35.
$$\#\{j < n : \sigma(j) \geq \sigma(n)\} = \sigma(n) \int_{\frac{\sigma(n)}{n}}^{\infty} (1 - D_\sigma(t)) t^{-2} dt + o(n),$$

$$\#\{j > n : \sigma(j) \leq \sigma(n)\} = \sigma(n) \int_0^{\frac{\sigma(n)}{n}} D_\sigma(t) t^{-2} dt + o(\sigma(n)),$$

where

$$D_\sigma(t) = \lim_{x \rightarrow \infty} \frac{1}{x} \left\{ n \leq x : \frac{\sigma(n)}{n} \leq t \right\}.$$

In [18] the authors prove the following theorems.

36. The system of inequalities

$$\begin{aligned} \varphi(n+1) &\geq \varphi(n+2) \geq \varphi(n+3) \geq \varphi(n+4) \geq \varphi(n+5), \\ \sigma(n+1) &\geq \sigma(n+2) \geq \sigma(n+3) \geq \sigma(n+4) \geq \sigma(n+5), \end{aligned}$$

has no solutions in positive integers n .

37. Let $\xi_1, \dots, \xi_k, \eta_1, \dots, \eta_k$ be positive real numbers. A sequence n_l ($l = 1, 2, \dots$) such that

$$\lim_{l \rightarrow \infty} \frac{\varphi(n_l + i)}{\varphi(n_l + i + 1)} = \xi_i, \quad \lim_{l \rightarrow \infty} \frac{\sigma(n_l + i)}{\sigma(n_l + i + 1)} = \eta_i, \quad (i = 1, \dots, k)$$

exists if and only if there exists a sequence n_l ($l = 1, 2, \dots$) such that

$$\lim_{l \rightarrow \infty} \frac{h^*(n_l + i)}{h^*(n_l + i + 1)} = \xi_i \eta_i,$$

where $h^*(n) = \sigma(n)\varphi(n)/n^2$.

38. Let i_1, i_2, i_3, i_4 and j_1, j_2, j_3, j_4 be any permutation of 1, 2, 3, 4. Then there exist infinitely many positive integers n such that

$$\begin{aligned} \varphi(n + i_1) &> \varphi(n + i_2) > \varphi(n + i_3) > \varphi(n + i_4), \\ \sigma(n + i_1) &> \sigma(n + i_2) > \sigma(n + i_3) > \sigma(n + i_4). \end{aligned}$$

The result 37 implies a theorem of the writer [26].

In [19] Erdős claims that

33*. the density of integers n such that $d(n) < d(n+1)$ and $\varphi(n) < \varphi(n+1)$ is $\frac{1}{4}$,

34*. the density of integers n such that $\varphi(n) < \varphi(n+1)$ and $\sigma(n) < \sigma(n+1)$ is strictly between $\frac{1}{4}$ and $\frac{1}{2}$.

In [20] the authors prove that

39. for large n , the number of solutions of $\varphi(n) = \varphi(n+1)$ not exceeding x is at most $x/\exp((\log x)^{1/3})$

and claim

35*. this is also true for $\sigma(n)$.

They conjecture:

C12. for every $\varepsilon > 0$ and $x > x_0(\varepsilon)$ the equations $\varphi(n) = \varphi(n+1)$, $\sigma(n) = \sigma(n+1)$ each have at least $x^{1-\varepsilon}$ solutions $n \leq x$.

The equation $\varphi(n) = \varphi(n+k)$ has been later considered in [13].

In [21] the authors denote by $k(n)$ the least k such that $\varphi_k(n) = 1$. They put $\phi(n) = n \prod_{k \geq 1} \varphi_k(n)$ and prove the following,

40. A version of the Elliott–Halberstam conjecture on primes in arithmetic progressions implies that $k(n)$ has a normal order $\alpha \log n$ for a certain constant $\alpha > 0$.
41. Let $\varepsilon(x) > 0$ tend to 0 arbitrarily slowly as $x \rightarrow \infty$. If $k \leq (\log_2 n)^{\varepsilon(x)}$, then the normal order of $\varphi_k(n)/\varphi_{k+1}(n)$ for $n \leq x$ is $ke^\gamma \log_3 x$.
42. For each $\varepsilon > 0$ the set of n with

$$\frac{s_k(n)}{s(n)} < \frac{s(n)}{n} + \varepsilon$$

has density 1.

43. Let $S_k(x)$ denote the number of odd numbers $m \leq x$ in the range of the function s_k . There is a positive number δ_0 such that

$$S_k(x) \ll x^{1-\delta_0}$$

uniformly for all natural numbers k and $x > 0$.

They conjecture the following.

- C13. For each prime p , let $N(x, p)$ denote the number of $n < x$ with $p|\varphi(n)$. Then for every $\varepsilon > 0$, $N(x, p) = o(x)$ uniformly in the region $p > (\log x)^{1+\varepsilon}$ and $N(x, p) \sim x$ uniformly in the region $p < (\log x)^{1+\varepsilon}$.
- C14. For each $\varepsilon > 0$, the upper density of the set of n with the property that the largest prime factor of $\varphi_k(n)$ exceeds n^ε tends to 0 as $k \rightarrow \infty$.
- C15. For each $\varepsilon > 0$ and k , the set of n with

$$\frac{s_{j+1}(n)}{s_j(n)} < \frac{s(n)}{n} + \varepsilon \quad \text{for } j = 1, \dots, k$$

has density 1.

- C16. If A is a set of natural numbers of positive upper density, then $s(A) = \{s(n) : n \in A\}$ also has positive upper density.

Finally a mistake committed in [50] is corrected.

II. Papers concerning the σ -function, but not the φ -function. Here belong papers:

- [22] On the density of the abundant numbers, *J. London Math. Soc.* 9 (1934), 278–280.
- [23] On primitive abundant numbers, *J. London Math. Soc.* 10 (1935), 49–58.
- [24] Note on consecutive abundant numbers, *J. London Math. Soc.* 10 (1935), 128–131.
- [25] On amicable numbers, *Publ. Math. Debrecen* 4 (1955), 108–111.
- [26] Remarks on number theory, I. On primitive α -abundant numbers, *Acta Arith.* 5 (1958), 25–33.
- [27] Asymptotic formulas for some arithmetical functions, *Canad. Math. Bull.* 1 (1958), 149–153.
- [28] On the sum $\sum_{d|2^n-1} d^{-1}$, *Israel J. Math.* 9 (1971), 43–48.
- [29] On abundant-like numbers, *Canad. Math. Bull.* 17 (1974), 599–602.
- [30] (with S. J. Benkoski) On weird and pseudoperfect numbers, *Math. Comp.* 28 (1974), 617–623.
- [31] (with G. J. Rieger) Ein Nachtrag über befreundete Zahlen, *J. Reine Angew. Math.* 273 (1975), 220.
- [32] (with J.-L. Nicolas) Répartition des nombres superabondants, *Sém. Delange-Pisot-Poitou 1973/74, Théorie des nombres, Fasc. 1, Exp. No. 5*, 18 pp., Secrétariat Math., Paris, 1975.
- [33] (with J.-L. Nicolas) Répartition des nombres superabondants, *Bull. Soc. Math. France* 103 (1975), 65–90.
- [34] On asymptotic properties of aliquot sequences, *Math. Comput.* 30 (1976), 641–645.
- [35] Sur la fonction “nombre de facteurs premiers de n ”, *Sém. Delange-Pisot-Poitou, 1978/1979. Théorie des nombres, Fasc. 2, Exp. No. 32*, 19 pp., Secrétariat Math., Paris, 1980.
- [36] Sur la fonction: nombre de facteurs premiers de N , *Enseign. Math* (2) 27 (1981), 3–27.
- [37] (with P. T. Bateman, C. Pomerance, E. G. Straus) The arithmetic mean of the divisors of an integer, *Analytic Number Theory (Proc. Conf., Temple Univ., Phila., 1980)*, *Lecture Notes in Math.* 899, pp. 197–220, Springer, Berlin–New York, 1981.
- [38] (with A. Sárközy) On isolated, respectively consecutive large values of arithmetic functions, *Acta Arith.* 66 (1994), 269–295.
- [39] Some of my favorite problems and results, *The Mathematics of Paul Erdős* (R. Graham and J. Nešetřil, eds.), 47–67, Springer 1997.

An α -abundant number is according to Erdős a positive integer n such that $\sigma(n) \geq \alpha n$, an abundant number is 2-abundant (usually, a positive integer is called abundant if $\sigma(n) > 2n$). A primitive α -abundant number is an α -abundant number no proper divisor of which is α -abundant. Let $N_\alpha(x)$ be the number of primitive α -abundant numbers $\leq x$,

[22] contains an elementary proof of Davenport's theorem [3] that abundant numbers have a density d . d has been estimated by Deléglise [4], who has proved: $d = 0.247\dots$

[23] contains a proof that

44. for x large enough

$$\frac{x}{\exp(8(\log x \log_2 x)^{1/2})} < N_2(x) < \frac{x}{\exp(\frac{1}{25}(\log x \log_2 x)^{1/2})}.$$

The constants 8 and $\frac{1}{25}$ have been improved by Ivić [15]. In **[39]** Erdős proposes the following problem.

P12. Does there exist a number c such that

$$N_2(x) = \frac{x}{\exp((c + o(1))(\log x \log_2 x)^{1/2})}?$$

[24] contains a proof of the following theorem.

45. There exist two positive constants c_1, c_2 such that for all sufficiently large n , there exist $c_1 \log_3 n$ consecutive integers all abundant and less than n , but not $c_2 \log_3 n$ consecutive integers all abundant and less than n .

It is also claimed that

36*. for every $\varepsilon > 0$ a constant $c(\varepsilon)$ exists such that $n > n_0(\varepsilon)$, then among $c(\varepsilon) \log_3 n$ consecutive integers there is at least one, say m , such that $\sigma(m)/m < 1 + \varepsilon$;

37*. if $\lim_{n \rightarrow \infty} \frac{f(n)}{\log_3 n} = \infty$, then the abundant numbers have the same density in the interval $(n, n + f(n))$ as in the interval $(1, n)$.

The result 45. has been improved by Pollack [20], who proved an asymptotic formula for the maximal length of a sequence of consecutive abundant numbers less than n .

[25] contains an estimate for the number $B(x)$ of pairs (a, b) called amicable, where positive integers a, b satisfy $a < \min\{b, x\}$ and $\sigma(a) = \sigma(b) = a + b$, namely

46. $B(x) = o(x)$.

This estimate is improved in **[31]** to the form

47. $B(x) = O(x/\log_3 x)$.

This result has been improved by Pomerance [23] to the form

$$B(x) = O(x \exp(-(\log x)^{1/3})).$$

Pomerance says that he has been inspired by **[23]**.

[25] contains besides the following claim.

38*. If $b = \sigma(a) - a$, then except for a sequence of density 0

$$\frac{\sigma(b)}{b} = \frac{\sigma(a)}{a} + O(1).$$

[26] contains the following result

48. $N_\alpha(x) = o\left(\frac{x}{\log x}\right).$

In [27] Erdős claims the following.

39*. For $0 < \alpha < \frac{1}{2}$

$$(7) \quad \sum_{n=1}^x \sigma((n, \lfloor n^\alpha \rfloor)) = (1 + o(1))x \log x.$$

40*. Necessary and sufficient condition that for a real $\alpha > 0$ we should have

$$\sum_{n=i}^x \sigma((n, \lfloor \alpha n \rfloor)) = \left(\frac{1}{2} + o(1)\right) x \log x$$

is that for every $\varepsilon > 0$ the number of solutions in positive integers a and b of $\left|\alpha - \frac{a}{b}\right| < \frac{1}{b^{\alpha+\varepsilon}}$ and of $\alpha < \frac{a}{b} < \alpha + \frac{\varepsilon}{b^2 \log b}$ is finite.

41*. The condition 40* is equivalent to the following condition on the continued fraction development of

$$\alpha = a_0 + \frac{1}{|a_1|} + \dots$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log a_n = 0, \quad \lim_{n \rightarrow \infty} \frac{1}{n} a_{2n+1} = 0.$$

He conjectures that

C17. (7) holds for all α in the interval $(\frac{1}{2}, 1)$.

The value $\alpha = 1/2$ is missing from the range given in 39* and C17. This is likely intentional, since, as Prof. I. Z. Ruzsa remarked, for $\alpha = \frac{1}{2}$ one gets elementarily the asymptotic $(1/2)x \log x$.

[28] contains the following theorem.

49. There exists a constant c_1 such that for every n

$$\frac{\sigma(2^n - 1)}{2^n - 1} < c_1 \log_2 n.$$

The proof works with 2 replaced by any integer $a > 1$. It is also claimed that

$$42^* . \quad \sum_{\substack{d|2^n-1 \\ d>n}} \frac{1}{d} \rightarrow 0,$$

43*. $\frac{\sigma(2^n - 1)}{2^n - 1}$ has a distribution function.

Finally the following problem is proposed.

P13. Is it true that for a certain constant c and every n

$$\sum_{d|2^n-3} \frac{1}{d} < c \log_2 n?$$

In [29] the following notation is used. Let $n_c(p)$ be the least number n divisible by p , but by no smaller prime that has $\sigma(n) \geq cn$. Erdős proves and claims the following.

50. For $c \geq 2$ the number $n_c(p)$ is cube-free, but not square-free apart from finitely many exceptions.

51. Let A, B be the set of c for which, with a certain p , $n_c(p)$ is or, respectively, is not square-free. A is everywhere dense in $(1, 2)$.

44*. B and $A \cap B$ are everywhere dense in $(1, 2)$.

He conjectures that

C18. for $1 < c < 2$ the result corresponding to 49 is false.

In [30] the following notions are used. A positive integer n is called pseudoperfect if it is the sum of some of its proper divisors. If n is abundant but not pseudoperfect, then it is called weird. The authors prove

52. the lower density of weird numbers is positive

and propose the following problems.

P14. Are there odd weird numbers?

P15. Can $\sigma(n)/n$ be arbitrarily large for weird n ?

In [32] and [33] the same notion of superabundant numbers is used as in [3] (p. 587). Let $Q(x)$ be the number of superabundant numbers $\leq x$. It is shown that

$$53. \quad \liminf_{x \rightarrow \infty} (\log Q(x) / \log_2 x) \geq 5.48.$$

In [34] the following theorem is proved.

54. For every k and $\delta > 0$ and for all n except a sequence of density 0

$$(1 - \delta)n \left(\frac{s(n)}{n} \right)^i < s_i(n), \quad 1 \leq i \leq k.$$

It is also claimed that

45*. for every k and $\delta > 0$ and for all n except a sequence of density 0

$$s_i(n) < (1 + \delta)n \left(\frac{s(n)}{n} \right)^i, \quad 1 \leq i \leq k.$$

The claim is withdrawn in [21].

In [35] and [36] the authors prove that

55.
$$\max_{n \leq x} (\sigma(n - 1) + \sigma(n)) \leq (1 + o(1))e^\gamma x \log_2 x.$$

In [37] the following notation is used. If $n = \prod_{p \text{ prime}} p^{e(p)}$, then

$$\langle n^\alpha \rangle = \prod_{p \text{ prime}} p^{\lfloor \alpha e(p) \rfloor}.$$

The authors prove the following theorems.

56. Let $N(x)$ denote the number of $n \leq x$ for which $\sigma(n)/d(n)$ is not an integer. Then

$$N(x) = x \exp(-(1 + o(1))2\sqrt{\log 2} \sqrt{\log_2 x}).$$

57. For every ε in $(0, 1)$ the set of n for which $\langle d(n)^{2-\varepsilon} \rangle | \sigma(n)$ has density 1, the set of n for which $\langle d(n)^{2+\varepsilon} \rangle | \sigma(n)$ has density 0, the set of n for which $d(n)^2 | \sigma(n)$ has density 1/2.

58. As $x \rightarrow \infty$

$$\sum_{n \leq x} \frac{\sigma(n)}{d(n)} = \frac{g(1)}{2\pi^{1/2}} \frac{x^2}{(\log x)^{1/2}},$$

where

$$g(s) = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s} \right)^{1/2} \left(1 + \frac{1}{2} \left(1 + \frac{1}{p} \right) p^{-s} + \frac{1}{3} \left(1 + \frac{1}{p} + \frac{1}{p^2} \right) p^{-2s} + \dots \right)$$

and the principal value of $(1 - \frac{1}{p^s})^{1/2}$ is taken.

59. As $n \rightarrow \infty$ we have

$$A \left\{ n : \frac{\sigma(n)}{d(n)} \leq x \right\} \sim Ax \log x,$$

where

$$A = \prod_{p \text{ prime}} \left(1 - \frac{1}{p}\right)^2 \left(1 + \frac{2}{p+1} + \frac{3}{p^2+p+1} + \dots\right).$$

In [38] the authors denote for a given arithmetic function f by $G(f, x)$ the greatest integer G for which there is a positive integer n with

$$f(n) > \sum_{0 < |i| < G} f(n+i)$$

and put

$$M(f, x) = \max_{n \leq x} f(n), \quad T(f, x) = \max_{n \leq x} (f(n-1) + f(n)).$$

They prove

$$60. \quad G(\sigma, x) = (1 + o(1))3e^\gamma \log_2 x,$$

$$61. \quad T(\sigma, x) \leq x \left(M\left(\frac{\sigma(n)}{n}, x\right) + 1 + O((\log_2 x)^{-1}) \right)$$

61 is an improvement on 55.

III. Papers concerning the φ -function, but not the σ -function.

Here belong the following papers

[40] On the normal number of prime factors of $p-1$ and some related problems concerning Euler's φ -function, *Quart. J. Math., Oxford Ser.* 6 (1935), 205–213.

[41] On the integers which are the totient of a product of three primes, *ibid.* 7 (1936), 16–19.

[42] On the integers which are the totient of a product of two primes, *ibid.*, 227–229.

[43] Some remarks on Euler's ϕ -function and some related problems, *Bull. Amer. Math. Soc.* 51 (1945), 540–544.

[44] Some asymptotic formulas in number theory, *J. Indian Math. Soc. (N.S.)* 12 (1948), 75–78.

- [45] (with H. N. Shapiro) The existence of a distribution function for an error term related to the Euler function, *Canad. J. Math.* 7 (1955), 63–76.
- [46] (with R. R. Hall) On the values of Euler's φ -function, *Acta Arith.* 22 (1973), 201–206.
- [47] (with R. R. Hall) Distinct values of Euler's φ -function, *Mathematika* 23 (1976), 1–3.
- [48] (with R. R. Hall) Euler's φ -function and its iterates, *ibid.* 24 (1977), 173–177.
- [49] Some problems and results in number theory, *Number Theory and Combinatorics, Japan 1984*, 65–87, Reidel 1985.
- [50] (with C. Pomerance) On the normal number of prime factors of $\varphi(n)$, *Rocky Mountain J. Math.* 15 (1985), 343–352.
- [51] (with F. Luca and C. Pomerance) On the proportion of numbers coprime to a given integer. De Koninck, Jean-Marie (ed.) et al., *Anatomy of integers*, Montreal, Canada, 2006. Providence, RI: American Mathematical Society (AMS), CRM Proc. and Lecture Notes 46 (2008), 47–64.

In [40] Erdős considers the number $V(x)$ of positive integers $m \leq x$ for which the equation $\varphi(n) = m$ is solvable and the number $A(m)$ defined above as the number of solutions of $\varphi(n) = m$. He proves that

62. for every $\varepsilon > 0$ and $x > x_0(\varepsilon)$

$$V(x) < \frac{x}{(\log x)^{1-\varepsilon}},$$

63. there exists a constant $c > 0$ such that for infinitely many n

$$A(n) \geq n^c.$$

Both results improve upon an earlier work of Pillai [19]. Wooldridge [28] gave later an admissible value of the constant c , as any number $< 3 - 2\sqrt{2}$.

Let $f_3(m)$, $f_2(m)$ be the number of representation of m as $\varphi(pqr)$, or $\varphi(pq)$, respectively, where p , q , r are distinct primes.

In [4] Erdős proves elementarily that

64.
$$\limsup_{m \rightarrow \infty} f_3(m) = \infty.$$

In [42] Erdős proves by a non-elementary method that

65. for infinitely many m , $f_2(m) > \exp(c\sqrt{\log m})$, where $c > 0$ is a constant. Almost the same argument applies to the equations $\sigma(pqr) = m$ and $\sigma(pq) = m$, but the fact is not mentioned in the papers.

In [43] Erdős proves that

66. $V(x) > \frac{cx}{\log x} \log_3 x$, c a positive constant, a result claimed in [40]. The claim that

$$V(x) > \frac{x}{\log x} (\log_3 x)^k \quad \text{for every } k \text{ and } x > x_0(k)$$

is substantiated in [47].

In [43] Erdős also proves a theorem due to him and P. Turán that

67. the number of integers n for which $\varphi(n) \leq x$ is $cx + o(x)$. The constant c has been determined by Dressler [5] as $\frac{\zeta(2)\zeta(3)}{\zeta(6)}$ (see also Bateman [1]).

In [44] Erdős proves the following theorem.

68. Denote by $A(n)$ the number of integers $m \leq n$ for which $(m, \varphi(m)) = 1$. Then

$$A(n) = (1 + o(1)) \frac{ne^{-\gamma}}{\log_3 n}.$$

He claims

- 46*. for every $\varepsilon > 0$ for which the inequality

$$(1 - \varepsilon) \log_4 m < \nu((m, \varphi(m))) < (1 + \varepsilon) \log_4 m$$

is not satisfied is $o(n)$. This claim is substantiated in [51].

- In [45] $H(x) = \sum_{n \leq x} \frac{\varphi(n)}{n} - \frac{6}{\pi^2} x$. The authors prove that

69. $H(x)$ has a continuous distribution function.

In [46] the authors prove that

70. for every $B > 2\sqrt{2}$

$$V(x) = O\left(\pi(x) \exp(B\sqrt{\log_2 x})\right).$$

In [47] the authors prove that

71. there exist positive constants A, C such that

$$V(x) \geq C\pi(x) \exp(A(\log_3 x)^2)$$

and ask the following question:

P16. is it true that for any $c > 1$

$$\lim_{x \rightarrow \infty} V(cx)/V(x) = c?$$

The results 62, 66, 70 and 71 have been superseded by the following theorem of Ford [8]:

$$V(x) = \frac{x}{\log x} \exp(C(\log_3 x - \log_4 x)^2 + D \log_3 x - (D + 1/2 - C) \log_4 x + O(1))$$

where C, D are constants.

This is, however, not strong enough to answer P16.

In [48]

$$V_r(x) = \#\{m \leq x : m = \varphi_r(n) \text{ for some } n\}.$$

The authors prove that

72. $V_2(x) \ll (x/(\log x)^2) \exp(D \log_2 x \log_4 x / \log_3 x)$, D a constant and conjecture that

C19. the relation

$$x/(\log x)^{r+\varepsilon} \ll V_r(x) \ll x/(\log x)^{r-\varepsilon}$$

holds for every fixed r and every $\varepsilon > 0$.

The right-hand side of this inequality for every r and its left-hand side for $r = 2$ have been proved by Luca and Pomerance [17].

In [49] Erdős makes the following conjectures.

C20. For every $k : \varphi(n) = \varphi(n + 1) = \dots = \varphi(n + k)$ has infinitely many solutions.

C21. For every $e > 0$ and $x > x_0(e)$ there are at least $(\log x)^e$ consecutive integers not exceeding x for which all the values $\varphi(n + i)$, $1 \leq i < (\log x)^e$ are distinct.

C22. If k_n is the maximal number k such that $\varphi(n + i)$, $1 \leq i \leq k_n$ are all distinct, then $k_n = O(n^\varepsilon)$ for every $\varepsilon > 0$.

He claims

47*. $k_n \ll n \exp(-(\log n)^{1/3})$.

The claim concerning $\varphi(n) = \varphi(n + 1)$ is substantiated in [20] (see 39).

In [50] the authors consider the function $\Omega(f(n))$, where $\Omega(n)$ is the total number of prime factors of n and $f(n) = \varphi(n)$ or $\lambda(n)$, the universal exponent mod n . They prove that the following theorems.

73. $\lim_{x \rightarrow \infty} \frac{1}{x} \#\left\{n \leq x : \Omega(f(n)) - \frac{1}{2}(\log_2 x)^2 \leq \frac{u}{\sqrt{3}}(\log_2 x)^{3/2}\right\} = G(u)$

where $G(u) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^u e^{-t^2/2} dt$.

$$74. \quad \lim_{x \rightarrow \infty} \frac{1}{x} \# \left\{ n \leq x : \omega(f(n)) - \frac{1}{2}(\log_2 x)^2 \leq \frac{u}{\sqrt{3}}(\log_2 x)^{3/2} \right\} = G(u).$$

They conjecture

C23.

$$\lim_{x \rightarrow \infty} \frac{1}{x} \# \left\{ n \leq x : \Omega(l_a(n)) - \frac{1}{2}(\log_2 x)^2 \leq \frac{u}{\sqrt{3}}(\log_2 x)^{3/2} \right\} = G(u) \frac{\varphi(a)}{a},$$

C24.

$$\lim_{x \rightarrow \infty} \frac{1}{x} \# \left\{ n \leq x : \omega(l_a(n)) - \frac{1}{2}(\log_2 x)^2 \leq \frac{u}{\sqrt{3}}(\log_2 x)^{3/2} \right\} = G(u) \frac{\varphi(a)}{a}.$$

where $l_a(n)$ is the exponent to which a belongs mod n ($a \neq 0, \pm 1, (a, n) = 1$).

An error committed in the proof of one of the lemmas is corrected in [21].

In [51] the authors consider the functions

$$f(a) = \#\{b : (a, b) = 1 \text{ and } a/b = \varphi(n)/n \text{ for some } n\},$$

$$g(b) = \#\{a : (a, b) = 1 \text{ and } a/b = \varphi(n)/n \text{ for some } n\}$$

and prove.

75. The inequality $f(a) \leq (1 + o(1))a \log_2 a / \log_3 a$ holds as $a \rightarrow \infty$. On the other hand, there exists a positive constant c_0 such that $f(a) > a^{c_0}$ for infinitely many a .

76. We have $f(a) = 0$ for almost all positive integers a .

77. We have $g(b) \leq b^{(1+o(1)) \log_3 b / \log_2 b}$ as $b \rightarrow \infty$.

78. The inequality

$$(n, \varphi(n)) \leq 2n \exp(-\sqrt{\log 2 \log n})$$

holds for all square-free $n \geq 1$. On the other hand, there is an infinite set S of square-free numbers n such that

$$(n, \varphi(n)) > n^{1-(1+o(1)) \log_3 n / \log_2 n} \quad \text{as } n \rightarrow \infty, n \in S.$$

79. For almost all n , $(n, \varphi(n))$ is the largest divisor of n supported on the prime divisors of n in the interval $[1, \log_2 n]$.

80. Let $A(x) = \frac{1}{x} \sum_{n \leq x} (n, \varphi(n))$. Then for any $k > 0$ we have

$$(\log x)^k \leq A(x) \leq x^{(1+o(1)) \log_3 x / \log_2 x} \quad \text{as } x \rightarrow \infty.$$

Thanks are due to Prof. P. Pollack for supplying several references and Prof. I. Z. Ruzsa for his penetrating report.

Note added in proof.

Ivić's results [15] have been improved by M. R. Avidon, *Acta Arith.* 77 (1996), 195–205.

REFERENCES

- [1] P. T. Bateman, *The distribution of values of the Euler function*, *Acta Arith.* 21 (1972), 329–345.
- [2] J. Browkin and A. Schinzel, *On integers not of the form $n - \varphi(n)$* , *Colloq. Math.* 68 (1995), 55–58.
- [3] H. Davenport, *Über numeri abundantes*, *Sber. Preuß. Akad. Wiss. Berlin*, 27 (1933), 830–837, also in the *Collected works*, vol. 4, 1834–1841.
- [4] M. Deléglise, *Bounds for the density of abundant integers*, *Exp. Math.* 7 (1998), 137–143.
- [5] R. E. Dressler, *A density which counts multiplicity*, *Pacific J. Math.* 34 (1970), 371–378.
- [6] P. D. T. A. Elliott, *Probabilistic number theory, I Mean-value theorems*, *Grundlehren der Mathematischen Wissenschaften* 239, Springer Verlag 1979.
- [7] A. Flammenkamp and F. Luca, *Infinite families of non-cototients*, *Colloq. Math.* 86 (2000), 37–41.
- [8] K. Ford, *The distribution of totients*, *Ramanujan J.* 2 (1998), 67–151.
- [9] —, *The number of solutions of $\varphi(x) = m$* , *Ann. of Math.* 150 (1999), 283–311.
- [10] —, *An explicit sieve bound and small values of $\sigma(\varphi(m))$* , *Period. Mat. Hungar.* 43 (2011), 15–29.
- [11] K. Ford, F. Luca and C. Pomerance, *Common values of the arithmetic functions φ and σ* , *Bull. London Math. Soc.* 42 (2010), 478–488.
- [12] K. Ford and P. Pollack, *On common values of $\phi(n)$ and $\sigma(n)$, I*, *Acta Math. Hungar.* 133 (2011), 251–271.
- [13] S. W. Graham, J. J. Holt and C. Pomerance, *On the solutions to $\phi(n) = \phi(n + k)$* , *Number theory in Progress*, vol. 2, 867–882, Walter de Gruyter 1999.
- [14] P. Loomis and F. Luca, *On totient abundant numbers*, *Integers. Electronic J. of Combinatorial Number Theory* 8 (2008), #A06.
- [15] A. Ivić, *The distribution of positive abundant numbers*, *Studia Sc. Math. Hungar.* 20 (1985), 183–187.
- [16] F. Luca and C. Pomerance, *On some problems of Mąkowski–Schinzel and Erdős concerning the arithmetical functions φ and σ* , *Colloq. Math.* 92 (2002), 111–130; *Acknowledgement of priority*, *ibid.* 126 (2012), 139.
- [17] —, —, *On the range of the iterated Euler function*, *Combinatorial number theory*, 101–106, Walter de Gruyter, Berlin, 2009.

- [18] H. Maier, *On the third iterates of the φ - and σ -function*, Colloq. Math. 49 (1984), 123–130.
- [19] S. S. Pillai, *On some functions connected with $\phi(n)$* , Bull. Amer. Math. Soc. 35 (1929), 832–836.
- [20] P. Pollack, *Long gaps between deficient numbers*, Acta Arith. 146 (2011), 33–43.
- [21] —, *Two remarks on iterates of Euler's totient function*, Arch. Math. 97 (2011), 449–453.
- [22] —, *On the greatest common divisor of a number and its sum of divisors*, Michigan Math. J. 60 (2011), 199–214.
- [23] C. Pomerance, *On the distribution of amicable numbers*, J. Reine Angew. Math. 293/294 (1977), 217–222.
- [24] —, *On the composition of the arithmetic functions σ and φ* , Colloq. Math. 58 (1989), 11–15.
- [25] P. Poulet, *Nouvelles suites arithmétiques*, Sphinx 2 (1932), 53–54.
- [26] A. Schinzel, *On functions $\varphi(n)$ and $\sigma(n)$* , Bull. Acad. Polon. Sci. Cl. III, 3 (1955), 415–419.
- [27] E. Wirsing, *Bemerkung zu der Arbeit über vollkommene Zahlen*, Math. Ann. 137 (1959), 316–318.
- [28] K. Wooldridge, *Values taken many times by Euler's phi-function*, Proc. Amer. Math. Soc. 76 (1979), 229–234.

Andrzej Schinzel

*Institute of Mathematics,
Polish Academy of Sciences,
Śniadeckich 8,
00-956 Warsaw,
Poland*

e-mail: schinzel@impan.pl

SOME RESULTS AND PROBLEMS IN THE THEORY OF WORD MAPS

ANER SHALEV

In recent years there has been much interest in word maps on groups, with various motivations and applications. Substantial progress has been made and many fundamental questions were solved, using a wide spectrum of tools, including representation theory, probability and geometry. This paper is an extended survey of the various developments in this field. We also suggest remaining open problems, conjectures and possible directions for further research.

Contents:

1. Introduction
2. Waring type problems
3. Surjective words and Ore Conjecture
4. Probabilistic aspects
5. Image size and fiber size
6. Conjugacy classes and Thompson Conjecture
7. Character methods
8. Infinite groups

1. INTRODUCTION

In recent years there has been a wealth of new results and methods in the study of word maps on groups, their properties and applications. By a *word* we mean an element $w = w(x_1, \dots, x_d)$ of the free group F_d on the free generators x_1, \dots, x_d . Given a word w and a group G we consider the *word map* $w = w_G : G^d \rightarrow G$ sending (g_1, \dots, g_d) to $w(g_1, \dots, g_d)$. The image of this map, namely the set of all group elements of the form $w(g_1, \dots, g_d)$ (where $g_i \in G$) is denoted by $w(G)$.

The author acknowledges the support of an Advanced ERC Grant 247034, a BSF grant 2008194, and the Miriam and Julius Vinik Chair in Mathematics which he holds.

Two important questions studied extensively are how large $w(G)$ is, and what is the w -width of G , which is the minimal k such that $w(G)^k = \langle w(G) \rangle$, namely every element of the subgroup generated by $w(G)$ is a product of length k of elements of $w(G)$ (there is also a slightly different definition of width which allows also inverses of elements of $w(G)$, see Section 2 below). Another interesting question is to study the fibers of the word map w , and in particular its kernel, namely the inverse image of 1. See Segal's recent book [87] for more background and further aspects.

There are several motivations for the research directions described above. One is related to the classical Waring problem in Number Theory. Hilbert's solution to this problem shows that every natural number is a sum of $g(k)$ k th powers, where g is a suitable function. Are there analogues of this result in highly non-commutative contexts? In particular for (nonabelian) finite simple groups?

Various such analogues have been provided. For example, in [68] and [83] it was shown that every element of a large finite simple group is a product of $f(k)$ k th powers, and this led to various generalizations, where powers are replaced by arbitrary words, and the unspecified function became explicit, and in fact very small. See [52], [89], [37], [38] and [41].

Another motivation is Serre's question from the 1960s, whether every finite index subgroup of a (topologically) finitely generated profinite group is open. Progress was made by many authors, and a general affirmative answer was provided by Nikolov and Segal in [73, 74]. A major tool in their proof are properties of word maps and results on word width in finite groups. See also [75, 76].

Other motivations for the study of word maps come from the study of residual properties of free groups [12], as well as certain questions in subgroup growth and representation varieties [39].

A useful result proved by Borel [6] in the 1980s states that word maps on simple algebraic groups are dominant maps. This can be applied in the study of word maps on finite simple groups of Lie type [35].

While the width of non-trivial words in large finite simple groups was eventually shown to be two by Larsen, Shalev and Tiep [41], even stronger results hold for some particular words. If w is the commutator word $[x_1, x_2]$ then it turns out that $w(G) = G$, namely every element of a finite simple group is a commutator. Indeed this was recently proved by Liebeck, O'Brien, Shalev and Tiep [47], establishing a longstanding conjecture of Ore from 1951 [78].

Extensions for quasisimple groups were given in [48], [42] and [24]. In particular it turns out that Ore Conjecture also holds for quasisimple groups, with finitely many given exceptions.

Various interdisciplinary methods were developed to study these problems, combining character theory, algebraic geometry, combinatorics, and sometimes analytic number theory and computational group theory as well.

Several important problems on $w(G)$ for finite (often simple) groups G are still open and will be formulated in the next sections. A new challenging direction is to study problems of similar flavor for infinite groups, and most notably arithmetic groups.

A first step is to find a way to “measure” the size of subsets of infinite groups. Of course if G is a topological group with a Haar measure then we have a natural way, but for infinite discrete groups there is no obvious way to measure subsets. This may be handled as follows. Let Γ be an infinite group generated by a finite set S . If Z is a subset of Γ , the asymptotics of the probability that a random walk of length k on the Cayley graph of (Γ, S) lands in Z provides a way to measure the “size” of Z . Furthermore, effective methods for estimating this asymptotics were recently developed in [33] and [62] under the title “Sieve methods in group theory”, imitating classical sieve methods in number theory, and adapting them to the non-commutative world.

The theory of expanders and especially the recent breakthrough on property τ and approximate subgroups (see [25], [26], [7], [82], [8], [86]) enabled these methods to work in some situations. For example, Lubotzky and Meiri [62] showed in this way that the set of all proper powers in a finitely generated linear group (over \mathbb{C}) which is not virtually solvable is “exponentially small”. This is a far reaching extension of some results of Hrushovski, Kropholler, Lubotzky and Shalev on powers in linear groups [27].

In spite of this progress for power maps, the study of word maps for various infinite groups is still in its very beginning, and many natural questions, e.g. on word width in arithmetic groups, are very much open. For example, does Ore Conjecture hold in $\mathrm{SL}_n(\mathbb{Z})$ for $n > 2$?

It is natural to study such questions first for local groups, for example p -adic groups. Progress was recently made in the study of word maps and commutator maps in p -adic groups such as $\mathrm{PSL}_n(\mathbb{Z}_p)$ [2]. In particular Ore Conjecture is proved there if n is a proper divisor of $p - 1$.

Currently there is also interest in word maps on other algebraic structures, such as rings and Lie algebras, but this goes beyond the scope of this survey.

Some words on the structure of this paper. In Section 2 we present results on word width and Waring type problems. Section 3 is devoted to word maps which are surjective on all (or almost all) finite simple groups. In particular this is where Ore Conjecture is discussed. Probabilistic aspects of word maps are discussed in Section 4. In Section 5 we study the size of

the image and the fibers of word maps, and some applications are given. Section 6 is devoted to products of conjugacy classes, Thompson Conjecture and related topics. In Section 7 we discuss character methods, a related zeta function, and their applications to the study of word maps. Finally, Section 8 is devoted to word maps on infinite groups, most notably p -adic and arithmetic groups.

Various open problems and conjecture are suggested throughout this article. We hope that they will inspire future research.

2. WARING TYPE PROBLEMS

A classical result in Number Theory, which goes back to Lagrange, states that every positive integer is a sum of four squares. Results for some larger powers were obtained, culminating in Hilbert's celebrated solution to Waring Problem, showing that every positive integer is a sum of $g(k)$ k th powers, where g is a suitable function (see, for instance, [69]).

Are there analogues of this phenomenon for interesting nonabelian groups, such as (nonabelian) finite simple groups? We are interested in situations where every group element can be expressed as a short product of elements in the image of a given word map.

For a group G and subsets $A, B \subseteq G$ let $AB = \{ab \mid a \in A, b \in B\}$ and $A^k = \{a_1 \cdots a_k \mid a_i \in G\}$.

We start with a fundamental result of Borel [6] (see also [35]) on word maps in algebraic groups.

Theorem 2.1. *Let G be a simple algebraic group over any field K . Let $w \in F_d$ be a non-identity word. Then the word map $w_G : G^d \rightarrow G$ is a dominant morphism.*

This means that the image $w(G)$ is not contained in any proper subvariety of G , and implies that $w(G)$ contains a non-trivial Zarsiki-open subset of G . As a consequence we easily obtain the following.

Corollary 2.2. *Let G be a simple algebraic group over an algebraically closed field K . Let $w_1, w_2 \in F_d$ be two non-identity words. Then we have $w_1(G)w_2(G) = G$.*

In particular it follows that $w(G)^2 = G$ if $w \neq 1$. To deduce Corollary 2.2 from Theorem 2.1, let $O_i \subseteq w_i(G)$ be non-trivial open subsets ($i = 1, 2$), and let $g \in G$. Then $O_1^{-1}g$ and O_2 are non-trivial Zariski-open subsets, and over an algebraically closed field this implies that $O_1^{-1}g \cap O_2$ is non-empty.

If g_2 is in the intersection then $g_2 = g_1^{-1}g$ for some $g_1 \in O_1$, so $g_1g_2 = g$ and $g_i \in w_i(G)$ for $i = 1, 2$, proving the corollary.

It will take much more work to establish a finitary version of this result (see Theorem 2.9 below).

We start with earlier relevant results on finite simple groups. By a theorem of Jones [31] and the classification of finite simple groups, there is no identity which holds in an infinite set of finite simple groups. In other words we have the following.

Theorem 2.3. *For every word $w \neq 1$ there exists a number $N = N(w)$ such that if G is a finite simple group of size at least N then $w(G) \neq \{1\}$.*

By the simplicity of G it follows that $\langle w(G) \rangle = G$ if G is large enough. Can we then find a constant c (which may depend on w but not on G) such that $w(G)^c = G$? This is equivalent to the verbal subgroup of the Cartesian product of all finite simple groups generated by values of w being a closed subgroup.

Various instances of this problem were considered in the past two decades. For $w = [x_1, x_2] = x_1^{-1}x_2^{-1}x_1x_2$, the commutator word, it was shown by Wilson [95] in 1994, using methods of mathematical logic, that indeed $w(G)^c = G$ for some absolute constant c . In 1996-7 Martinez and Zelmanov [68], and independently Saxl and Wilson [83], solved the problem for the power word $w = x_1^k$.

Theorem 2.4. *Given a positive integer k there is a number $f(k)$ such that if G is a finite simple group of exponent not dividing k then every element of G is a product of $f(k)$ k th powers.*

The results on commutators and powers were extended to arbitrary words by Liebeck and myself in [52] as follows.

Theorem 2.5. *For every word $w \neq 1$ there is a positive integer $c = c(w)$ such that if G is a finite simple group such that $w(G) \neq \{1\}$ then $w(G)^c = G$.*

For several years it was not known whether the number $c(w)$ in the theorem above can be replaced by an absolute constant. This problem was very recently solved by Kassabov and Nikolov [32] who provided a negative answer.

Proposition 2.6. *For every positive integer c there is a word $w \neq 1$ and a finite simple group G such that $w(G) \neq \{1\}$ and $w(G)^c \neq G$.*

In this result G can be taken to be an alternating group A_n with n arbitrarily large. Indeed it is shown in [32] that for any $n \geq 7$ there is a

word w such that the image $w(A_n)$ of the word map on A_n is all 3-cycles and the identity element. This easily implies Proposition 2.6, since $|w(A_n)|^c$ is smaller than $|A_n|$ for sufficiently large n .

The construction in [32] was greatly generalized by Lubotzky [61], characterizing images of word maps in finite simple groups. Clearly these images are characteristic subsets (closed under the group automorphisms) containing 1. It turns out that this is the only restriction.

Theorem 2.7. *For every finite simple group G and a characteristic subset S of G with $1 \in S$ there is a word w such that $w(G) = S$.*

Lubotzky's result confirms a conjecture I made during the conference Words and Growth (Jerusalem 2012), and was proved during that conference.

While $c(w)$ in Theorem 2.5 genuinely depends on w , it was shown in [89] that for finite simple groups of sufficiently large order there is no such dependence, and $c(w)$ can be taken to be a surprisingly small number.

Theorem 2.8. *Let $w \neq 1$ be a word. Then there exists a positive integer $N = N(w)$ such that for every finite simple group G with $|G| \geq N(w)$ we have $w(G)^3 = G$.*

This result is obtained in [89]. We actually prove a somewhat stronger result, that there exists a conjugacy class C of G contained in $w(G)$ such that $C^3 = G$. The proof uses probabilistic arguments and character theory, in order to study the distribution of the random variable $y_1 y_2 y_3$ where y_i are randomly chosen elements of the (suitably chosen) conjugacy class $C \subset w(G)$.

It is intriguing that the proof of Theorem 2.8 for alternating groups depends, among other tools, on the Erdős-Turán theory of random permutations (see [13, 14]).

A new proof of Theorem 2.8 was given by Nikolov and Pyber in [77], using the so-called Gowers' trick [21]. Roughly speaking, Gowers' trick shows that if S is a large subset of a finite group G (namely $|S| > |G|/k^{1/3}$ where k is the minimal degree of a non-identity character of G) then $S^3 = G$. Results from [37] and additional work show that $w(G)$ is large enough if G is a (large) finite simple group of Lie type, and this provides a simplified proof for the Lie type case of Theorem 2.8.

Is Theorem 2.8 best-possible, or can its conclusion be strengthened to $w(G)^2 = G$? This was a major open problem for a several years. An affirmative answer was recently given by Larsen, Shalev and Tiep [41]. In fact we obtain a more general result, dealing with two words w_1, w_2 .

Theorem 2.9. *For every two non-identity words $w_1, w_2 \in F_d$ there exists a number $N = N(w_1, w_2)$ such that if G is a finite simple group of order at least N then $w_1(G)w_2(G) = G$.*

In particular we have $w(G)^2 = G$ provided $|G| \geq N(w)$.

Theorem 2.9 may be regarded as a finitary analogue of Corollary 2.2 for simple algebraic groups. It is a best possible solution to this Waring type problem for finite simple groups. Indeed, some words, such as x^2 , are never surjective on finite simple groups, so the word width is sometimes 2. We also note that there is no chance to prove Theorem 2.9 using Gowers' trick and its variations, which are not strong enough to prove results of type $S^2 = G$.

Theorem 2.9 gives rise to the following far reaching improvement of Theorem 2.4.

Theorem 2.10. *For every positive integer k there exists a number $N = N(k)$ such that if G is a finite simple group of order at least N then every element of G is a product of two k th powers.*

This can be regarded as a direct non-commutative analogue of Waring problem in Number Theory. In the original context we need a sum of $g(k)$ k th powers to cover everything, while in the universe of finite simple groups a product of two k th powers suffices. Thus non-commutativity is sometimes an advantage.

The proof of Theorem 2.9 is rather long and complex. The case of alternating groups and groups of Lie type of bounded rank appears in [37, 38]. The proof for classical groups of unbounded rank (and hence for all finite simple groups) appears in [41]. All these proofs employ character methods, and often some additional methods from algebraic geometry and analytic number theory.

Some of the results above can be extended to finite quasisimple groups, sometimes with small modifications. Recall that a group is called *quasisimple* if it is perfect, and simple modulo its center. For example $\mathrm{SL}_n(q)$ are quasisimple (unless $n = 2$ and $q = 2, 3$). In [42] we establish the following.

Theorem 2.11. *For every three non-identity words $w_1, w_2, w_3 \in F_d$ there exists a number $N = N(w_1, w_2, w_3)$ such that if G is a finite quasisimple group of order at least N then $w_1(G)w_2(G)w_3(G) = G$.*

In particular we have $w(G)^3 = G$ if $|G| \geq N(w)$ for quasisimple groups G . Our proof relies on Gowers' trick mentioned above.

Can one obtain a stronger result for quasisimple groups G , namely $w(G)^2 = G$ if $|G| \gg 0$? The answer turns out to be negative: there are words

$w \neq 1$ and infinitely many finite quasisimple groups G for which $w(G)^2 \neq G$. For example, it is shown in [42] and independently in [4], that $x_1^4 x_2^4$ is not surjective on $\mathrm{SL}_2(q)$ for $q \equiv 3, 5 \pmod{8}$. In this sense Theorem 2.11 is best possible.

However, for some families of quasisimple groups, such as the universal covers of alternating groups, we do obtain width 2 results. See [42] for more details.

In [24] Guralnick and Tiep have just proved that for every words $w_1, w_2 \neq 1$ there exists $N = N(w_1, w_2)$ such that if G is a finite quasisimple group of order at least N then $G \setminus Z(G) \subseteq w_1(G)w_2(G)$.

Let us now consider finite groups in general. Here we cannot expect results of the form $w(G)^k = G$ for small k , or indeed for any k , since $w(G)$ need not generate G even if $w(G) \neq \{1\}$. But we can study the w -width of G , defined as the minimal k satisfying $(w(G) \cup w(G)^{-1})^k = \langle w(G) \rangle$, the verbal subgroup generated by $w(G)$. A fundamental question is the following.

Problem 2.12. Which words w have bounded width in all finite groups G with a given number of generators, namely width $\leq f(w, d(G))$ for some function f ?

This is a difficult problem. Some words, such as the metabelian word $[[x_1, x_2], [x_3, x_4]]$, do not have this property. Deep results by Nikolov and Segal establish the following.

Theorem 2.13. *Suppose one of the following holds:*

- (i) $w = [x_1, \dots, x_d]$, a left-normed commutator of length d .
- (ii) w is a locally finite word, namely the variety of groups it defines is locally finite.
- (iii) $w = x_1^k$ for some k .
- (iv) $w \notin F'_d$.

Then w has the property stated in Problem 2.12 above.

See [73, 74] for parts (i) and (ii), and [75] for parts (iii) and (iv) (which generalizes (iii) and follows from it).

Note that this result for $w = x_1^k$ yields an immediate solution to Serre's problem whether finite index subgroups of finitely generated profinite groups are open, simplifying the original proof in [73, 74].

Indeed let G be a (topologically) finitely generated profinite group, and let $H \leq G$ be a finite index subgroup. To show that H is open it suffices to show that its core in G is open, so we may assume that H is normal in G . Let $k = |G : H|$. Then it follows that $g^k \in H$ for all $g \in G$, so $H \geq \langle w(G) \rangle$ where $w = x_1^k$. Set $N = \langle w(G) \rangle$. By Theorem 2.13 every element of N is a

bounded product of k th powers, which implies that N is closed. Thus G/N is a finitely generated profinite group satisfying the identity $x_1^k = 1$. By Zelmanov's solution to the Restricted Burnside Problem [98, 99] it follows that G/N is finite. Hence N is open, and so is H .

See also [76] for various generalizations to finite groups and compact groups.

While this section is devoted mostly to upper bounds on word width, interesting lower bounds are also given in various contexts. For example, Nikolov constructs in [71] finite perfect groups of arbitrarily large commutator width. For certain power words $w = x^k$ the w -width of every non-solvable group is at least 2, namely w -width 1 implies solvability. For example, in [56] we obtain the following somewhat surprising result.

Theorem 2.14. *Let G be a finite group and suppose the set of 12th powers in G is a subgroup of G . Then G is solvable.*

The proof requires the classification of finite simple groups and other tools. It can be shown that 12 is the minimal number with this property, but there are various other numbers k with the property that if the x^k -width of a finite group G is 1 then G is solvable. These numbers are characterized in [57].

In spite of these results, we are still far from a full understanding of finite groups of x^k -width 1, or of w -width 1 for other words w .

3. SURJECTIVE WORDS AND ORE CONJECTURE

In spite of considerable progress in the study of word maps on finite – often simple – groups, some important natural problems are still very much open. Recall that for a word $w \neq 1$ and a large enough finite simple group G we have $w(G)^2 = G$ by Theorem 2.9 above. However, for some words w we may expect more, namely $w(G) = G$. This leads to the following challenging problem.

Problem 3.1. Which words are surjective on all finite simple groups?

A word $w \in F_d$ is called *primitive* if it is part of a free basis for F_d , namely it is an image of the word x_1 by some automorphism of F_d .

Clearly, primitive words are surjective on any group, but other words also have this property. Indeed the words which are surjective on any group were characterized by Segal (see [87], Lemma 3.1.1) as follows.

Proposition 3.2. *A word $w \in F_d$ is surjective on every group if and only if $w \in x_1^{k_1} \cdots x_d^{k_d} F_d'$ for some integers k_1, \dots, k_d which satisfy $\gcd(k_1, \dots, k_d) = 1$.*

For example, $x_1^2 x_2^3$ is not primitive but always surjective.

Various words are not surjective on all groups, and yet they are surjective on all finite simple groups. A longstanding conjecture of Ore [78], posed in 1951, states that the commutator word has this property.

Conjecture 3.3 (Ore Conjecture). *Every element of every finite simple group is a commutator.*

This conjecture attracted much attention, and various related results were obtained over the years. The case of alternating groups was handled by Ore [78]. Special linear groups were treated by Thompson [91, 92, 93] in the 1960s. A breakthrough was obtained in 1998 by Ellers and Gordeev [15] who proved the conjecture for simple groups of Lie type over fields of size exceeding 8. Finally, in 2010 the conjecture was fully proved by Liebeck, O'Brien, Tiep and myself [47], so we have

Theorem 3.4. *Ore Conjecture holds.*

The proof in [47] combines 3 ingredients: representation theory and character methods, induction on the dimension, and computational group theory.

Roughly speaking we show, by combining Frobenius character formula (see Proposition 7.1 below) with various (some new) results on characters of finite groups of Lie type, that elements with small centralizers are commutators. Then we show that the remaining elements are breakable, in the sense that they lie in a product of subgroups of Lie type of smaller dimension, and deduce by induction that they too are commutators. This argument works for groups in large enough dimension, and for various groups of smaller dimension Ore Conjecture should be verified directly (the induction base). This is done using computational group theory, and in fact required 3 years of CPU time.

Note that it follows from Theorem 3.4 that longer commutators in distinct variables are also surjective on all finite simple groups. However when variables repeat this not clear. A challenging case is that of Engel words $e_n = [x, y, y, \dots, y]$, where y occurs n times.

A classical theorem of Zorn states that a finite group is nilpotent if and only if it satisfies the identity $e_n = 1$ for some n . Simple groups may be considered as the opposite of nilpotent groups, and for them we expect the Engel words to have the maximal possible image, namely the whole group. We therefore propose the following.

Conjecture 3.5. *All Engel words are surjective on all finite simple groups.*

This was verified by computer for various finite simple groups by O'Brien and others. Some theoretical work was done by Puder and Schul on Engel words in alternating groups, but Conjecture 3.5 is still open even in this case.

Bandman, Garion and Grunewald [3] studied Conjecture 3.5 for $SL_2(q)$ and $PSL_2(q)$ using the trace method, obtaining the following.

Proposition 3.6. *For every $n \geq 2$ there is a number $q_0(n)$ such that if $q \geq q_0(n)$ is a prime power then the Engel word e_n is surjective on $PSL_2(q)$.*

Let us now turn to other words. In [41] it is conjectured that the word x^2y^2 is always surjective on finite simple groups. This was verified independently in [49] and in [22], using completely different methods, so we have:

Theorem 3.7. *Every element of every finite simple group is a product of two squares.*

This result may be regarded as a non-commutative analogue of Lagrange four squares theorem in Number Theory. It is intriguing that it holds for all finite quasisimple groups too, as shown in [42].

Theorem 3.7 can be generalized as follows.

Theorem 3.8. *Let $n = p^k$ be any prime power. Then every element of every finite simple group is a product of two n th powers.*

This was proved in [22] in general, and in [49] for $p > 7$.

Recall that, by a classical result of Burnside, groups of order $p^a q^b$ (where p, q are primes) are always solvable. Thus if n is divisible by at most two primes then x^n is not an identity in any finite simple group. We suggest the following conjecture which is a common generalization of Burnside's $p^a q^b$ Theorem and Theorem 3.8 above.

Conjecture 3.9. *Let $n = p^a q^b$ where p, q are primes. Then every element of every finite simple group is a product of two n th powers.*

Work in progress by Guralnick, Liebeck, O'Brien, Tiep and myself settles this in many cases, and we hope a proof of Conjecture 3.9 will be completed soon.

A related problem is the following.

Problem 3.10. Which words are surjective on almost all (namely all but finitely many) finite simple groups?

Note that words of the form w_1w_2 where w_1, w_2 are any two non-trivial words in disjoint sets of variables have this property, by Theorem 2.9 above.

On the other hand power words $w = v^k$ for $k > 1$ and some word v do not have this property, since there are infinitely many finite simple groups whose image is not coprime to k , hence the k th power map on them is not surjective.

For some time it was speculated that perhaps every word which is not a proper power is surjective on almost all finite simple groups. A breakthrough was recently made by Jambor, Liebeck and O'Brien in [29].

Theorem 3.11. *There are words w which are not proper powers and which are not surjective on infinitely many finite simple groups.*

The idea of the proof is to focus on simple groups of type $\mathrm{PSL}_2(q)$ for some infinite collection of q and to use trace methods for $G = \mathrm{SL}_2(q)$, as developed in [3]. Words w are then constructed with the property that the trace of elements of $w(G)$ is never zero, which implies that w is not surjective on $\mathrm{PSL}_2(q)$. An example of such a word is $x_1^2[x_1^{-2}, x_2^{-1}]^2$.

We conjecture that such examples do not exist in large rank.

Conjecture 3.12. *Let w be a word which is not a proper power. Then there exists a positive integer $N = N(w)$ such that w is surjective on all alternating groups of degree at least N , and on all finite simple classical groups of rank at least N .*

4. PROBABILISTIC ASPECTS

A word map $w : G^d \rightarrow G$ gives rise to a probability distribution $P_{w,G}$ on the finite group G by defining

$$P_{w,G}(X) = |w^{-1}(X)|/|G|^d$$

for subsets $X \subseteq G$. If $P_{w,G}$ is uniform on G we say that w is *uniform* on G (namely $|P_{w,G}(X)| = |X|/|G|$ for any subset $X \subseteq G$).

Deep results on the distribution of $P_{w,G}$ for symmetric groups $G = S_n$ were obtained by Nica [70] in 1994. Suppose $w = v^k$ where $k \geq 1$ is maximal (k is then called the exponent of w). Roughly speaking Nica shows that the behavior of P_{w,S_n} for large n is similar to the behavior of $P_{x_1^k, S_n}$ if we focus on the distribution of the number of m -cycles (where m is bounded) in $w(g_1, \dots, g_d)$ where $g_i \in S_n$ are randomly chosen.

Certain words w are uniform on all finite groups G for an obvious reason. These are the primitive words.

A natural conjecture, posed by several people at the Hebrew University, is that the converse also holds. The case of F_2 was proved by Puder [80]. Parzanchevski and Puder [81] then proved the full conjecture, so we have the following.

Theorem 4.1. *A word $w \in F_d$ is uniform on all finite groups if and only if w is primitive.*

In fact it is shown in [81] that if w is uniform on infinitely many symmetric groups S_n then w is primitive. This is obtained by studying the random variable Y_n which counts the number of fixed points of $w(g_1, \dots, g_d)$ where $g_i \in S_n$ are random permutations. It is shown (by combining Nica's methods with additional tools) that if w is not primitive then the expectation of Y_n for large n is not 1. This implies that w is not uniform on S_n .

One may ask a more general question: for which words $u, v \in F_d$ we have $P_{u,G} = P_{v,G}$ for all finite groups G ? If $v = \phi(u)$ for some $\phi \in \text{Aut}(F_d)$ then u and v clearly induce the same distribution on every finite group. We conjecture that the converse also holds.

Conjecture 4.2. *Words $u, v \in F_d$ induce the same distribution on every finite group if and only if they are equivalent under the $\text{Aut}(F_d)$ -action.*

If true, this would be a far reaching extension of Theorem 4.1 (which confirms the conjecture in the case $u = x_1$).

Next we discuss almost uniformity of words. If $P_{w,G}$ converges in the L_1 -norm to a uniform distribution on G as G ranges over a family of finite groups we say that w is *almost uniform* on this family.

We can show that various non-primitive words are almost uniform on finite simple groups. The first result of this type was obtained by Garion and myself [18]:

Theorem 4.3. *The commutator word $[x_1, x_2]$ is almost uniform on finite simple groups.*

The proof is character-theoretic (see Section 7 below). Combining Theorem 4.3 with the proven Dixon Conjecture that two random elements of a finite simple group generate the group with probability tending to 1 with the group order (see [51] and the references therein) we obtained the following consequence.

Corollary 4.4. *Almost every element of a finite simple group G can be written as a commutator $[g, h]$ where (g, h) is a generating pair for G .*

This is used in [18] to prove a conjecture of Guralnick and Pak [23] showing that the number of connected components of the Product Replacement Graph of a finite simple group G on two generators tends to infinity as $|G| \rightarrow \infty$. This demonstrates yet another application of the theory of word maps.

We also show in [18] that the word $x_1^2 x_2^2$ is almost uniform on finite simple groups. See also [3, 4] for results on almost uniformity of words of type $x_1^n x_2^m$ and of Engel words e_n on $\mathrm{PSL}_2(q)$.

Other words might have similar properties. We suggest the following.

Conjecture 4.5. *If w_1, w_2 are non-trivial words in disjoint variables, then $w_1 w_2$ is almost uniform on finite simple groups.*

Positive evidence is given below.

Theorem 4.6.

- (i) *Conjecture 4.5 holds for alternating groups A_n .*
- (ii) *Conjecture 4.5 holds for finite simple groups of Lie type of bounded rank.*
- (iii) *For any positive integers n, m , the word $x_1^n x_2^m$ is almost uniform on finite simple groups.*
- (iv) *Any admissible word is almost uniform on finite simple groups.*

Here a reduced word $w(x_1, \dots, x_d) \neq 1$ is called *admissible* if each x_i occurs twice in w , once with exponent 1 and once with exponent -1 . Thus commutators are admissible, as well more general words such as $x_1 x_2 \cdots x_d x_1^{-1} x_2^{-1} \cdots x_d^{-1}$, and so on.

Part (i) of Theorem 4.6 is obtained in [38, Theorem 1.18], and parts (ii), (iii) and (iv) are the main results of [40]. Again character methods play a crucial role in all these proofs.

In spite of the results mentioned above, the following general question is very much open.

Problem 4.7. Which words are almost uniform on finite simple groups?

Note that power words $w = v^k$ ($k > 1$) are not almost uniform, since $|w(G)|/|G|$ is bounded away from 1 on infinitely many finite simple groups G . Are non-power words always almost uniform? This seems too strong to be true, but at present we have no counter example.

Character methods are also useful in studying random walks on finite simple groups G with respect to $w(G)$ as the generating set. Such a walk starts with 1 and at each time it multiplies the group element reached so far by a random element chosen (uniformly) from $w(G)$. Denote the

distribution of such a walk at time t by P_t . If the L_1 -distance between P_t and the uniform distribution U on G is smaller than some prescribed ε (say $1/e$) we say that the *mixing time* $T(w(G), G)$ of this random walk is at most t .

We can now state the following.

Theorem 4.8. *Let w be a non-identity word, and let G be a finite simple group. Then the mixing time $T(w(G), G)$ is equal to 2 if G is large enough. In fact, if g_1, g_2 are two randomly chosen elements of $w(G)$, then the distribution of the random variable g_1g_2 tends (in the L_1 -norm) to the uniform distribution on G as $|G| \rightarrow \infty$.*

This result for alternating groups was obtained in [38, Theorem 1.17], and the result for groups of Lie type was proved by Schul and myself [85, Theorem 1.1]. Theorem 4.8 immediately implies that $|w(G)^2| \geq (1 - \varepsilon)|G|$ for any given $\varepsilon > 0$ and a large enough finite simple group G (but proving that $w(G)^2 = G$ is harder).

5. IMAGE SIZE AND FIBER SIZE

In previous sections we discussed word width and surjectivity questions. This section is devoted to the study of the size of the image of a word map, as well as the size of the kernel (and arbitrary fibers) of this map. Some of the results presented here are useful in the proof of results stated in previous sections.

It turns out that, for large finite simple groups G , $|w(G)|$ is in some sense quite close to $|G|$.

We start with the interesting particular case of the power words in symmetric groups (the asymptotics in alternating groups is similar). In [34] Larsen proved the following.

Proposition 5.1. *Let $k \geq 1$ be an integer and let $w = x_1^k$. Then*

$$|w(S_n)| \sim n^{-b}n!,$$

where $b = 1 - \phi(k)/k$ and ϕ is the Euler function.

This implies, for example, that the probability that a random element of S_n (or A_n) is a square is roughly $n^{-1/2}$. Note that, choosing k suitably (e.g. as the product of the first m primes) we may arrange that $b \geq 1 - \varepsilon$ for any fixed $\varepsilon > 0$.

We now turn from powers to arbitrary words in alternating groups. A result of Larsen and myself [37] shows the following.

Theorem 5.2. *For each non-trivial word w and $\varepsilon > 0$, there exists $N = N(w, \varepsilon)$ such that if $n \geq N$ then*

$$|w(A_n)| \geq n^{-\frac{29}{9}-\varepsilon} |A_n|.$$

It follows from Proposition 5.1 that this bound is tight up to the value of the exponent, which must be at most -1 . It is intriguing that the proof of Theorem 5.2 depends also on Analytic Number Theory and Vinogradov's three primes theorem.

We now turn to simple groups of Lie type. In those which are of bounded rank the word values have positive proportion. Indeed by a result of Larsen [35] we have:

Theorem 5.3. *Let G be a finite simple group of Lie type of rank r and let $w \neq 1$ be a word. Then there exists a constant $c = c(r, w) > 0$ depending only on r and w such that*

$$|w(G)| \geq c|G|.$$

The proof relies on algebraic geometry (see also [39]). Another result from [35] deals with finite simple groups in general.

Theorem 5.4. *Let $w \neq 1$ be a word. Then*

$$\liminf \log |w(G)| / \log |G| = 1,$$

as G ranges over all finite simple groups.

Thus the Hausdorff dimension of $w(G)$ tends to 1. This means that for every $\varepsilon > 0$ and every sufficiently large finite simple group G we have $|w(G)| \geq |G|^{1-\varepsilon}$.

Eventually better lower bounds on the size of $w(G)$ were obtained in [37]. For groups of Lie type we have:

Theorem 5.5. *For every word $w \neq 1$ there is a number $N = N(w)$ such that if G is a finite simple group of Lie type of rank r which is not of type A_r or 2A_r , and $|G| \geq N$, then*

$$|w(G)| \geq cr^{-1}|G|,$$

for some absolute constant $c > 0$.

Weaker lower bounds on $|w(G)|$ for groups G of Lie type A_r or 2A_r are obtained by Nikolov and Pyber [77].

We conjecture that Theorem 5.5 holds for all Lie types, and also for alternating groups, in the following sense.

Conjecture 5.6. For every word $w \neq 1$ there exists a number $N = N(w)$ such that if G is an alternating group of degree n or a finite simple group of Lie type of rank n , and $|G| \geq N$, then

$$|w(G)| \geq cn^{-1}|G|,$$

where $c > 0$ is an absolute constant.

For words which are not proper powers stronger conclusions might follow.

Problem 5.7. Suppose w is not a power word, and let G be a finite simple group.

- (i) Is it true that there exist $N, c > 0$ depending on w such that if $|G| \geq N$ then $|w(G)| \geq c|G|$?
- (ii) Is it true that for every $\varepsilon > 0$ there exists $N = N(w, \varepsilon)$ such that if $|G| \geq N$ then $|w(G)| \geq (1 - \varepsilon)|G|$?

Note that Conjecture 3.12 on the surjectivity of word maps in large rank implies an affirmative answer to part (i) above (using Theorem 5.3).

We now turn to the kernel of word maps and related topics. We sometimes denote the word map w on G^d by w_G .

For a finite group G define $P_G(w) = |\text{Ker}(w_G)|/|G^d|$, the probability that $w(g_1, \dots, g_d) = 1$ where $g_i \in G$ are randomly chosen. Thus $P_G(w) = 1$ if and only if G is an identity in G . In [12] it was shown that, for $w \neq 1$ and finite simple groups G ,

$$P_G(w) \rightarrow 0 \text{ as } |G| \rightarrow \infty,$$

and this result yielded a simplified solution of a problem posed by Magnus, showing that the free group F_d is residually S for any infinite set S of finite simple groups.

How fast does $P_G(w)$ tend to zero? If w is a primitive word then it is uniform on all finite groups (see Section 4) and so $P_G(w) = |G|^{-1}$. If $w = [x_1, x_2]$, the commutator word, then it is well known that $P_G(w) = k(G)/|G|$, where $k(G)$ is the number of conjugacy classes of G . For power words $w = x_1^k$, and G an alternating group of large degree or a classical group of large rank one can show that $P_G(w)$ is close to $|G|^{-1/k}$ (see [96] and Theorem 1.4 in [54]). Can we find estimates for general words?

The following result from [39] provides a strong bound on $P_G(w)$ for general words w and finite simple groups G . In fact it holds for general fibers of the word map, not just the fiber above 1.

Theorem 5.8. *For every word $1 \neq w \in F_d$ there exist $\varepsilon = \varepsilon(w) > 0$ and $N = N(w) > 0$ such that for every finite simple group G of order at least N and every element $g \in G$ we have*

$$|w_G^{-1}(g)| \leq |G|^{d-\varepsilon}.$$

In particular we have (under the above assumptions)

$$P_G(w) \leq |G|^{-\varepsilon}.$$

This bound is best possible: the power word example mentioned above shows that ε indeed depends on w , and may be arbitrarily close to zero.

We present two applications of Theorem 5.8.

Recall that, for a group H and a positive integer n , $a_n(H)$ denotes the number of index n subgroups of H . For background on subgroup growth, see the book [63] by Lubotzky and Segal. It is well known (see [63, 1.2]) that $a_n(F_d) \sim n \cdot (n!)^{d-1}$. It turns out that non-free groups on d generators have significantly smaller subgroup growth.

Corollary 5.9. *Let H be a non-free group with d generators. Then there exists $\varepsilon > 0$ such that*

$$a_n(H) \leq (n!)^{d-1-\varepsilon}$$

for all sufficiently large n .

Since the deduction of this corollary is rather short we outline it below. It is well known (see [63, 1.1]) that

$$a_n(H) \leq |\mathrm{Hom}(H, S_n)| / (n-1)!.$$

Suppose H is generated by h_1, \dots, h_d . Since H is not free there is a non-trivial word $w \in F_d$ such that $w(h_1, \dots, h_d) = 1$. Therefore every $\phi \in \mathrm{Hom}(H, S_n)$ satisfies $w(\phi(h_1), \dots, \phi(h_d)) = 1$, which implies that

$$|\mathrm{Hom}(H, S_n)| \leq |\mathrm{Ker}(w_{S_n})|.$$

By Theorem 5.8 (more precisely, by its version to S_n which holds too) there exists $\delta > 0$ such that for all large n we have

$$|\mathrm{Ker}(w_{S_n})| \leq (n!)^{d-\delta}.$$

This implies

$$a_n(H) \leq |\mathrm{Hom}(H, S_n)| / (n-1)! \leq n(n!)^{d-1-\delta} \leq (n!)^{d-1-\varepsilon}$$

for any fixed $0 < \varepsilon < \delta$ and sufficiently large n . This concludes the proof of Corollary 5.9.

Our next application concerns representation varieties. Let K be an algebraically closed field, and let H be a finitely generated group. Then for each positive integer n one may form the representation variety $\text{Hom}(H, \text{GL}_n(K))$. These varieties and their dimensions have been widely studied for various groups, see for instance [54] for the case of Fuchsian groups. Clearly $\dim \text{Hom}(F_d, \text{GL}_n(K)) = dn^2$. Applying Theorem 5.8 we show that these dimensions for non-free d -generated groups are substantially smaller. We also obtain a similar result for more general algebraic groups.

Corollary 5.10. *Let H be a non-free group with d generators. Then there exists $\varepsilon > 0$ such that*

- (i) $\dim \text{Hom}(H, \text{GL}_n) \leq (d - \varepsilon)n^2$ for all $n > 1$;
- (ii) $\dim \text{Hom}(H, \underline{G}) \leq (d - \varepsilon) \dim \underline{G}$ for any semisimple algebraic group \underline{G} .

The above result shows that the subvariety of \underline{G}^d defined by the equation $w(g_1, \dots, g_d) = 1$ (where $1 \neq w \in F_d$) is not only proper (hence of codimension at least 1), but has large codimension (at least $\varepsilon \dim \underline{G}$).

We emphasize that these results hold over algebraically closed fields K of arbitrary characteristic.

There is some interest in the fibers of some special word maps such as the commutator map. This can be translated to the study of certain character sums. See Section 7 below for details.

6. CONJUGACY CLASSES AND THOMPSON CONJECTURE

Let C be a conjugacy class in a finite simple group G . If $C \neq \{1\}$ then there exists a natural number k such that $C^k = G$. Hence there exists k so that $C^k = G$ for all non-trivial classes C , and the minimal such k is defined to be the *covering number* of G .

While the covering number of alternating groups was found long ago (see [1]) the case of groups of Lie type was solved (asymptotically) much later, see [16] and [43].

Theorem 6.1.

- (i) *The covering number of A_n is $\lceil n/2 \rceil$.*
- (ii) *The covering number of a finite simple group of Lie type of rank r is at most cr , where c is an absolute constant.*

The upper bound in (ii) is sharp up to the value of the constant c .

The covering number represents the worst case situation: for most conjugacy classes we have $C^k = G$ for much smaller values of k .

Note that $C^k = G$ implies $k \geq \log |G| / \log |C|$ by a trivial counting argument. The following theorem from [52] provides a similar upper bound on the minimal k satisfying $C^k = G$. In fact it deals not just with conjugacy classes but also with their unions, namely normal subsets in general.

Theorem 6.2. *There exists an absolute constant c such that for any finite simple group G and a normal subset S of G of size > 1 we have $S^k = G$ for some $k \leq c \log |G| / \log |S|$.*

Theorem 6.2 was a main tool in proving Theorem 2.5 above, since it can be applied to the normal subset $S = w(G)$. Moreover, combining it with Theorem 5.4 which implies e.g. that $|w(G)| \geq |G|^{1/2}$ if $|G| \geq N(w)$, we obtain (under the same assumption) $w(G)^{2c} = G$ where c is the absolute constant in Theorem 6.2.

While Theorem 6.1 represents the worst case, and Theorem 6.2 the general case, the following conjecture deals with the best case, namely with the minimal k for which $C^k = G$ for some class C .

Conjecture 6.3 (Thompson Conjecture). *Every finite simple group G has a conjugacy class C such that $C^2 = G$.*

Note that Thompson Conjecture implies Ore Conjecture discussed in Section 3. Indeed, suppose $C^2 = G$ for a class C in G . Then $1 \in C^2$, hence $C^{-1} = C$. We obtain $C^{-1}C = G$, so each $g \in G$ can be written as $g = x^{-1}x^y = [x, y]$ for some $x \in C$ and $y \in G$.

Much work has been done on Thompson Conjecture, see [15] and the references therein. The conjecture is still open for simple groups of Lie type over small fields. In recent years some approximations to Thompson Conjecture have been obtained.

In [89] we show, using probabilistic and character methods, that every sufficiently large finite simple group G has a conjugacy class C such that $C^3 = G$.

This has just been improved in the preprint [24] by Guralnick and Tiep as follows.

Theorem 6.4. *Every finite simple group G has a conjugacy class C of elements of prime order such that $C^3 = G$.*

In [90] we obtain the following probabilistic approximation to Thompson Conjecture.

Theorem 6.5. *Let $x \in G$ be a randomly chosen element of a finite simple group G , and let $C = x^G$ be its conjugacy class. Then for any $\varepsilon > 0$ the probability that $|C^2| \geq (1 - \varepsilon)|G|$ tends to 1 as $|G| \rightarrow \infty$.*

Therefore there is a conjugacy class C whose square almost covers G (for large G).

More progress was recently made by Larsen, Shalev and Tiep [41] and by Guralnick and Malle [22], leading to the following.

Theorem 6.6. *Every finite simple group G has conjugacy classes C_1 and C_2 such that $C_1 C_2 \supseteq G \setminus \{1\}$.*

In [41] this is proved for all finite simple groups whose order exceeds some (explicit) constant. Subsequently this was shown in [22] to hold with no exceptions.

In the proof in [22] for groups of Lie type one may choose C_1, C_2 to be semisimple classes. This yields the following interesting consequence.

Corollary 6.7. *Let G be a finite simple group of Lie type. Then every element of G can be written as a product of two semisimple elements.*

It was shown earlier by Ellers and Gordeev that every element of a finite simple group of Lie type is a product of two unipotent elements (see [15] and the references therein).

In addition to covering results of the type $C^k = G$, there is considerable interest in random walks on finite (almost) simple groups G with respect to a conjugacy class C as a generating set. See Diaconis and Shahshahani [11] for transpositions in symmetric groups, Lulov [64] and Vishne [94] for homogeneous classes in symmetric groups, Lulov and Pak [65] for cycles in symmetric groups, and [52], [55] for groups of Lie type. A main problem investigated is determining the mixing time $T(C, G)$ of the random walk (see Section 4 above).

The following result from [38] gives rather sharp bounds on mixing times in A_n .

Theorem 6.8. *Let $\sigma \in A_n$, and $C = \sigma^{S_n}$, and let $T = T(C, A_n)$ denote the mixing time of the associated random walk on A_n .*

- (i) *The mixing time T is bounded if and only if σ has at most n^α fixed points, where $\alpha < 1$ is bounded away from 1.*
- (ii) *If σ has n^α fixed points where $\alpha < 1$ then*

$$(1 - \alpha)^{-1} \leq T \leq 2(1 - \alpha)^{-1} + 1.$$

- (iii) *If σ is fixed-point-free or has $n^{o(1)}$ fixed points then $T \leq 3$.*
- (iv) *If σ has at most $n^{o(1)}$ cycles of length 1 and 2 then $T = 2$.*

Parts (iii) and (iv) are best possible, and extend Lulov's result [64] for permutations σ which consist of n/m m -cycles (where the mixing time is 3 if $m = 2$ and 2 if $m \geq 3$). Part (iii) confirms a conjecture of Lulov and Pak in [65] that the mixing time of a fixed-point-free class of permutations is 2 or 3.

In [90] we obtain a somewhat surprising result for general simple groups G , showing that the mixing time $T(C, G)$ is usually the smallest possible, namely 2.

Theorem 6.9. *Let G be a finite simple group, let $x \in G$ be randomly chosen, and let $C = x^G$ be its conjugacy class. Then the probability that $T(C, G) = 2$ tends to 1 as $|G| \rightarrow \infty$.*

We show a bit more, namely that the product of two random elements of a "typical" class C is almost uniformly distributed on G . Note that Theorem 6.9 implies Theorem 6.5 above, and this is how the latter result is proved.

Next we consider class expansion in finite simple groups G . The following was obtained in [89].

Theorem 6.10. *Let G be a finite simple group, and let C be a conjugacy class of G . Then we have*

- (i) $|C^3| \geq \min(|C|^{1+\varepsilon}, |G|)$ where $\varepsilon > 0$ is some absolute constant.
- (ii) If G is of Lie type of bounded rank then $|C^2| \geq |C|^{1+\varepsilon}$ for some absolute constant $\varepsilon > 0$.

Note that, by the celebrated Product Theorem obtained later in [82] and [7], the conclusion in part (i) above holds for any generating set C of G , provided G is of Lie type of bounded rank.

A recent result from [19] extends part (ii) of Theorem 6.10 for all finite simple groups, provided the class C is small. In fact it deals with normal subsets in general.

Theorem 6.11. *There exist absolute constants $\varepsilon, \delta > 0$ with the property that if G is any finite simple group and C is a normal subset of G such that $|C| \leq |G|^\delta$ then $|C^2| \geq |C|^{1+\varepsilon}$.*

This confirms a conjecture I made a few years ago.

We conjecture below that for small conjugacy classes C much more can be said, namely, the size of C^2 is almost maximal, namely almost $|C|^2$.

Conjecture 6.12. *For every $\varepsilon > 0$ there exists $\delta > 0$ such that if G is any finite simple group and $C \subset G$ is any conjugacy class of G satisfying $|C| \leq |G|^\delta$ then $|C^2| \geq |C|^{2-\varepsilon}$.*

It suffices to prove the conjecture for alternating groups and for classical groups of large rank. Indeed, in the bounded rank case we may choose δ small enough which would imply $C = \{1\}$, which trivially satisfies the required conclusion.

Work in progress by Gili Schul [84] confirms the conjecture in some cases, including that of alternating groups.

We conclude this section by discussing a recent very general conjecture posed by Liebeck, Nikolov and myself in [46].

Conjecture 6.13. *There exists an absolute constant c such that if G is a finite simple group and A is any subset of G of size at least two, then G is a product of N conjugates of A for some $N \leq c \log |G| / \log |A|$.*

Note that we must have $N \geq \log |G| / \log |A|$ by order considerations, and so the bound above is tight up to a multiplicative constant.

Conjecture 6.13 may be regarded as a far reaching generalization of Theorem 6.2 dealing with the case where A is a normal subset (so all its conjugates coincide with itself). It is also a stronger version of a recent conjecture we posed in [45], where A was assumed to be a subgroup of G . Positive evidence for the latter conjecture is provided by [50] (when A is a Sylow subgroup) and [72, 44, 60] (when A is of type SL_n), with applications to bounded generation and expanders. Further results were proved in [45] in various cases where A is a maximal subgroup of G , but the general case is still open.

The following results provide positive evidence for the stronger conjecture stated above, regarding subsets. The first result, from [46], can be stated as follows.

Theorem 6.14. *Conjecture 6.13 holds if G any finite simple group and the subset A has bounded size.*

We actually show that there is an absolute constant c such that every subset A of G with $|A| \geq 2$ has $N \leq c \log |G|$ conjugates whose product is G .

The second result was proved subsequently by Gill, Pyber, Short and Szabó [19].

Theorem 6.15. *Conjecture 6.13 holds for all subsets A provided G is a finite simple group of Lie type of bounded rank.*

Tools from the recent theory of approximate subgroups (see [82] and [7]) are very useful in this context.

7. CHARACTER METHODS

The proofs of the results described above require various tools, and we will not attempt to describe all of them in this survey paper. Rather, we shall highlight a major tool, namely character methods, which plays an essential role in many of these proofs. See Isaacs [28] for background on general character theory and Lusztig [66] for characters of groups of Lie type.

Let $\text{Irr } G$ denote the set of complex irreducible characters of the finite group G . To explain the connections to character theory, we start by discussing the so called non-commutative Fourier transform. Recall that with a word $w \in F_d$ and a finite group G we associated a probability measure $P_{w,G}$ on G . For $g \in G$ we have $P_{w,G}(g) = |w^{-1}(g)|/|G|^d$, the probability to obtain g as $w(g_1, \dots, g_d)$ where $g_i \in G$ are randomly chosen.

Clearly, $P_{w,G}$ is a class function on G , and as such it can be expressed uniquely as a linear combination of irreducible characters. For convenience we write

$$P_{w,G} = |G|^{-1} \sum_{\chi \in \text{Irr } G} a_{w,\chi} \chi,$$

where $a_{w,\chi} \in \mathbb{C}$ are the so called Fourier coefficients. Using an inverse Fourier transform we may reconstruct the Fourier coefficients $a_{w,\chi}$ as follows.

$$a_{w,\chi} = \frac{1}{|G|^d} \sum_{g_1, \dots, g_d \in G} \chi(w(g_1, \dots, g_d)^{-1}),$$

which is the average value of the character χ on $w(\bar{g})^{-1}$. In particular we have $a_{w,1} = 1$ for all words w . Unfortunately, the formula above is not easy to use, and the Fourier coefficients are not known for arbitrary words, but they have been determined in some important cases.

Consider first the commutator word $w = [x_1, x_2]$. By a classical result of Frobenius from 1896 we have $a_{w,\chi} = 1/\chi(1)$. In other words we have

Proposition 7.1. *Let G be a finite group, and let $g \in G$. Then the number $N(g)$ of pairs $(x, y) \in G \times G$ such that $[x, y] = g$ satisfies*

$$N(g) = |G| \sum_{\chi \in \text{Irr } G} \frac{\chi(g)}{\chi(1)}.$$

Consequently, an element $g \in G$ is a commutator if and only if $\sum_{\chi} \frac{\chi(g)}{\chi(1)} \neq 0$.

Now, if G is a finite simple group, and $g \in G$ is an element with a small centralizer, then one can use recent advances in representation theory (based on Deligne-Lusztig theory [66] and other tools) to show that the main contribution to the character sum in Proposition 7.1 comes from the trivial character $\chi = 1$, and all the other characters altogether contribute marginally. This means that, for such elements g we have $\sum_{\chi} \frac{\chi(g)}{\chi(1)} = 1 + o(1)$ and so $N(g) = |G|(1 + o(1))$ and in particular g is a commutator when G is large enough.

This argument shows that elements with a small centralizer are commutators. To prove Ore Conjecture (Theorem 3.4 above) one needs to show that the remaining elements are also commutators, and this is done using a (somewhat complex) inductive argument.

Frobenius formula is also a key tool in the proof of Theorem 4.3 on the almost uniformity of the commutator map on finite simple groups. In fact the method in [18] yields a similar result for arbitrary finite groups whose representation growth is very small.

A central concept in many of our character-theoretic arguments is the so called *Witten zeta function* ζ^G encoding the character degrees of a finite group G . For a real number s define

$$\zeta^G(s) = \sum_{\chi \in \text{Irr } G} \chi(1)^{-s}.$$

This function was defined by Witten [97] for real Lie groups and was studied and applied extensively in [53, 54, 55] for finite simple groups. See also [36] for the case of algebraic, arithmetic and linear groups.

It turns out that the value of $\zeta^G(s)$ for certain numbers s encodes key information on various properties of G . To illustrate this in the context of the commutator map, let U be the uniform distribution on G , and let $P = P_{w,G}$ be the commutator distribution (so that $P(g) = N(g)/|G|^2$). Using non-commutative Fourier techniques one can show that

$$\|P - U\| \leq \left(\sum_{\chi \neq 1} \chi(1)^{-2} \right)^{1/2} = (\zeta^G(2) - 1)^{1/2},$$

where the above norm is the L_1 -norm. Hence if G is a finite group such that $\zeta^G(2)$ is very close to 1, then the commutator map from $G \times G$ to G is almost uniform.

Now, in [54] we show the following.

Theorem 7.2. *Fix a real number $s > 1$, and let G be a finite simple group. Then $\zeta^G(s) \rightarrow 1$ as $|G| \rightarrow \infty$.*

Using this for $s = 2$ we obtain a proof of Theorem 4.3.

While the commutator distribution P is almost uniform on finite simple groups G , we still do not know enough about the values of $P(g)$ for elements $g \in G$ and how close they are to $|G|^{-1}$. Note that $P(1) = k(G)/|G|$ where $k(G)$ is the number of conjugacy classes of G , which tends to infinity as $|G| \rightarrow \infty$. However, if $g \neq 1$, can we show that $P(g)$ is close to $|G|^{-1}$? By Proposition 7.1 we have $P(g) = |G|^{-1}S(g)$, where $S(g) = \sum_{\chi \in \text{Irr } G} \chi(g)/\chi(1)$.

Problem 7.3. Study the character sum $S(g)$ as G ranges over natural families of finite simple groups. For which elements $g \in G$ we have $S(g) \geq 1 - \varepsilon$, or $S(g) \geq \varepsilon$ for a fixed $\varepsilon > 0$?

For some time it was conjectured that $S(g) \geq 1 - \varepsilon$ for all large finite simple groups G and $g \in G$, so that (asymptotically) no elements are under-represented as commutators. This was refuted by Liebeck and myself. We showed that transvections g in $\text{PSU}_3(q)$ and $\text{PSU}_5(q)$ are under-represented as commutators in the sense that $S(g) = o(1)$, and these are the only exceptions in these groups. It would be interesting to find out whether this is a special case of some general phenomenon.

Estimation of the character sums $S(g)$ requires understanding cancellation phenomena, since in many cases $\sum |\chi(g)|/\chi(1)$ is very large while $S(g)$ is still very close to 1.

Having discussed commutators let us examine Fourier expansion for other words. For $w = x^2$ the Fourier coefficients are known, and $a_{w,\chi}$ coincides with the Schur indicator of the character χ , which is 1, 0 or -1 .

Now, if u, v are words in disjoint sets of variables, then we have

$$P_{uv,G} = P_{u,G} * P_{v,G},$$

the convolution of the two distributions. This yields the basic relation

$$a_{uv,\chi} = \frac{a_{u,\chi}a_{v,\chi}}{\chi(1)},$$

from which it is easy to compute $a_{w,\chi}$ if w is a product of commutators and squares in disjoint variables. This is useful in the study of Fuchsian groups, see [53, 54].

Fourier coefficients of admissible words (see Section 4 above) were studied by Das and Nath [9] and by Parzanchevski and Schul [79]. This is useful in showing that admissible words are almost uniform (see Theorem 4.6 above). In [79] there is a formula to obtain the Fourier coefficients which

avoids summation over dismissible variables (those appearing once with exponent 1 and once with exponent -1) and over square variables (those appearing twice with exponent 1).

Character methods are also relevant in the context of studying powers of conjugacy classes and the Thompson Conjecture in particular. Here a main tool is the following classical result (see e.g. [88], §7.2, or [1], 10.1, p. 43).

Proposition 7.4. *Let G be a finite group, and let $C_1, \dots, C_k \subset G$ be conjugacy classes of G . For each element $g \in G$ let $M(g)$ be the number of k tuples (x_1, \dots, x_k) where $x_i \in C_i$ ($i = 1, \dots, k$) such that $x_1 \cdots x_k = g$. Then*

$$M(g) = \frac{|C_1| \cdots |C_k|}{|G|} \sum_{\chi \in \text{Irr } G} \frac{\chi(C_1) \cdots \chi(C_k) \chi(g^{-1})}{\chi(1)^{k-1}}.$$

Here we write $\chi(C_i)$ for the (common) value of $\chi(x)$ for $x \in C_i$.

Consequently we see that $g \in C^2$ if and only if

$$\sum_{\chi \in \text{Irr } G} \frac{\chi(C)^2 \chi(g^{-1})}{\chi(1)} \neq 0.$$

In particular Thompson Conjecture amounts to saying that any finite simple group G has a class C such that the character sum above is non-zero for all $g \in G$.

In general estimating this sum for all $g \in G$ is quite a formidable task. However, to show that $G \setminus \{1\} \subseteq C_1 C_2$ for classes C_1, C_2 , as in Theorem 6.6, it suffices to show that

$$\sum_{\chi \in \text{Irr } G} \frac{\chi(C_1) \chi(C_2) \chi(g^{-1})}{\chi(1)} \neq 0$$

for all $g \neq 1$, and the freedom in choosing the classes C_1, C_2 is very helpful. In the main case, where G is a classical group, we choose for C_1, C_2 classes of suitable regular semisimple elements $s_1, s_2 \in G$ lying in maximal tori $T_1, T_2 \subset G$. The tori T_i are chosen to be weakly orthogonal (see Section 2 of [41] for the precise definition), and this ensures that if χ is an irreducible character of G such that $\chi(s_1) \chi(s_2) \neq 0$ then χ is unipotent; moreover, there will be a small (in particular, bounded) number of such unipotent characters.

The choice of s_1, s_2 ensures that the number of non-zero summands in the character sum above is small. However, in order to control these

summands we need information on the character ratios $|\chi(g)|/\chi(1)$. Gluck's bounds (see for instance [20]) are useful but they do not suffice, and we have to establish sharper character bounds for elements of large support (see Section 4 of [41] for precise definition). We prove the following.

Theorem 7.5. *If G is a finite quasisimple classical group over \mathbb{F}_q and $g \in G$ is an element of support at least N , then*

$$|\chi(g)|/\chi(1) < q^{-\sqrt{N}/481}$$

for all $1 \neq \chi \in \text{Irr } G$.

This character theoretic result seems to be of independent interest, and may have further applications. This approach leads eventually to the proof of Theorem 6.6 above. It also plays a crucial role in the proof of Theorem 2.9, which is based on finding conjugacy classes $C_i \subset w_i(G)$ ($i = 1, 2$) such that $C_1 C_2$ covers all of G except the identity element (or sometimes elements of small support).

The proof of Theorem 6.9 above, that the mixing time with respect to a random class is 2, uses again character methods and the Witten zeta function ζ^G . It is intriguing that in this context the value of $\zeta^G(s)$ at $s = 2/3$ plays a key role. We show that if G ranges over any family of finite groups such that $\zeta^G(2/3) \rightarrow 1$, then for a random $x \in G$ the mixing time $T(x^G, G)$ is 2 with probability tending to 1.

Now, for most families of finite simple groups $\zeta^G(2/3) \rightarrow 1$ (see [55]), and the few remaining families (which are $L_2(q)$, $L_3(q)$ and $U_3(q)$) are dealt with by ad-hoc methods.

Our results on classes and word maps in symmetric and alternating groups are based on new sharp bounds on character values in symmetric groups obtained in the joint work [38] with Larsen. These bounds have the form $|\chi(\sigma)| \leq \chi(1)^{E(\sigma)+o(1)}$ for all $\chi \in \text{Irr } S_n$, where $0 \leq E(\sigma) \leq 1$ depends on the cycle structure of the permutation σ .

The detailed bound is quite technical, but we present here some of its main consequences.

Theorem 7.6. *Let $\sigma \in S_n$.*

(i) *If σ has at most n^α fixed points, then*

$$|\chi(\sigma)| \leq \chi(1)^{1/2+\alpha/2+o(1)} \quad \text{for all } \chi \in \text{Irr } S_n.$$

(ii) *If σ has at most $n^{o(1)}$ cycles of length $< m$ then*

$$|\chi(\sigma)| \leq \chi(1)^{1/m+o(1)} \quad \text{for all } \chi \in \text{Irr } S_n.$$

(iii) If σ has at most n^α cycles then

$$|\chi(\sigma)| \leq \chi(1)^{\alpha+o(1)} \quad \text{for all } \chi \in \text{Irr } S_n.$$

These bounds are essentially best possible. Part (ii) above is a far reaching generalization of a result of Fomin and Lulov [17]. Theorem 7.6 is very useful in bounding mixing times $T(C, G)$ of random walks. Indeed, if P_t is the distribution of this walk at time t , then by the upper bound lemma of Diaconis and Shahshahani [11] we have

$$\|P_t - U\|^2 \leq \sum_{1 \neq \chi \in \text{Irr } G} \frac{\chi(C)^{2t}}{\chi(1)^{2t-2}}.$$

Using Theorem 7.6 we can bound $|\chi(C)|$ by $\chi(1)^\alpha$ for appropriate α , which reduces the right hand side above to the value of the Witten zeta function ζ^G at a certain point s (depending on t). We can then find the minimal integer $t > 1$ such that $\zeta^G(s)$ tends to 0, and conclude that $T(C, G) \leq t$. This is how Theorem 6.8 is proved.

Theorem 7.6 also plays a major role in proving Theorem 2.9 for alternating groups.

Are there analogues of Theorem 7.6 for finite groups of Lie type? Such analogues could be extremely useful in solving various outstanding problems in such groups. Some results in this direction for semisimple elements were obtained some years ago in the preprint [5] with Bezrukavnikov and Liebeck. The general theorem is too technical to state here, but we illustrate it with the following special case.

Theorem 7.7. *Let $G = \text{GL}_n(q)$ and let $g \in G$ be a diagonal matrix with eigenvalues $\lambda_1, \dots, \lambda_m$ each occurring n/m times. Then we have*

$$|\chi(g)| \leq c\chi(1)^{1/m} \quad \text{for all } \chi \in \text{Irr } G,$$

where c is a constant depending only on n (and not on q).

This result could be regarded as a linear analogue of 7.6(ii) above and of the Fomin-Lulov bound [17] for S_n .

8. INFINITE GROUPS

In previous sections we discussed words and Waring type problems in finite groups. In this section we propose new challenges and discuss similar problems for infinite groups, such as p -adic groups, arithmetic groups and linear groups.

For background on word width in infinite groups, see [87].

Results on commutators in special linear groups over fields, and in Lie groups, were obtained long ago. For example, R.C. Thompson proved the following in [91, 92, 93].

Theorem 8.1. *Let $n \geq 2$ and let K be any (possibly infinite) field. If $n = 2$ suppose $|K| > 3$. Then every element of $\mathrm{SL}_n(K)$ is a commutator.*

Here the main difficulty was the case of small fields.

A similar result holds for semisimple Lie groups.

However, similar problems over various commutative rings (instead of fields) are still very much open, and we shall address them below.

Let us now move from commutators to powers $w = x^k$. What can be said about its image, namely the set of k th powers in various infinite groups? A classical result of Mal'cev [67] shows that $w(G)$ is very large, in the sense that it contains a finite index subgroup of G , provided G is virtually nilpotent and finitely generated. In [27] we study powers in linear groups, obtaining a converse to this result of Mal'cev.

Theorem 8.2. *Let G be a finitely generated linear group over any field K , and let $w = x^k$.*

- (i) *If $w(G)$ contains a finite index subgroup of G then G is virtually nilpotent.*
- (ii) *If G is covered by finitely many translates $g_i w(G)$ of $w(G)$ then G is virtually solvable.*

The proof involves strong approximation, Number Theory (the S -unit equation) as well as invariant measures on amenable groups.

It follows from part (i) that if the x^k -width of G is 1 then G is virtually nilpotent. See Theorem 2.14 and the comments following it for results of somewhat similar flavor in finite groups (which only hold for specific values of k). Part (ii) shows that finitely generated non-virtually solvable linear groups have few k th powers (namely $w(G)$ has “infinite index” in G). A stronger result is given in Theorem 8.15 below.

The behavior of the p th power map in pro- p groups is particularly interesting. Let $w = x^p$. If G is a finite powerful p -group then it is known

that $w(G)$ is a subgroup, namely G has x^p -width 1 (see [58]). Since p -adic analytic pro- p groups G are virtually powerful (see [59]) it follows that $w(G)$ contains a finite index (open) subgroup in this case. We conjecture that the converse also holds.

Conjecture 8.3. *Let G be a finitely generated pro- p group and let $w = x^p$. Then $w(G)$ contains an open subgroup of G if and only if G is p -adic analytic.*

The conjecture implies that if the set of p th powers of a finitely generated pro- p group G is a subgroup then G is p -adic analytic. Even this weaker version of the conjecture is very much open.

Word width in pro- p groups and p -adic analytic groups was studied by Jaikin-Zapirain [30], where the following two fundamental results are obtained.

Theorem 8.4. *A word $w \in F_d$ has finite width in all finitely generated pro- p groups if and only if $w \notin (F'_d)^p F''_d$.*

Here $(F'_d)^p$ is the subgroup of the commutator subgroup F'_d generated by all its p th powers.

Theorem 8.5. *Every word has finite width in every compact p -adic analytic group.*

In this result the width depends on the group and on the word. Can we, in some cases, bound the width by an absolute constant? With some luck, can this constant be 3, or even 2, as in Theorems 2.8 and 2.9 for large finite simple groups?

Word maps in p -adic (and adelic) groups were recently studied by Avni, Gelander, Kassabov and Shalev [2]. Suppose that G is a semisimple, simply connected, algebraic group over \mathbb{Q} , and consider the p -adic group $G(\mathbb{Z}_p)$. A typical example is $\mathrm{SL}_n(\mathbb{Z}_p)$. Such infinite profinite groups are extensions of an infinite pro- p group by a finite simple (or quasisimple) group, and they may share some properties with finite simple groups. We propose the following.

Problem 8.6. For a word w , study the image $w(G(\mathbb{Z}_p))$ and the w -width of $G(\mathbb{Z}_p)$.

It turns out that these questions may be studied using variations on Hensel Lemma, lifting solutions using the derivative dw of w . The following results from [2] shed some light on these questions.

Proposition 8.7. *For every non-trivial word $w \in F_d$, $w(G(\mathbb{Z}_p))$ contains a non-empty open subset of $G(\mathbb{Z}_p)$. In particular, $w(G(\mathbb{Z}_p))$ has positive Haar measure.*

Indeed, by Theorem 2.1 above, the map $w : G^d \rightarrow G$ is dominant. Over an algebraically closed field of characteristic 0, like $\overline{\mathbb{Q}_p}$, this is equivalent to the existence of a point for which the derivative dw of w is surjective (as a map of $\overline{\mathbb{Q}_p}$ vector spaces). Since the set of points in G^d for which dw is surjective is Zariski open and $G(\mathbb{Z}_p)$ is Zariski dense, there is a tuple $\vec{g} = (g_1, \dots, g_d) \in G(\mathbb{Z}_p)^d$ such that $dw|_{\vec{g}}$ is surjective. The Proposition now follows from the p -adic version of the open mapping theorem.

It is interesting that this result does not hold for local rings of positive characteristic: for example, the image of $SL_n(\mathbb{F}_p[[t]])$ under the map $x \mapsto x^p$ is contained in the set of all matrices all of whose eigenvalues are p th powers, which is a nowhere dense set.

The next result shows that the w -width of any given word in p -adic groups is at most 3, provided p is large enough.

Theorem 8.8. *For any non-trivial words w_1, w_2, w_3 there exists a number N depending only on these words, such that such if $p \geq N$ is a prime then*

$$w_1(G(\mathbb{Z}_p))w_2(G(\mathbb{Z}_p))w_3(G(\mathbb{Z}_p)) = G(\mathbb{Z}_p).$$

In particular $w(G(\mathbb{Z}_p))^3 = G(\mathbb{Z}_p)$ if $w \neq 1$ and $p \geq N(w)$.

This may be regarded as a p -adic analogue of Theorem 2.8 above. The condition on p is necessary, since for small p the word w may be an identity in some finite quotients of $G(\mathbb{Z}_p)$.

Can we strengthen Theorem 8.8 and obtain $w(G(\mathbb{Z}_p))^2 = G(\mathbb{Z}_p)$ in line of Theorem 2.9 for finite simple groups? It turns out that the answer is no, since such a conclusion fails for quasisimple groups (which may be quotients of the given p -adic group). However, our next result shows that $w(G(\mathbb{Z}_p))^2$ is very large.

Theorem 8.9. *For any non-trivial words w_1, w_2 there exists a number $N = N(w_1, w_2)$ such that if $p \geq N$ is a prime then*

$$w_1(G(\mathbb{Z}_p))w_2(G(\mathbb{Z}_p)) \supseteq G(\mathbb{Z}_p) \setminus Z(G(\mathbb{Z}_p))G^1(\mathbb{Z}_p).$$

Here $G^1(\mathbb{Z}_p)$ denotes the first congruence subgroup of $G(\mathbb{Z}_p)$. Theorem 8.9 shows that elements which are non-central modulo $G^1(\mathbb{Z}_p)$ must lie in $w(G(\mathbb{Z}_p))^2$ if $w \neq 1$ and $p \geq N(w)$.

We now turn from general words to the commutator word.

We propose to study the following p -adic version of Ore Conjecture.

Conjecture 8.10. *Suppose $n > 2$ or $p > 3$. Then all elements of $\mathrm{SL}_n(\mathbb{Z}_p)$ are commutators.*

A partial solution is given by the following result from [2].

Theorem 8.11.

- (i) *Every element of $\mathrm{SL}_n(\mathbb{Z}_p)$ whose image in $\mathrm{SL}_n(p)$ is not central is a commutator, provided $n \geq p + 2$.*
- (ii) *If n is a proper divisor of $p - 1$ then every element of $\mathrm{PSL}_n(\mathbb{Z}_p)$ is a commutator.*

A more general conjecture may be stated, dealing with other p -adic groups.

We also state another problem, in the spirit of Thompson Conjecture.

Problem 8.12. Study powers of conjugacy classes in p -adic groups $G(\mathbb{Z}_p)$. Can we find a conjugacy class C such that $C^2 = G(\mathbb{Z}_p)$?

We now switch to arithmetic groups, where much less is known. Such groups have the form $G(O_S)$, where G is a simply connected semisimple algebraic group over a number field K , S is a finite set of primes of K containing all the archimedean ones, and O_S is the ring of S -integers in K .

A main case to be considered is $\mathrm{SL}_n(\mathbb{Z})$. We propose the following intriguing question.

Problem 8.13. Is it true that all elements of $\mathrm{SL}_n(\mathbb{Z})$ ($n > 2$) are commutators?

We emphasize that, unlike Conjecture 8.10 above, we do not conjecture here that all elements are commutators. It may well be that the set of commutators in $\mathrm{SL}_n(\mathbb{Z})$ is not only a proper subset but even an “exponentially small” subset (see below and [62] for the precise definition). Problem 8.13 is a major test case reflecting the difficulty of deducing from local to global.

More generally, it would be interesting to find the commutator width of $\mathrm{SL}_n(\mathbb{Z})$; for large n it is at most 6, by [10].

The same problems may be considered for general arithmetic groups. Problem 8.13 is a special case of the following.

Problem 8.14. For a word w and an arithmetic group Γ study $w(\Gamma)$ and the w -width of Γ .

It is also interesting to study powers of conjugacy classes in arithmetic groups, and analogues of Thompson Conjecture.

A new main tool in the study of these problems is the group theoretic sieve method.

The basic idea, which appears in Kowalski [33], and in full generality in Lubotzky and Meiri [62], is as follows: We want to estimate “the size” of a subset Z of a finitely generated group Γ generated by a set S . We do this by looking at a sequence of finite index normal subgroups N_i of Γ . If these quotients are “sufficiently independent” and the family of Cayley graphs $\text{Cay}(\Gamma/N_i; S)$ form a family of expanders, and if the image of Z in all (or in sufficiently many) of them is of size less than $(1 - \delta)|\Gamma/N_i|$ for some fixed $\delta > 0$, then the size of Z is *exponentially small*. This means that a random walk on $\text{Cay}(\Gamma; S)$ meets Z at step n with probability $\leq \alpha^n$ for some fixed $\alpha < 1$.

The wealth of recent new results on approximate subgroups, expanders and property τ enables one to apply the sieve method in some situations, for arithmetic and linear groups.

In [62] Lubotzky and Meiri develop and apply sieve methods to prove the following.

Theorem 8.15. *Let Γ be a finitely generated linear group over \mathbb{C} . Suppose Γ is not virtually solvable. Then the set of all proper powers in Γ is exponentially small.*

This strengthens part (ii) of Theorem 8.2 above in two ways: we deal with all proper powers simultaneously, and we show that not only they have infinite index, they are exponentially small. However, while Theorem 8.2 holds for arbitrary fields, in Theorem 8.15 the case of linear groups in positive characteristic is not yet known.

Problem 8.16.

- (i) Does Theorem 8.15 hold for finitely generated linear groups over any field?
- (ii) Are there analogues of it for other words w instead of power words?

More specifically, for arbitrary words w , it would be interesting to measure the size of $w(\Gamma)$ for arithmetic and linear groups, and to find out when it is exponentially small.

As for word width, in lattices in simple groups of rank 1 one may expect infinite width of various words, while in lattices in higher rank the word width might be finite.

We conclude this paper with a provocative question on the surjectivity of arbitrary word maps on certain algebraic groups. Note that, by Corollary 2.2, if w is any non-identity word, then $w(\text{SL}_n(\mathbb{C}))^2 = \text{SL}_n(\mathbb{C})$. However,

w need not be surjective in this case. There are non-identity words which are non-surjective on $\mathrm{SL}_2(\mathbb{C})$, indeed $w = x^2$ is such a word (see [6] and the references therein).

Problem 8.17. Is every word $w \neq 1$ surjective on $\mathrm{PSL}_n(\mathbb{C})$?

REFERENCES

- [1] Z. Arad and M. Herzog (Eds), *Products of Conjugacy Classes in Groups*, Springer Lecture Notes **1112**, Springer-Verlag, Berlin, 1985.
- [2] N. Avni, T. Gelander, M. Kassabov and A. Shalev, Word values in p -adic and adelic groups, arXiv:math/1303.1161.
- [3] T. Bandman, S. Garion and F. Grunewald, On the surjectivity of Engel words on $\mathrm{PSL}(2, q)$, *Groups, Geom. Dyn.* **6** (2012), 409–439.
- [4] T. Bandman and S. Garion, Surjectivity and equidistribution of the word $x^a y^b$ on $\mathrm{PSL}_2(q)$ and $\mathrm{SL}_2(q)$, *Internat. J. Algebra Comput.* **22** (2012), 33pp.
- [5] R. Bezrukavnikov, M.W. Liebeck and A. Shalev, Character bounds, random walks and covering in finite Chevalley groups, Preprint 2005.
- [6] A. Borel, On free subgroups of semisimple groups, *Enseign. Math.* **29** (1983), 151–164.
- [7] E. Breuillard, B. Green and T. Tao, Approximate subgroups of linear groups, *Geom. Funct. Anal.* **21** (2011), 774–819.
- [8] J. Bourgain and A. Gamburd, Uniform expansion bounds for Cayley graphs of $\mathrm{SL}_2(\mathbb{F}_p)$, *Ann. of Math.* **167** (2008), 625–642.
- [9] A. K. Das and R. K. Nath, On solutions of a class of equations in a finite group, *Comm. in Algebra* **37** (2009), 3904–3911.
- [10] R. K. Dennis and L. N. Vaserstein, On a question of M. Newman on the number of commutators, *J. Algebra* **118** (1988), 150–161.
- [11] P. Diaconis and M. Shahshahani, Generating a random permutation with random transpositions, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **57** (1981), 159–179.
- [12] J. D. Dixon, L. Pyber, Á. Seress, A. Shalev, Residual properties of free groups and probabilistic methods, *J. reine angew. Math. (Crelle's)* **556** (2003), 159–172.
- [13] P. Erdős and P. Turán, On some problems of a statistical group theory. I, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **4** (1965), 175–186.
- [14] P. Erdős and P. Turán, On some problems of a statistical group theory. II, *Acta Math. Acad. Sci. Hungaricae* **18** (1967), 151–163.
- [15] E. W. Ellers and N. Gordeev, On conjectures of J. Thompson and O. Ore, *Trans. Amer. Math. Soc.* **350**, 3657–3671.
- [16] E. W. Ellers, N. Gordeev and M. Herzog, Covering numbers for Chevalley groups, *Israel J. Math.* **111** (1999), 339–372.
- [17] S. V. Fomin and N. Lulov, On the number of rim hook tableaux, *J. Math. Sci. (New York)* **87** (1997), 4118–4123.

- [18] S. Garion and A. Shalev, Commutator maps, measure preservation, and T -systems, *Trans. Amer. Math. Soc.* **361** (2009), 4631–4651.
- [19] N. Gill, L. Pyber, I. Short and E. Szabó, On the product decomposition conjecture for finite simple groups, *Groups, Geom. Dyn.*, to appear.
- [20] D. Gluck, Sharper character value estimates for groups of Lie type, *J. Algebra* **174** (1995), 229–266.
- [21] W. T. Gowers, Quasirandom groups, *Combin. Probab. Comput.* **17** (2008), 363–387.
- [22] R. Guralnick and G. Malle, Products of conjugacy classes and fixed point spaces, *J. Amer. Math. Soc.* **25** (2012), 77–121.
- [23] R. M. Guralnick and I. Pak, On a question of B.H. Neumann, *Proc. Amer. Math. Soc.* **131** (2002), 2021–2025.
- [24] R. Guralnick and Ph. Tiep, The Waring problem for finite quasisimple groups. II, arXiv:math/1302.0333.
- [25] H. A. Helfgott, Growth and generation in $SL_2(\mathbb{Z}/p\mathbb{Z})$, *Ann. of Math.* **167** (2008), 601–623.
- [26] E. Hrushovski, Stable group theory and approximate subgroups, *J. Amer. Math. Soc.* **25** (2012), 189–243.
- [27] E. Hrushovski, P. H. Kropholler, A. Lubotzky and A. Shalev, Powers in finitely generated groups, *Trans. Amer. Math. Soc.* **348** (1996), 291–304.
- [28] I. M. Isaacs, *Character Theory of Finite Groups*, Academic Press, New York – San Francisco – London, 1976.
- [29] S. Jambor, M. W. Liebeck and E. A. O’Brien, Some word maps that are non-surjective on infinitely many finite simple groups, *Bull. London Math. Soc.*, to appear.
- [30] A. Jaikin-Zapirain, On the verbal width of finitely generated pro- p groups, *Rev. Mat. Iberoam.* **24** (2008), 617–630.
- [31] G. A. Jones, Varieties and simple groups, *J. Austr. Math. Soc.* **17** (1974), 163–173.
- [32] M. Kassabov and N. Nikolov, Words with few values in finite simple groups, arXiv:math/1112.5484.
- [33] E. Kowalski, *The large sieve and its applications*, Arithmetic geometry, random walks and discrete groups. Cambridge Tracts in Mathematics **175**, Cambridge University Press, Cambridge, 2008. xxii+293 pp.
- [34] M. Larsen, How often is a partition an n 'th power?, arXiv:math.CO/9712223.
- [35] M. Larsen, Word maps have large image, *Israel J. Math.* **139** (2004), 149–156.
- [36] M. Larsen and A. Lubotzky, Representation growth for linear groups, *J. Europ. Math. Soc.* **10** (2008), 351–390.
- [37] M. Larsen and A. Shalev, Word maps and Waring type problems, *J. Amer. Math. Soc.* **22** (2009), 437–466.
- [38] M. Larsen and A. Shalev, Characters of symmetric groups: sharp bounds and applications, *Invent. Math.* **174** (2008), 645–687.
- [39] M. Larsen and A. Shalev, Fibers of word maps and some applications, *J. Algebra* **354** (2012), 36–48.

- [40] M. Larsen and A. Shalev, Character estimates and almost measure preserving word maps, Preprint 2012.
- [41] M. Larsen, A. Shalev and Ph. Tiep, The Waring problem for finite simple groups, *Ann. of Math.* **174** (2011), 1885–1950.
- [42] M. Larsen, A. Shalev and Ph. Tiep, Waring problem for finite quasisimple groups, *IMRN* rns109 (2012), 26 pages.
- [43] R. Lawther and M. W. Liebeck, On the diameter of a Cayley graph of a simple group of Lie type based on a conjugacy class, *J. Comb. Theory, Ser. A* **83** (1998), 118–137.
- [44] M. W. Liebeck, N. Nikolov and A. Shalev, Groups of Lie type as products of SL_2 subgroups, *J. Algebra* **326** (2011), 201–207.
- [45] M. W. Liebeck, N. Nikolov and A. Shalev, A conjecture on product decompositions in simple groups, *Groups, Geom. Dyn.* **4** (2010) (Magnus Issue), 799–812.
- [46] M. W. Liebeck, N. Nikolov and A. Shalev, Product decompositions in finite simple groups, *Bull. London Math. Soc.* **44** (2012), 469–472.
- [47] M. W. Liebeck, E. A. O’Brien, A. Shalev and Ph. Tiep, The Ore Conjecture, *J. Europ. Math. Soc.* **12** (2010), 939–1008.
- [48] M. W. Liebeck, E. A. O’Brien, A. Shalev and Ph. Tiep, Commutators in finite quasisimple groups, *Bull. London Math. Soc.* **43** (2011), 1079–1092.
- [49] M. W. Liebeck, E. A. O’Brien, A. Shalev and Ph. Tiep, Products of squares in simple groups, *Proc. Amer. Math. Soc.* **140** (2012), 21–33.
- [50] M. W. Liebeck and L. Pyber, Finite linear groups and bounded generation, *Duke Math. J.* **107** (2001), 159–171.
- [51] M. W. Liebeck and A. Shalev, The probability of generating a finite simple group, *Geom. Ded.* **56** (1995), 103–113.
- [52] M. W. Liebeck and A. Shalev, Diameters of finite simple groups: sharp bounds and applications, *Ann. of Math.* **154** (2001), 383–406.
- [53] M. W. Liebeck and A. Shalev, Fuchsian groups, coverings of Riemann surfaces, subgroup growth, random quotients and random walks, *J. Algebra* **276** (2004), 552–601.
- [54] M. W. Liebeck and A. Shalev, Fuchsian groups, finite simple groups, and representation varieties, *Invent. Math.* **159** (2005), 317–367.
- [55] M. W. Liebeck and A. Shalev, Character degrees and random walks in finite groups of Lie type, *Proc. London. Math. Soc.* **90** (2005), 61–86.
- [56] M. W. Liebeck and A. Shalev, Powers in finite groups and a criterion for solubility, *Proc. Amer. Math. Soc.*, to appear.
- [57] M. W. Liebeck and A. Shalev, Power sets and soluble subgroups, *Proc. Amer. Math. Soc.*, to appear.
- [58] A. Lubotzky and A. Mann, Powerful p -groups. I. Finite groups, *J. Algebra* **105** (1987), 484–505.
- [59] A. Lubotzky and A. Mann, Powerful p -groups. II. p -adic analytic groups, *J. Algebra* **105** (1987), 506–515.

- [60] A. Lubotzky, Finite simple groups of Lie type as expanders, *J. Europ. Math. Soc.* **13** (2011), 1331–1341.
- [61] A. Lubotzky, Images of word maps on finite simple groups, arXiv:math/1211.6575.
- [62] A. Lubotzky and C. Meiri, Sieve methods in group theory: I. powers in linear groups, *J. Amer. Math. Soc.* **25** (2012), 1119–1148.
- [63] A. Lubotzky and D. Segal, *Subgroup Growth*, Progress in Mathematics **212**, Birkhäuser, Basel, 2003.
- [64] N. Lulov, Random walks on symmetric groups generated by conjugacy classes, Ph.D. Thesis, Harvard University, 1996.
- [65] N. Lulov and I. Pak, Rapidly mixing random walks and bounds on characters of the symmetric groups, *J. Algebraic Combin.* **16** (2002), 151–163.
- [66] G. Lusztig, *Characters of reductive groups over finite fields*, Ann. of Math. Studies, Princeton University Press, Princeton, 1984.
- [67] A. I. Mal'cev, Homomorphisms onto finite groups, *Ivanov. Gos. Ped. Inst. Ucen. Zap. Fiz.-Mat. Nauki* **8** (1958), 49–60.
- [68] C. Martinez and E. I. Zelmanov, Products of powers in finite simple groups, *Israel J. Math.* **96** (1996), 469–479.
- [69] M. B. Nathanson, *Additive Number Theory: the classical bases*, Graduate Texts in Mathematics **164**, Springer, 1996.
- [70] A. Nica, On the number of cycles of given length of a free word in several random permutations, *Random Structures Algorithms* **5** (1994), 703–730.
- [71] N. Nikolov, On the commutator width of perfect groups, *Bull. London Math. Soc.* **36** (2004), 30–36.
- [72] N. Nikolov, A product decomposition for the classical quasisimple groups, *J. Group Theory* **10** (2007), 43–53.
- [73] N. Nikolov and D. Segal, On finitely generated profinite groups, I: strong completeness and uniform bounds, *Ann. of Math.* **165** (2007), 171–238.
- [74] N. Nikolov and D. Segal, On finitely generated profinite groups, II: product decompositions of quasisimple groups, *Ann. of Math.* **165** (2007), 239–273.
- [75] N. Nikolov and D. Segal, Powers in finite groups, *Groups Geom. Dyn.* **5** (2011), 501–507.
- [76] N. Nikolov and D. Segal, Generators and commutators in finite groups; abstract quotients of compact groups, *Invent. Math.* **190** (2012), 513–602.
- [77] N. Nikolov and L. Pyber, Product decompositions of quasirandom groups and a Jordan type theorem, *J. Europ. Math. Soc.* **13** (2011), 1063–1077.
- [78] O. O. Ore, Some remarks on commutators, *Proc. Amer. Soc.* **272** (1951), 307–314.
- [79] O. Parzanchevski and G. Schul, On the Fourier expansion of word maps, arXiv:math/1207.0453.
- [80] D. Puder, Primitive words, free factors and measure preservation, *Israel J. Math.*, to appear.
- [81] D. Puder and O. Parzanchevski, Measure preserving words are primitive, arXiv:math/1202.3269.

- [82] L. Pyber and E. Szabó, Growth in finite simple groups of Lie type, arXiv:math/1001.4556.
- [83] J. Saxl and J. S. Wilson, A note on powers in simple groups, *Math. Proc. Camb. Phil. Soc.* **122** (1997), 91–94.
- [84] G. Schul, Expansion in finite simple groups, Preprint.
- [85] G. Schul and A. Shalev, Words and mixing times in finite simple groups, *Groups, Geom. Dyn.* **5** (2011), 509–527.
- [86] A. Salehi-Golsefidy and P. Varjú, Expansion in perfect groups, *Geom. Funct. Anal.* **22** (2012), 1832–1891.
- [87] D. Segal, *Words: notes on verbal width in groups*, London Math. Soc. Lecture Note Series **361**, Cambridge University Press, Cambridge, 2009.
- [88] J.-P. Serre, *Topics in Galois Theory*, Research notes in math **1**, Jones and Bartlett, Boston-London, 1992.
- [89] A. Shalev, Word maps, conjugacy classes, and a non-commutative Waring-type theorem, *Ann. of Math.* **170** (2009), 1383–1416.
- [90] A. Shalev, Mixing and generation in simple groups, *J. Algebra* **319** (2008), 3075–3086.
- [91] R. C. Thompson, Commutators in the special and general linear groups, *Trans. Amer. Math. Soc.* **101** (1961), 16–33.
- [92] R. C. Thompson, On matrix commutators, *Portugal. Math.* **21** (1962), 143–153.
- [93] R. C. Thompson, Commutators of matrices with coefficients from the field of two elements, *Duke Math. J.* **29** (1962), 367–373.
- [94] U. Vishne, Mixing and covering in symmetric groups, *J. Algebra* **205** (1998), 119–140.
- [95] J. S. Wilson, First-order group theory, in *Infinite Groups 1994*, de Gruyter, Berlin, 1996, pp. 301–314.
- [96] H. S. Wilf, The asymptotics of $e^{P(z)}$ and the number of elements of each order in S_n , *Bull. Amer. Math. Soc.* **15** (1986), 228–232.
- [97] E. Witten, On quantum gauge theories in two dimensions, *Comm. Math. Phys.* **141** (1991), 153–209.
- [98] E. I. Zelmanov, Solution of the restricted Burnside problem for groups of odd exponent, (Russian) *Izv. Akad. SSSR Ser. Mat.* **54** (1990), 42–59; translation in *Math USSR-Izv.* **36** (1990), 41–60.
- [99] E. I. Zelmanov, Solution of the restricted Burnside problem for 2-groups, (Russian) *Mat. Sb.* **182** (1991), 568–592; translation in *Math USSR-Sb* **72** (1992), 543–565.

Aner Shalev

*Institute of Mathematics,
The Hebrew University,
Jerusalem 91904,
Israel*

e-mail: shalev@math.huji.ac.il

SOME OF ERDŐS' UNCONVENTIONAL PROBLEMS IN NUMBER THEORY, THIRTY-FOUR YEARS LATER

GÉRALD TENENBAUM

There are many ways to recall Paul Erdős' memory and his special way of doing mathematics. Ernst Straus described him as “the prince of problem solvers and the absolute monarch of problem posers”. Indeed, those mathematicians who are old enough to have attended some of his lectures will remember that, after his talks, chairmen used to slightly depart from standard conduct, not asking if there were any questions but if there were any answers.

In the address that he forwarded to Miklós Simonovits for Erdős' funeral, Claude Berge mentions a conversation he had with Paul in the gardens of the Luminy Campus, near Marseilles, in September 1995. After Paul's opening lecture for this symposium on Combinatorics, Berge asked him to specify his beauty criteria for a conjecture in discrete mathematics. Erdős mainly retained the following five:

- (i) The *simplicity* of the statement;
- (ii) The expected *difficulty* of the solution (which Paul liked to measure in dollars);
- (iii) The *posterity* of the subsequent theorem, i.e. the set of results arising either directly from the solution or from the methods designed to obtain it;
- (iv) The *future* of the path opened by the problem, which I would rather call the set of *descendants* of the problem, in other words the family of new questions opened up by the statement or the solution of the conjecture;
- (v) The *intuitive representability* of the specific mathematical property that is being dealt with.

Apart, perhaps, the last, for which an adequate transposition should be described with further precision, these criteria are equally relevant to a classification for a conjecture in analytic and/or elementary number theory.

My purpose here mainly consists in illustrating these criteria by revisiting some of the problems stated by Erdős in his profound article [24].

Aside from updating the status of a number of interesting questions, my hope is to convince the reader that Erdős' conjectures, although stated in a condensed and seemingly particular form, were problematic rather than problems. Day after day, year after year, each of his questions appears, in the light of discussions and partial progress, as a node in a gigantic net, designed not for a single prey but for a whole species.

In the sequel of this paper, quotes from the article [24] are set in italics. I took liberties to correct obvious typographic errors and to slightly modify some notations in order to fit with subsequent works. Erdős' paper starts with the following.

First of all I state a very old conjecture of mine: the density of integers n which have two divisors d_1 and d_2 satisfying $d_1 < d_2 < 2d_1$ is 1. I proved long ago [20] that the density of these numbers exists but I have never been able to prove that it is 1. I claimed [21] that I proved that almost all integers n have two divisors

$$(1) \quad d_1 < d_2 < d_1 \{ 1 + (e/3)^{(1-\eta) \log \log n} \}$$

and that (1) is best possible, namely it fails if $1 - \eta$ is replaced by $1 + \eta$. R. R. Hall and I confirmed this later statement but unfortunately we cannot prove (1). We are fairly sure that (1) is true and perhaps it is not hopeless to prove it by methods of probabilistic number theory that are at our disposal.

This is an edifying example of a conjecture meeting the above five requirements. However, before elaborating on this, it may be worthwhile try understanding the process that led Erdős to this simple and deep statement.

An integer n is called *perfect* if it is equal to the sum of its proper divisors. Thus $6 = 1 + 2 + 3$ and $28 = 1 + 2 + 4 + 7 + 14$ are perfect. In modern notation, a perfect integer n satisfies $\sigma(n) = 2n$ where $\sigma(n)$ stands for the sum of all divisors. This is an interesting formulation since $\sigma(n)$ is a multiplicative function of n . In the third century before our era, Euclid proved (IX.36) that $2^{p-1}(2^p - 1)$ is perfect whenever $2^p - 1$ is prime, which of course implies that p itself is prime.

An integer n is called *abundant* if $\sigma(n) > 2n$. In the early thirties, in a book on number theory, Erich Bessel-Hagen asks whether abundant integers have a natural density. Davenport [12], Chowla [11], Erdős [16] and Behrend [3] all gave, independently, a positive answer. All proofs, except that of Erdős, rest on the method of (real or complex) moments. Erdős attacks the problem from another viewpoint: primitive abundant numbers, i.e. abundant numbers having no abundant proper divisor. Writing $f(n)$ for $\sigma(n)/n$, any primitive abundant integer n satisfies

$$2 \leq f(n) \leq f(n/p)f(p) < 2(1 + 1/p)$$

whenever $p|n$. Since the largest prime factor of n is usually large, this restricts the cardinality of primitive abundant numbers not exceeding x , which can be shown to be $o(x/(\log x)^2)$. The proof is then completed by noticing that, if we write

$$\mathcal{M}(\mathcal{A}) := \{ma : a \in \mathcal{A}, m \geq 1\}$$

for the so-called *set of multiples* of the set \mathcal{A} and $d, \bar{d}, \underline{d}$ for natural, upper and lower density respectively, then

$$d\mathcal{M}(\mathcal{A}_T) \leq \underline{d}\mathcal{M}(\mathcal{A}) \leq \bar{d}\mathcal{M}(\mathcal{A}) \leq d\mathcal{M}(\mathcal{A}_T) + \sum_{\substack{a>T \\ a \in \mathcal{A}}} \frac{1}{a}$$

holds for any integer sequence \mathcal{A} such that $\sum_{a \in \mathcal{A}} 1/a < \infty$, with $\mathcal{A}_T := \mathcal{A} \cap [1, T]$.

This was the starting point of the fruitful concept of set of multiples.

It was once suspected that any set of multiples should have a natural density. However, Besicovitch [5] soon disproved this conjecture by showing that

$$(2) \quad \liminf_{T \rightarrow \infty} d\mathcal{M}([T, 2T]) = 0.$$

Indeed, it is easy to deduce from this that, given any $\varepsilon > 0$ and a sequence $\{T_j\}_{j=0}^\infty$ increasing sufficiently fast, then $\mathcal{A} := \cup_j]T_j, 2T_j]$ satisfies $\underline{d}\mathcal{M}(\mathcal{A}) < \varepsilon, \bar{d}\mathcal{M}(\mathcal{A}) \geq \frac{1}{2}$.

The reader might ask at this stage: interesting indeed, but how does this link to (1)? We still need a few more steps inside Erdős' peculiar way of thinking.

It is one of the marks of the great: not to accept an obstruction before understanding it completely. This holds outside of mathematics as well as inside. Erdős did not accept Besicovitch's counter-example for itself and continued the quest.

First [18], he improved (2) to the optimal

$$(3) \quad \lim_{T \rightarrow \infty} d\mathcal{M}([T, T^{1+\varepsilon_T}]) = 0$$

provided $\varepsilon_T \rightarrow 0$ as $T \rightarrow \infty$.

With this new, crucial piece of information, he progressed in two connected directions: first, to show, with Davenport [13]—see also [14] for

another, very interesting proof—that any set of multiples has a logarithmic density, equal to its lower asymptotic density,¹ and, second, to show [20]² that Besicovitch-type constructions are essentially the only obstacles to the existence of $dM(\mathcal{A})$: writing $d_1(n, \mathcal{A}) := \inf\{d|n : d \in \mathcal{A}\}$ with the convention that $d_1(n, \mathcal{A}) = \infty$ whenever $n \notin M(\mathcal{A})$, a necessary and sufficient condition that $M(\mathcal{A})$ has a natural density is

$$(4) \quad \lim_{\varepsilon \rightarrow 0} \overline{d}\{n \geq 1 : n^{1-\varepsilon} < d_1(n, \mathcal{A}) \leq n\} = 0.$$

Now, consider the set

$$(5) \quad \mathcal{E} := \{m \in \mathbb{N}^* : m = dd', d < d' < 2d\}.$$

Then the double inequality $n^{1-\varepsilon} < d_1(n, \mathcal{E}) \leq n$ plainly implies that n has a divisor in $]n^{1/2-\varepsilon}, n^{1/2}]$ so it is easy to deduce (4) from (3).

So we now know that the set of integers with two close divisors has a natural density. (By ‘close’ we mean here that the ratio of the two divisors should lie in $]1, 2[$.) Moreover, as seen above, the existence property follows in a natural way from the theory of sets of multiples: the sequence \mathcal{E} defined above is one of simplest examples one can think of that meets the criterion (4).

But what should the density be? Erdős stated, as early as 1948 (and probably much before) [20], that this density should be equal to 1. Here again, a seemingly anecdotal conjecture is actually based on a profound assumption—any answer to it, positive or negative, is bound to enlighten our understanding of the multiplicative structure of integers.

Let us make the convention to use the suffix pp to indicate that a relation holds on a set of asymptotic density 1. As we shall see later in this paper, Erdős had known for long that sufficiently far prime factors behave almost independently pp. Specifically, if we denote by

$$(6) \quad \{p_j(n)\}_{j=1}^{\omega(n)}$$

the increasing sequence of distinct prime factors of an integer n and if we write

$$(7) \quad U_j(n) := \{\log_2 p_j(n) - j\} / \sqrt{j},$$

then, to a first approximation, $U_j(n)$ and $U_h(n)$ resemble independent Gaussian random variables pp provided that $j/h \rightarrow \infty$. (Here and in the sequel, we let \log_k denote the k -fold iterated logarithm.) Having this in

¹We shall make use of this extra information later on.

²See [29] for a short proof.

mind, it is reasonable to believe that, in first approximation, the quantities $\log(d'/d)$ are evenly distributed pp in the interval $[-\log n, \log n]$. Since these quantities are $3^{\omega(n)}$ in number, we deduce from the Hardy–Ramanujan estimate $\omega(n) \sim \log_2 n$ pp that the smallest of these numbers should be of size $(\log n)^{1-\log 3+o(1)}$ pp.

This is, perhaps no more, certainly no less, what is hidden behind conjecture (1).

This conjecture, which is now a theorem, due to Erdős–Hall [27] for the lower bound and to Maier–Tenenbaum [55] for the upper bound, has had a wide posterity and many descendants.

In his doctoral dissertation supervised by the author [65], Stef proves that the number R_x of exceptional integers not exceeding x and which do not belong to $\mathcal{M}(\mathcal{E})$ satisfies

$$(8) \quad x/(\log x)^{\beta+o(1)} \ll R_x \ll xe^{-c\sqrt{\log_2 x}}$$

for a suitable constant $c > 0$, with $\beta = 1 - (1 + \log_2 3)/\log 3 \approx 0,00415$. These are the best known estimates to date.

To the chapter of posterity certainly belong all results involving the still mysterious Erdős–Hooley Delta-function and the so-called propinquity functions

$$E_r(n) := \min_{1 \leq j \leq \tau(n)-r} \log\{d_{j+r}(n)/d_j(n)\} \quad (r \geq 1),$$

where $\{d_j(n)\}_{j=1}^{\tau(n)}$ stands for the increasing sequence of the divisors of an integer n .

One of the most recent achievements in this direction is a very precise confirmation of the heuristic principle leading to (1), as described above: Raouj, Stef and myself prove in [62] that

$$E_1(n) = \frac{\log n}{3^{\omega(n)}} (\log_2 n)^{\vartheta_n} \quad \text{pp,}$$

where $-5 \leq \vartheta_n \leq 10$. Many more precise and connected results are actually proved in [62].

The situation is much less satisfactory regarding the functions E_r when $r \geq 2$, for which the precise pp behaviour is still unknown. Using techniques similar to that of the proof of theorem 3 of [36], it can be shown that

$$E_2(n) > (\log n)^{-\gamma_2+o(1)} \quad \text{pp}$$

for some $\gamma_2 < \log 3 - 1$. Moreover, the methods and results of [56] yield

$$E_r(n) \leq (\log n)^{-\beta_r + o(1)} \quad \text{pp},$$

with

$$\beta_r := \frac{(\log 3 - 1)^m}{(3 \log 3 - 1)^{m-1}}, \quad 2^{m-1} < r + 1 \leq 2^m.$$

Thus, we have

$$\beta_1 = \log 3 - 1 \approx 0.09861, \quad \beta_2 = \beta_3 \approx 0.00423, \quad \beta_r \approx 0.00018 \quad (4 \leq r \leq 7).$$

Also, it is proved in [56] (th. 1.1) that $E_r(n) > \tau(n)^{-1/r + o(1)}$ holds pp uniformly in $r \geq 1$, and thus

$$E_r(n) = 1/(\log n)^{o(1)} \quad \text{pp} \quad (r = r(n) \rightarrow \infty),$$

a result which might look surprising at first sight.

We conjecture the existence of a strictly decreasing sequence $\{\alpha_r\}_{r=1}^\infty$ such that

$$E_r(n) = (\log n)^{-\alpha_r + o(1)} \quad \text{pp}.$$

It is particularly irritating, for instance, to be unable to find a better pp upper bound for $E_2(n)$ than for $E_3(n)$.

We also mention as a posterity result the proof by Raouj [61] of Erdős' conjecture asserting that

$$d\mathcal{M}(\cup_{d|n}]d, 2d]) = 1 + o(1) \quad \text{pp}.$$

This is established in the following fairly strong (and optimal) form. Put $\lambda^* := \log 4 - 1$ and $\delta_n := d\mathcal{M}(\cup_{d|n}]d, (1 + 1/(\log n)^\lambda)d])$. Then

$$\frac{1}{(\log n)^{F(\lambda) + o(1)}} < 1 - \delta_n < e^{-c_\lambda \sqrt{\log n}} \quad (0 \leq \lambda < \lambda^*) \quad \text{pp},$$

$$\delta_n = (\log n)^{-F(\lambda) + o(1)} \quad (\lambda > \lambda^*)$$

where $F(\lambda) := \beta \log \beta - \beta + 1$ with $\beta := -1 + (1 + \lambda)/\log 2$ if $\lambda \leq 3 \log 2 - 1$, and $F(\lambda) := \lambda - \log 2$ if $\lambda > 3 \log 2 - 1$.

The Erdős–Hooley function is defined as

$$\Delta(n) := \sup_{u \in \mathbb{R}} \sum_{\substack{d|n \\ e^u < d \leq e^{u+1}}} 1 \quad (n \geq 1).$$

It first appears (implicitly) in [23] and (explicitly) in [30], [31] in the early seventies. It was next studied by Hooley [50] with the aim of developing a variety of applications to several branches of number theory.

The ratio $\Delta(n)/\tau(n)$ has an immediate probabilistic interpretation: with Lévy's 1937 definition, it is the value at 1 of the concentration function of the random variable D_n taking the values $\log d$ ($d|n$) with uniform probability $1/\tau(n)$. It is noteworthy to state here that $D_n = \sum_{p^\nu || n} D_{p^\nu}$ where the D_{p^ν} are independent.

If we replace the factor 2 by e , which is irrelevant to all intents and purposes, Erdős' initial conjecture

$$(9) \quad dM(\mathcal{E}) = 1$$

is equivalent to the statement that

$$(10) \quad \Delta(n) > 1 \quad \text{pp,}$$

so that (8) provides quantitative estimates for the number of exceptions.

The best pp-bounds to date for the Δ -function appear in a joint article with Maier [56]. We prove that

$$(\log_2 n)^{\gamma+o(1)} < \Delta(n) < (\log_2 n)^{\log 2+o(1)} \quad \text{pp,}$$

where the exponent $\gamma := (\log 2)/\log\left(\frac{1-1/\log 27}{1-1/\log 3}\right) \approx 0.33827$ is conjectured to be optimal.

To show the existence and determine the value of the exact exponent is a challenging problem in probabilistic number theory. There is no doubt that such a result would imply deeper ideas on the structure of the set of divisors of a normal integer.

However, as shown by Hooley in [50], it is mainly information on the average order

$$s(x) := \frac{1}{x} \sum_{n \leq x} \Delta(n)$$

that has applications to other arithmetical topics such as Waring-type problems [75], Diophantine approximation [50], [69], and Chebyshev's problem on the greatest prime factor of polynomial sequences [71]. It is thus proved in [71], as a consequence of an average estimate for a variant of $s(x)$, that, for any $\alpha < 2 - \log 4 \approx 0.61370$, the bound

$$P^+ \left(\prod_{n \leq x} F(n) \right) > x e^{(\log x)^\alpha} \quad (x > x_0(F))$$

holds for any irreducible polynomial $F(X) \in \mathbb{Z}[X]$ with degree > 1 . This is currently the best available result valid for polynomials of arbitrary degree. Here and in the sequel $P^+(m)$ denotes the largest prime factor of the integer m with the convention that $P^+(1) = 1$.

Established in [44] and [68], the best bounds for $s(x)$ at the time of writing are

$$(11) \quad \log_2 x \ll s(x) \ll e^{c\sqrt{\log_2 x \log_3 x}} \quad (x \rightarrow \infty)$$

where c is a suitable constant. See [46], [69] and, for instance, [63] for further references and descriptions on this question.

Still in the area of descendants of the conjecture (1), we mention the recent paper [8] in collaboration with La Bretèche and where sharp, weighted average bounds are given for functions of the type

$$(12) \quad \Delta(n, f) := \sup_{u \in \mathbb{R}, 0 \leq v \leq 1} \left| \sum_{d|n, e^u < d \leq e^{u+v}} f(d) \right|$$

where f is an oscillating function, typical cases being those of a non principal Dirichlet character or of the Möbius function. All suitably weighted finite integral, even moments are also studied. This is the key step to the proof, given in [10], of Manin’s conjecture, in the strong form conjectured by Peyre and with effective remainder term, for all Châtelet surfaces.

Maier established in [53] normal upper and lower bounds for (12) in the case $f = \mu$, the Möbius function, and his method is equally applicable in the case $f = \chi$, a real, non principal Dirichlet character.

Short averages have also been investigated, by Nair–Tenenbaum [57], Henriot [48], and La Bretèche–Tenenbaum [9]. These may have numerous, sometimes surprising applications. For instance, writing $\langle t \rangle$ for the fractional part of a real number t , we have [57], for any given $\varepsilon > 0$,

$$\sup_{D \geq 1} \left| \sum_{D \leq d \leq 2D} \left\langle \frac{x+y}{d} \right\rangle - \left\langle \frac{x}{d} \right\rangle \right| \ll y(\log x)^{o(1)} \quad (x^\varepsilon \leq y \leq x),$$

a bound which known exponential sums methods, by far, will fail to meet.

This ends our comments and update on conjecture (1).

The next problem in [24] is described as follows.

Denote by $\tau^+(n)$ the number of integers k for which n has a divisor d satisfying $2^k < d \leq 2^{k+1}$. I conjecture that for almost all n

$$(13) \quad \tau^+(n)/\tau(n) \rightarrow 0$$

which of course implies that almost all integers have two divisors satisfying $d_1 < d_2 < 2d_1$. It would be of some interest to get an asymptotic formula for

$$(14) \quad \mathcal{T}(x) := \sum_{n \leq x} \tau^+(n).$$

It is easy to prove that $\mathcal{T}(x)/(x \log x) \rightarrow 1$.

This is an example of Erdős' way of attacking conjectures from many different angles. Indeed, it is often the case that a stronger statement is more accessible than a weaker one, because it reveals a deeper feature. Here, $\tau^+(n) < \tau(n)$ would suffice to prove the desired conjecture, but Erdős asks for much more. As it turns out, hypothesis (13) is wrong (and the constant 1 in the last statement should be replaced by 0, most certainly a lapsus digiti), but the idea of considering the measure of the set $\cup_{d|n} (\log d + [-\frac{1}{2}, \frac{1}{2}])$ was precisely that which eventually led to the solution in [55].

Improving on an estimate of [33] that was already sufficient to invalidate (13), it was shown in [46] (Chapter 4) that the arithmetic function $\tau^+(n)/\tau(n)$ has a limiting distribution $\nu(z)$ satisfying

$$(15) \quad \frac{z}{\sqrt{\log(2/z)}} \ll \nu(z) \ll z \log(2/z) \quad (0 < z < 1).$$

Thus, ν is certainly continuous at the origin. Two interesting open problems are (i) to improve upon (15) and (ii) to determine, if any, the discontinuity points of the distribution function ν .

Regarding the second question, I can prove the following.

Theorem 1. *The distribution function ν is continuous at $z = 1$.*

Proof. We know from theorem 51 of [46] (but this already follows from the analysis given in [55]) that, for every $\varepsilon > 0$, there exists $T_\varepsilon > e^{1/\varepsilon}$ such that all integers n except at most those from a sequence of upper density $\leq \varepsilon/3$ have two divisors d, d' , such that $d < d' < 2^\varepsilon d < T_\varepsilon$. We may of course assume that T_ε increases with $1/\varepsilon$.

Write $n_\varepsilon := \prod_{p^j || n, p \leq T_\varepsilon} p^j$. For a non-exceptional integer n and each $m|(n/n_\varepsilon)$, the two divisors md and md' belong to the same interval $]2^k, 2^{k+1}[$ ($k \in \mathbb{N}$) unless $|(\log md)/\log 2 - k - 1| < \varepsilon$. However, as shown in lemma 48.1 of [46], the discrepancy of the sequence $\{(\log m)/\log 2 : m|(n/n_\varepsilon)\}$ does not exceed ε on a subsequence of lower density $1 - \varepsilon/3$. Thus, if we discard a sequence of integers n of upper density at most $2\varepsilon/3$, we have

$$\tau^+(n) \leq \tau(n) - (1 - \varepsilon)\tau(n/n_\varepsilon).$$

Since, for instance, $\tau(n_\varepsilon) \leq \log T_\varepsilon$ holds on a sequence of lower density $1 - \varepsilon/3$, we get that, writing $\eta := 1/\{2 \log T_\varepsilon\}$,

$$\tau^+(n) \leq \tau(n)\{1 - \eta\}$$

except at most on a sequence of upper density ε . We have therefore proved that $\nu(1 - \eta) \geq 1 - \varepsilon = \nu(1) - \varepsilon$. Observing that ε tends to 0 as a function of η , we obtain the required result. ■

According to a copy of the galley-proof that Nicolas forwarded to me at the time, the statement concerning $\mathcal{J}(x)$ is probably due to some last-minute confusion. It is nevertheless linked to another very interesting problem in probabilistic number theory.

Let $H(x, y, z)$ denote the number of integers not exceeding x having a divisor in $]y, z]$, so that, with the notation (14),

$$\mathcal{J}(x) = \sum_{2^k \leq x} H(x, 2^k, 2^{k+1}).$$

There is a large literature on $H(x, y, z)$, starting with (2) and (3), which can already be seen as evaluations of

$$\limsup_{T \rightarrow \infty} \lim_{x \rightarrow \infty} H(x, T, 2T)/x, \quad \text{and} \quad \lim_{T \rightarrow \infty} \lim_{x \rightarrow \infty} H(x, T, T^{1+\varepsilon_T})/x,$$

respectively. We refer the reader to the recent paper [38] for the history of estimates of $H(x, y, z)$ in the various ranges of the parameters. Here, we only quote the evaluation

$$(16) \quad H(x, y, 2y) \asymp \frac{x}{(\log y)^\delta (\log_2 y)^{3/2}} \quad (2 \leq y \leq \sqrt{x})$$

with $\delta := 1 - (1 + \log_2 2)/\log 2 \approx 0.08607$. These bounds improve on those of [67], where it is shown by a much simpler analysis that

$$e^{-c_1 \sqrt{\log_2 y}} \leq H(x, y, 2y)(\log y)^\delta / x \leq c_2 / \sqrt{\log_2 y}$$

for suitable constants c_1, c_2 . Using the symmetry of the divisors of n around \sqrt{n} , we easily deduce from (14) and (16) the following estimate proved in [38]:

$$(17) \quad \mathcal{J}(x) \asymp \frac{x(\log x)^{1-\delta}}{(\log_2 x)^{3/2}}.$$

Thus, we still fall short of an asymptotic formula for $\mathcal{T}(x)$, although we are now fairly close to one—another challenging problem from an old paper.

Let us continue.

Another interesting and unconventional problem states as follows: let $1 = d_1 < d_2 < \dots < d_{\tau(n)} = n$ be the set of divisors of n . Put

$$\mathcal{G}(n) := \sum_{1 \leq i < \tau(n)} \frac{d_i}{d_{i+1}}.$$

I conjecture that $\mathcal{G}(n) \rightarrow \infty$ if we disregard a sequence of integers n of density 0. This again would imply the conjecture on $d_1 < d_2 < 2d_1$, but needless to say I cannot prove it.

It would be of interest to determine the normal order of $\tau^+(n)$ and of $\mathcal{G}(n)$ (or at least of $\log \tau^+(n)$ and $\log \mathcal{G}(n)$). Also an asymptotic formula for

$$\sum_{n \leq x} \mathcal{G}(n)$$

would be of interest. It is easy to prove that $(1/x) \sum_{n \leq x} \mathcal{G}(n) \rightarrow \infty$.

It turns out to be almost trivial that $\mathcal{G}(n) \rightarrow \infty$ pp. Indeed, if p is the smallest prime factor of n , then $pd_i|n$ for at least $\frac{1}{2}\tau(n)$ values of i and hence $\mathcal{G}(n) > \tau(n)/2p$. In particular, we have $\mathcal{G}(n) > \tau(n)/\xi(n)$ pp whenever $\xi(n) \rightarrow \infty$. It is, however, not true that this lower bound implies (9). Erdős probably had in mind the correct statement that (9) follows from $\mathcal{G}(n) > \frac{1}{2}\tau(n)$ pp, in other words that the distribution function of $\mathcal{G}(n)/\tau(n)$, if it exists, is supported on $[\frac{1}{2}, 1]$.

Erdős and I proved in [34] that $\mathcal{G}(n)/\tau(n)$ does have a distribution function. We actually established a fairly general statement: given any bounded real function ϑ defined on $]0, 1[$, the arithmetical function

$$F(n; \vartheta) := \frac{1}{\tau(n)} \sum_{1 \leq i < \tau(n)} \vartheta\left(\frac{d_i}{d_{i+1}}\right)$$

has a limiting distribution.³

However it is not true that the distribution function of $\mathcal{G}(n)/\tau(n)$ is supported on $[\frac{1}{2}, 1]$. Indeed, we can show that

$$d\{n \geq 1 : \mathcal{G}(n)/\tau(n) \leq \varepsilon\} > 0 \quad (0 < \varepsilon \leq 1).$$

³Note that, in the case $\vartheta := \mathbf{1}_{[1/2, 1]}$, the continuity at 0 of this distribution follows from 1 above and in turn implies (9). This, however, does not yield a new proof of (9) since we actually used a refinement of (9) to establish 1.

This follows from the fact that most integers n free of small prime factors are such that $d_i < \frac{1}{2}\varepsilon d_{i+1}$ for most indices i . We omit the details, which can easily be reconstructed from lemma 4 of [33] and lemma 3 of [34].

As far as average orders are concerned, it is proved in [34] that

$$\sum_{n \leq x} F(n; \vartheta) = \vartheta(1)x \log x + O\left(\frac{x(\log x)^{1-\delta} \log_3 x}{\sqrt{\log_2 x}}\right),$$

provided ϑ is twice continuously differentiable on $[0, 1]$. Here δ is as in (16) and the exponent of $\log x$ is optimal. Moreover, by theorem 3 of [34] and (16), we obtain the improvement

$$\frac{c_1 x (\log x)^{1-\delta}}{(\log_2 x)^{3/2}} \leq x \log x - \sum_{n \leq x} \mathcal{G}(n) \leq \frac{c_2 x (\log x)^{1-\delta}}{(\log_2 x)^{3/2}},$$

valid for suitable positive constants c_1, c_2 .

After a discussion on the normal size of the k -th prime factor $p_k(n)$ of an integer n and a simple proof, via the Turán–Kubilius inequality, of the asymptotic formula

$$(18) \quad \log_2 p_k(n) \sim k \quad (k \rightarrow \infty) \quad \text{pp,}^4$$

Erdős describes a problem on fractional parts of Bernoulli numbers, which does not fit with the focus of this survey. Then, he states two problems related to densities of integer sequences.

Denote by $\lambda_k(p)$ the density of the integers n whose k -th prime factor is p . $\lambda_k(p)$ can easily be calculated by the exclusion-inclusion principle (essentially the sieve of Eratosthenes). By (18), for almost all integers, $p_k(n)$ is about $\exp \exp k$. On the other hand, it is easy to see that the largest value of $\lambda_k(p)$ is assumed for much smaller values of p , in fact for

$$e^{k(1-\varepsilon)} < p < e^{k(1+\varepsilon)}.$$

By more careful computation it would easily be possible to obtain better estimates. The simple explanation for this apparent paradox is that there are very many more values of p at e^{e^k} than at e^k . It is not impossible that $\lambda_k(p)$ is unimodal, i.e. it first increases with p , then assumes its maximum

⁴We do not reproduce this and refer the reader to [46] (chapter 1) and to [74] (theorem III.3.10).

and then decreases. I in fact doubt that $\lambda_k(p)$ behaves so regularly but have not disproved it.

The same problems arise if $\Lambda_k(d)$ denotes the density of the integers m whose k -th divisor is d . Here I obtain that if $d_1(n) < d_2(n) < \dots$ are the consecutive divisors of n then for all but εx integers $n \leq x$ for $k > k_0(\varepsilon, n)$

$$\exp \{ k^{(1/\log 2)-\varepsilon} \} < d_k(n) < \exp \{ k^{(1/\log 2)+\varepsilon} \}.$$

On the other hand, for fixed k , $\Lambda_k(d)$ is maximal for

$$(19) \quad e^{(1-\varepsilon) \log k \log_2 k} < d < e^{(1+\varepsilon) \log k \log_2 k}.$$

It can be shown that $\Lambda_k(d)$ is not unimodal.

The existence of the densities $\lambda_k(p)$ and $\Lambda_k(d)$ immediately follows from the fact that the sequences under consideration are finite unions of congruence classes. The idea of considering the local laws of the distributions of $p_k(n)$ and $d_k(n)$ stems naturally from the law of iterated logarithm underlying (18) (and based upon the fact that the variables $U_j(n)$ defined in (7) are almost Gaussian): indeed, Erdős announced in 1969 [22] that

$$(20) \quad \sum_{\log_2 p \leq k+z\sqrt{k}} \lambda_k(p) = \Phi(z) + o(1) \quad (k \rightarrow \infty),$$

where $\Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt.$

Thus, the study of the $\lambda_k(p)$ is another way of looking at the asymptotic independence of the small prime factors, while, as it turns out, the study of the $\Lambda_k(d)$ is a (positive) test of the dependence of the divisors.

By the sieve of Eratosthenes, we have

$$(21) \quad \lambda_k(p) = \frac{1}{p} \prod_{q < p} \left(1 - \frac{1}{q} \right) s_{k-1}(p) \quad (k \geq 1),$$

where q denotes a prime number and we have put

$$s_j(p) := \sum_{\substack{P^+(m) < p \\ \omega(m) = j}} \frac{1}{m} \quad (j \geq 0).$$

Thus, we have identically

$$F(z, p) := \prod_{q < p} \left(1 + \frac{z}{q-1} \right) = \sum_{j \geq 0} s_j(p) z^j.$$

As noted by Balazard,⁵ this settles, in the affirmative, the question of the unimodality of the sequence $\{s_j(p)\}_{j \geq 1}$ and hence of $\{\lambda_k(p)\}_k$ for all p . Indeed, it is well known (see, e.g., [59], Part V, problem 47) that, if a polynomial has only real roots, then the number of sign changes in the sequence of its coefficients is equal to the number of positive roots. Since, for all positive numbers a_1, \dots, a_n , the polynomial

$$(1 - x) \prod_{1 \leq j \leq n} (x + a_j) = \sum_{0 \leq r \leq n+1} (\sigma_{n-r} - \sigma_{n+1-r})x^r$$

where $\sigma_h := \sum_{1 \leq j_1 < j_2 < \dots < j_h \leq n} a_{j_1} \cdots a_{j_h}$ ($0 \leq j \leq n + 1$), has exactly one positive root, it follows that the sequence $\{\sigma_h\}_{h=0}^n$ of elementary symmetric functions of the a_j is unimodal. Applying this with

$$\{a_j\}_{j=1}^n := \{1/(q - 1) : q < p\}$$

yields the stated property.

Of course the above argument tells us nothing about the mode. An analysis of $\lambda_k(p)$ by the saddle-point method has been achieved by Erdős and myself in [35]. I only quote a few results from this work. Write

$$L := \log \left(\frac{\log p}{\log(k + 1)} \right), \quad M := \log \left(\frac{\log p}{1 + \log^+(k/L)} \right),$$

$$R := L \{ 1 + \log^+(k/L) \}.$$

Then, given any $\varepsilon > 0$, we have

$$\lambda_k(p) = \frac{1}{p} \prod_{q < p} \left(1 - \frac{1}{q} \right) \frac{M^{k-1}}{(k-1)!} e^{O((k-1)/R)} \quad (1 \leq k \leq p^{1-\varepsilon}).$$

$$\frac{\lambda_{k+1}(p)}{\lambda_k(p)} = \frac{M}{k} \left\{ 1 + O\left(\frac{M}{R}\right) \right\}$$

Moreover, we have, for all primes p ,

$$\max_{k \geq 1} \lambda_k(p) = \frac{1 + O(1/\log_2 p)}{p \sqrt{2\pi \log_2 p}}$$

and any value of k realizing this maximum satisfies $k = \log_2 p + O(1)$.

⁵Private communication, February 28, 1989.

For fixed k , the result we found was slightly different from that foreseen by Erdős, probably through a hasty computation. We actually have

$$\max_p \lambda_k(p) = e^{-kQ(\log k)}$$

where $Q(z) := z - \log z - 1 + \frac{2 \log z + 1}{z} + \frac{2(\log z)^2 - \log z + O(1)}{z^2}$; furthermore any value of p realizing this maximum satisfies

$$\log p = \frac{k}{\log k} \left\{ 1 + \frac{2 \log_2 k}{\log k} + \frac{2(\log_2 k)^2 - 3 \log_2 k + O(1)}{(\log k)^2} \right\}.$$

It remains that the phenomenon described by Erdős does hold: modal values of the sequence $\{\lambda_k(p)\}_p$ occur at relatively small values. In other words, in the series

$$\sum_p \lambda_k(p) = 1$$

the decrease of the general term as a function of p is so slow that the contribution of the very numerous terms around $\exp \exp k$ dominate, while the 'large' values around $e^{k/\log k}$ are too few, and indeed not sufficiently large, to contribute significantly to the sum.

To my knowledge, the problem of the (probably non) unimodality of the sequence $\{\lambda_k(p)\}_p$ is still open.

In [15], De Koninck and I improve on (20). Uniformly for $k \geq 1, z \in \mathbb{R}$, we have

$$\sum_{\log_2 p \leq k + z\sqrt{k}} \lambda_k(p) = \Phi(z) + \frac{\Phi_0(z)}{\sqrt{2\pi k}} + O\left(\frac{1}{k}\right)$$

with

$$\Phi_0(z) := e^{-z^2/2} \left\{ \frac{1}{3} + A - \frac{1}{3}z^2 \right\},$$

$$A := \gamma - \sum_p \left\{ \log \left(\frac{1}{1 - 1/p} \right) - \frac{1}{p} \right\} \approx 0.26150.$$

Here γ denotes Euler's constant.

This yields estimates for the median value of the distribution of the k -th prime factor, defined as the largest prime $p^* = p_k^*$ such that

$$\sum_{p \leq p_k^*} \lambda_k(p) < \frac{1}{2}.$$

We find that

$$(22) \quad \log_2 p_k^* = k - b + O(1/k) \quad (k \geq 1)$$

with $b = \frac{1}{3} + A \approx 0.59483$ and numerical computations provide $p_2^* = 37$, $p_3^* = 42719$.

A clear descendant of this problem is the following formula, also proved in [15], which turns out to be an application of the estimate for partial sums of the exponential series—an ancient problem of Ramanujan—needed to prove (22). We have

$$\sum_{\substack{n \leq x \\ \Omega(n) \leq \log_2 x}} 1 = \frac{1}{2}x - x \frac{C + \langle \log_2 x \rangle}{\sqrt{2\pi \log_2 x}} + O\left(\frac{1}{\log_2 x}\right) \quad (x \geq 3),$$

where $C := A - \frac{2}{3} - \sum_p 1/\{p(p-1)\} \approx 0.36798$ and $\langle t \rangle$ denotes the fractional part of the real number t .

As is to be expected, the results on $\Lambda_k(d)$ are much less precise. Erdős' pp-estimate for $\{d_k(n)\}_{1 \leq k \leq \tau(n)}$ immediately follows from the law of iterated logarithm for the prime factors. We obtain in particular, for all $\varepsilon > 0$,

$$\sum_{|\log_2 d - (\log k) / \log 2| > R_k} \Lambda_k(d) = o(1) \quad (k \rightarrow \infty),$$

with $R_k := \sqrt{\{(2 + \varepsilon) / \log 2\} \log k \log_3 k}$. Thus, we can consider that the problem of normal order of $d_k(n)$ is essentially solved. In (19), Erdős raises the problem of modal values of $\Lambda_k(d)$ i.e. of determining as precisely as possible those d such that

$$\Lambda_k(d) = \Lambda_k^* := \max_m \Lambda_k(m).$$

He announces a result which we shall see to be slightly incorrect but nevertheless unveils a rather deep phenomenon.

Let $\tau(n, z)$ denote the number of divisors of n not exceeding z . The following formula, proved in [35], is the analogue of (21):

$$\Lambda_k(d) = \frac{1}{d} \prod_{p \leq d} \left(1 - \frac{1}{p}\right) \sum_{\substack{P^+(m) \leq d \\ \tau(md, d) = k}} \frac{1}{m}.$$

Here, the m -sum obviously depends on the arithmetic structure of m and seemingly harmless questions may reveal to be quite delicate, such as the proof given in [35] of the equivalence

$$(23) \quad \Lambda_k(d) > 0 \iff \tau(d) \leq k \leq d.$$

Let us put

$$K_j := k^{(\log_{j+2} k) / \log 2} \quad (j \geq 0).$$

It is well known that $\min_{\tau(d) \geq k} = K_0^{1+o(1)}$. Now let $N_y := \prod_{p \leq y} p$, where y is the smallest integer such that $\tau(N_y) = 2^{\pi(y)} \geq k$. By selecting $d = d_k(N_y)$ and reducing the m -sum above to the single value $m = N_y/d$, we obtain the left-hand side of the double inequality

$$\frac{k^{O(1)}}{K_0 K_1} \leq \Lambda_k^* \leq \frac{k^{O(1)} K_1}{K_0}$$

proved in [35], while the upper bound already needs a rather involved analysis of the sum. This led Erdős and I in [35] to express the belief that the correct version of (19) should be $d = K_0^{1+o(1)}$.

Indeed, there are essentially two sound models for the structure of those d realizing the mode. Either $\tau(d) \approx k$ and hence $d \approx K_0$ and therefore the m -sum has size $\asymp 1$, or m and d contribute evenly to the divisors counted by $\tau(md, d)$ and $\tau(d) \approx \tau(m, d) \approx \sqrt{k}$, so that d and the values of m appearing in the sum are all at least of size $\sqrt{K_0}$. This latter possibility is of course much more complex than the former, since it implies the existence of many integers m having divisors combining with those of d in such a way that $\tau(md, d) = d$. The above belief corresponded to the conviction that the simplest situation did prevail. However, in [7], La Bretèche and I show that this is not the case: for large k , we have

$$\frac{k^{O(1)}}{K_0 \sqrt{K_1} K_2} \leq \Lambda_k^* \leq \frac{\sqrt{K_2} k^{O(1)}}{K_0 \sqrt{K_1}}, \quad \Lambda_k(d) = \Lambda_k^* \Rightarrow d = K_0^{1/2+o(1)}.$$

(See [7] for a more precise statement and some further information.)

Here again, Erdős' question led to a deeper understanding of the structure of the set of divisors of certain classes of integers and revealed an unexpected phenomenon.

The conjecture (19), although inaccurate, clearly satisfies all criteria quoted at the beginning of this paper. As far as criterion (iv) is concerned,

we quote from [7] the following estimate, where $\Psi_1(x, y)$ denotes the number of y -friable squarefree integers not exceeding x , viz.

$$\Psi_1(x, y) := \sum_{n \leq x, P^+(n) \leq y} \mu(n)^2.$$

Given any $\kappa \geq 1$, we have, for $x \geq 2, y \geq 2, 1 \leq z \leq \min(x, y^\kappa)$,

$$\Psi_1\left(x + \frac{x}{z}, y\right) - \Psi_1(x, y) \ll \frac{\Psi_1(x, y)}{z}.$$

The statement concerning the non-unimodality of $\{\Lambda_k(d)\}_d$ follows easily from (23), since, for any $\varepsilon > 0$, we can construct four integers such that

$$K_0^{1+\varepsilon} < p_1 < d_1 < p_2 < d_2 < 2K_0^{1+\varepsilon},$$

where the p_j are primes and the d_j satisfy $\tau(d_j) > k$ and hence $\Lambda_k(d_j) = 0$ ($j = 1, 2$).

In the next paragraphs of [24], Erdős quotes a number of results related to the normal distribution of prime factors, some of which are stated in [22]. For instance, he explains that, with the notation (7), the statement that $U_j(n)$ and $U_h(n)$ are asymptotically independent provided $j/h \rightarrow \infty$ follows from the methods of [28], his epoch-making paper with Kac on the Gaussian distribution of prime factors. He also comments on the fact that (18) shouldn't be taken too literally by stating the following theorem, which I reproduce with a few changes in the notation.

Let $\{\alpha_k\}_{k=0}^\infty$ tend monotonically to 0 as $k \rightarrow \infty$. Denote by $h_\alpha(n)$ the number of k such that $|\log_2 p_k(n) - k| \leq \alpha_k$. Then, if $\sum_k \alpha_k / \sqrt{k} < \infty$, for every integer m the set $\{n \geq 1 : h_\alpha(n) = m\}$ has a natural density β_m and $\sum_m \beta_m = 1$, in other words $h_\alpha(n)$ has a limiting distribution, while, if $\sum_k \alpha_k / \sqrt{k} = \infty, h_\alpha(n) \rightarrow \infty$ pp.

As far as I know, none of these results has ever been proved in full detail and no effective versions of the statements have been investigated. It would be quite interesting to pursue these tasks with the powerful analytical tools that have been devised since Erdős' paper was written.

The next section of [24] introduces a fundamental concept.

Let $p_1 < p_2 < \dots$ be an infinite sequence of primes. It is quite easy to prove that

$$\sum \frac{1}{p_i} = \infty$$

is the necessary and sufficient condition that almost all integers n should have a prime factor p_i . It seems very difficult to obtain a necessary and sufficient condition that if $a_1 < a_2 < \dots$ is a sequence of integers then almost all integers n should be a multiple of one of the a 's.

I just want to illustrate the difficulty by a simple example. Let

$$n_{i+1} > (1 + c)n_i.$$

Consider the integers m which have a divisor d satisfying

$$n_k < d \leq n_k(1 + \eta_k).$$

If $\sum_{k \geq 1} \eta_k < \infty$, then it is easy to see that the density of these integers exists and is less than 1. If $\sum_{k \geq 1} \eta_k = \infty$, it seems difficult to get a general result, e.g. if $\eta_k = 1/k$ the density in question exists and is less than 1. It seems certain that there is an α , $0 < \alpha < 1$, so that if $\beta < \alpha$ and $\eta_k = 1/k^\beta$ the density of the m having a divisor d , $n_k < d \leq n_k(1 + \eta_k)$ is 1 and if $\beta > \alpha$ it is less than 1.

The problem raised here may be reformulated as follows: characterise those integer sequences \mathcal{A} such that $d\mathcal{M}(\mathcal{A}) = 1$. Following Hall [41], we call such a sequence \mathcal{A} a *Behrend sequence*. This concept has been a constant concern for Erdős during more than fifty years: while, as he remarks in the above excerpt, the corresponding problem is easy when one considers a sequence of primes, or, more generally, a sequence of pairwise coprime integers, delicate and interesting questions arise immediately in the general case, corresponding to the study of strongly dependent random variables.

By the Davenport–Erdős theorem [13] quoted earlier, a necessary and sufficient condition that \mathcal{A} should be a Behrend sequence is that $\delta\mathcal{M}(\mathcal{A}) = 1$ where δ stands for the logarithmic density. Thus, we have obviously that $\delta\mathcal{A} = 1$ is a sufficient condition for \mathcal{A} to be a Behrend sequence. For a long time, I thought that this should have a simple, direct proof, but I could not find one that wasn't essentially equivalent to the Davenport–Erdős general and deep result that $\underline{d}\mathcal{M}(\mathcal{A}) = \delta\mathcal{M}(\mathcal{A})$ for any sequence \mathcal{A} . I eventually came up with the following.

Theorem 2. *Let \mathcal{A} be an integer sequence such that $\delta\mathcal{A} = 1$. Then $d\mathcal{M}(\mathcal{A}) = 1$.*

Proof. Recall that we defined $P^+(n)$ as the largest prime factor of an integer n with the convention that $P^+(1) = 1$. Symmetrically we let $P^-(n)$ denote the smallest prime factor of n and set $P^-(1) = \infty$. For $y \geq 1$, let us write

$$\mathcal{A}_y := \{n \in \mathcal{A} : P^+(n) \leq y\}, \quad n_y := \prod_{\substack{p^\nu \parallel n \\ p \leq y}} p^\nu.$$

As $n_y \in \mathcal{M}(\mathcal{A})$ implies $n \in \mathcal{M}(\mathcal{A})$, we plainly have, for any fixed $y \geq 1$ and $x \rightarrow \infty$,

$$\begin{aligned} \frac{1}{x} \sum_{\substack{n \leq x \\ n \in \mathcal{M}(\mathcal{A})}} 1 &\geq \frac{1}{x} \sum_{\substack{r \in \mathcal{M}(\mathcal{A}_y) \\ P^+(r) \leq y}} \sum_{\substack{s \leq x/r \\ P^-(s) > y}} 1 = \frac{1}{x} \sum_{\substack{r \in \mathcal{M}(\mathcal{A}_y) \\ P^+(r) \leq y}} \left\{ \frac{x}{r} \prod_{p \leq y} \left(1 - \frac{1}{p}\right) + O(1) \right\} \\ &\rightarrow m(y) := \prod_{p \leq y} \left(1 - \frac{1}{p}\right) \sum_{\substack{r \in \mathcal{M}(\mathcal{A}_y) \\ P^+(r) \leq y}} \frac{1}{r}. \end{aligned}$$

Thus, we only have to show that $m(y) \rightarrow 1$ as $y \rightarrow \infty$. We have trivially

$$m(y) \geq \prod_{p \leq y} \left(1 - \frac{1}{p}\right) \sum_{r \in \mathcal{A}_y} \frac{1}{r}.$$

Writing a for an element of \mathcal{A} , we deduce from our hypothesis $\delta\mathcal{A} = 1$ that there is a non-increasing function $\epsilon(x)$ tending to 0 as $x \rightarrow \infty$ such that, for $1 \leq y \leq x$,

$$(24) \quad \{1 - \epsilon(x)\} \log x \leq \sum_{a \leq x} \frac{1}{a} \leq \sum_{r \in \mathcal{A}_y} \frac{1}{r} + \sum_{\substack{n \leq x \\ P^+(n) > y}} \frac{1}{n}.$$

Setting $u := (\log x)/\log y$, we may rewrite the last sum in (24) as

$$\begin{aligned} &\sum_{n \leq x} \frac{1}{n} - \sum_{P^+(n) \leq y} \frac{1}{n} + \sum_{\substack{n > x \\ P^+(n) \leq y}} \frac{1}{n} \\ &\leq \log x + O(1) - \prod_{p \leq y} \left(1 - \frac{1}{p}\right)^{-1} + x^{-1/\log y} \prod_{p \leq y} \left(1 - \frac{1}{p^{1-1/\log y}}\right)^{-1} \\ &\leq \{1 + O(e^{-u})\} \log x + O(1) - \prod_{p \leq y} \left(1 - \frac{1}{p}\right)^{-1}. \end{aligned}$$

Inserting back into (24), we get

$$\sum_{r \in \mathcal{A}_y} \frac{1}{r} \geq \prod_{p \leq y} \left(1 - \frac{1}{p}\right)^{-1} + O(1 + \{e^{-u} + \epsilon(x)\}u \log y).$$

It remains to select $u = 1/\sqrt{\epsilon(y)}$ and let $y \rightarrow \infty$ to obtain $\lim_y m(y) = 1$.

■

The paper [14] (see also [74], Exercises 247-249) contains another fundamental formula, viz.

$$(25) \quad \underline{d}\mathcal{M}(A) = \lim_{T \rightarrow \infty} d\mathcal{M}(A \cap [1, T]).$$

We call the right-hand side the *sequential density* of the set of multiples $\mathcal{M}(A)$. From Behrend's fundamental inequality, valid for finite sequences, we hence deduce from (25) that

$$(26) \quad 1 - \underline{d}\mathcal{M}(A \cup B) \geq \{1 - \underline{d}\mathcal{M}(A)\}\{1 - \underline{d}\mathcal{M}(B)\}$$

holds for all integer sequences A, B .⁶ It follows in particular that

$$(27) \quad \sum_{a \in A} \frac{1}{a} = \infty$$

is a necessary condition for A to be a Behrend sequence, and that any tail $A \setminus [1, T]$ of a Behrend sequence is still a Behrend sequence.

The structure of Behrend sequences long intrigued Erdős. The problem is indeed quite intricate and even seemingly innocent questions, such as that of a criterion for A to be a Behrend sequence in the special case when the members of A only have a bounded number of, or even at most two, prime factors, do not have a simple answer: such a criterion is given in Ruzsa–Tenenbaum [64] in the case of two prime factors; in Erdős–Hall–Tenenbaum [29], it is shown that $d\mathcal{M}(A)$ always exists when the number of prime factors is bounded but that this condition is optimal.

Another interesting feature of Behrend sequences, proved in [47], is that, if A is a Behrend sequence, then $\sum_{d|n, d \in A} 1 \rightarrow \infty$ pp.

Since it seems hopeless to find an effective criterion for the general situation, we are led to consider sequences with a special structure. The sequence \mathcal{E} in (5) is one example. Another instance is that of block sequences, appearing implicitly in Erdős' formulation above. As in [47], we formally define a block sequence by the property that it can be written in the form

$$A = \bigcup_{j \geq 1} \mathcal{A}_j, \quad \mathcal{A}_j :=]T_j, H_j T_j] \cap \mathbb{N}^* \quad (j \geq 1),$$

where the (disjoint) blocks \mathcal{A}_j satisfy some growth condition that guarantees some local regularity, namely that, for some fixed parameter $\eta > 0$,

$$(28) \quad 1 + 1/T_j^{1-\eta} \leq H_j \leq \min(T_j, T_{j+1}/T_j) \quad (j \geq 1).$$

⁶This has been nicely improved by Ahlswede and Khachatryan [1].

When the T_j grow sufficiently fast, we might then expect a Borel–Cantelli type criterion enabling us to decide whether \mathcal{A} is a Behrend sequence according to whether a certain series involving the quantities $d\mathcal{M}(\mathcal{A}_j)$ diverges or not.

These questions have had a fairly wide posterity and many descendants. We refer the reader, in particular, to the papers [41], [47], [64], [73] and to the book [42], for a number of results and conjectures on Behrend sequences and uniform distribution on divisors. Here, we only quote two significant results which confirm, at least in the case of block sequences, that a criterion of Borel–Cantelli type is relevant.

In order to avoid technical hypotheses, we restrict to special cases which still reflect the general picture. We start with a result concerning the situation when the blocks are somewhat short.⁷ The necessity part is due to Hall–Tenenbaum [47], and the sufficiency to Tenenbaum [72].

Theorem 3 ([47], [72]). *Let $\mathcal{A} = \cup_j \mathcal{A}_j$ be a block sequence such that, for suitable real constants $\alpha, \gamma, \sigma, \tau$, with $\sigma > -1$, we have*

$$\log(T_{j+1}/T_j) \asymp j^\sigma (\log j)^\tau, \quad \log H_j \asymp (\log j)^\gamma / j^\alpha \quad (j \rightarrow \infty).$$

Put $\sigma_0 := (\log 2)/(1 - \log 2)$ and define

$$\alpha_0(\sigma) := \begin{cases} (1 - \log 2)(\sigma_0 - \sigma) & \text{if } -1 < \sigma \leq \sigma_0, \\ \sigma_0 - \sigma & \text{if } \sigma > \sigma_0. \end{cases}$$

Then \mathcal{A} is a Behrend sequence if $\alpha < \alpha_0(\sigma)$ and \mathcal{A} is not a Behrend sequence if $\alpha > \alpha_0(\sigma)$.

Note that (28) implies $\sigma + \alpha > 0$ or $\sigma + \alpha = 0$ and $\gamma \leq \tau$.

If we set $\sigma = \tau = \gamma = 0$, we obtain that, provided we have

$$1 + c_1 \leq T_{j+1}/T_j \leq 1 + c_2$$

for suitable constants $c_1 > 0, c_2 > 0$, and $H_j := 1 + 1/j^\alpha$ ($j \geq 1$), the block sequence \mathcal{A} is a Behrend sequence if $\alpha < \log 2$ and is not if $\alpha > \log 2$. This settles Erdős’ conjecture quoted above. His original claim was that the critical exponent α_0 should exist under the sole condition $T_{j+1}/T_j > 1 + c_1$, but this cannot hold as it stands since it follows from theorem 1 of [47] that \mathcal{A} is not a Behrend sequence for any α if we set, for instance, $T_j := \exp \exp j$ ($j \geq 1$). However, he explained later on, in private conversation, that he really had in mind a two-sided condition.

⁷See [72] for an explanation of the fact that any criterion for block Behrend sequences can be split into one in which the block are assumed to be short, in some precise way, and one in which the blocks are assumed to be long.

From Behrend's inequality (26), the condition

$$\sum_j d\mathcal{M}(\mathcal{A}_j) = \infty$$

is necessary for a block sequence to be a Behrend sequence. However, this is in general much weaker than the sufficient condition obtained in [47]. For instance, if we assume, in the setting of Theorem 3, that $-\sigma < \alpha \leq 0$ or that $\alpha = -\sigma \leq 0$ and $\gamma < \tau$, then we have from Ford's estimates in [38] that

$$d\mathcal{M}(\mathcal{A}_j) \asymp \frac{(\log 2j)^{(\gamma-\tau)\delta-3/2}}{j^{(\sigma+\alpha+1)\delta}} \quad (j \geq 1),$$

where δ is as in (16), while Theorem 3 tells us that \mathcal{A} is a Behrend sequence if

$$\sum_j \frac{1}{j^{(\sigma+\alpha+1)\beta}} = \infty$$

for some $\beta > 1 - \log 2$ and, moreover, that \mathcal{A} is not a Behrend sequence if the above series converges for some $\beta < 1 - \log 2$. Hence, we have a pseudo Borel–Cantelli criterion of the shape

$$\sum_j \{d\mathcal{M}(\mathcal{A}_j)\}^{c+o(1)} = \infty,$$

with $c := (1 - \log 2)/\delta \approx 3.566509$. It would be very interesting to have a probabilistic interpretation for conditions of this type.

For the special sequence

$$\mathcal{A}_\lambda := \bigcup_{j \geq 1}] \exp j^\lambda, 2 \exp j^\lambda] \cap \mathbb{N}^*,$$

a refined approach of the same technique yields in [47] a complete proof of Erdős' so called \mathcal{B}_λ -conjecture⁸ dating at least from the seventies and referred to in [46] pp. 49 and 63: \mathcal{A}_λ is a Behrend sequence if, and only if, $\lambda \leq 1/(1 - \log 2)$. This is heuristically justified by the assumption that, for almost all n , the numbers $(\log d)^{1/\lambda}$ are uniformly distributed modulo 1 when d runs through the divisors of n .⁹ However, the limiting case $\lambda = 1/(1 - \log 2)$ is not covered by this argument and indeed needs a more delicate proof.

⁸The name of the conjecture comes from the former notation $\mathcal{B}(\lambda) = \mathcal{M}(\mathcal{A}_\lambda)$.

⁹This is actually proved in [66]. See also [45] and [73].

In the same spirit, and as a clear descendant of this class of problems, I quote the theorem of Kerner and myself [51], according to which

$$(29) \quad \min_{d|n} \|d\vartheta\| = 1/\tau(n)^{1+o(1)} \quad \text{pp,}$$

provided the sequence of convergents $\{p_j/q_j\}_{j=0}^\infty$ of the real number ϑ satisfies

$$(30) \quad \log q_{j+1} < (\log q_j)^{1+o(1)}.$$

Here we used the standard notation $\|t\| = \min_{n \in \mathbb{Z}} |t - n|$. Note that, as explained in [51], it is easy to construct real numbers ϑ contravening (29). A challenging open question is to determine precisely the set of real numbers ϑ such that (29) holds. We know from [51] that (30) cannot be replaced by $\log q_{j+1} < q_j^{(1-\varepsilon)/\log 2}$ with some $\varepsilon > 0$.

When the blocks are long, in a suitable sense, we obtain a similar pseudo-criterion, but with $c = 1$ —hence closer to a classical probabilistic approach.

Theorem 4 ([47]). *Let \mathcal{A} be a block sequence. Assume that, for some $\varepsilon > 0$, we have*

$$\log H_{j+1} > 2(\log T_{j+1})^\varepsilon (\log T_j)^{1-\varepsilon} \quad (j \geq 1).$$

Then $\sum_j \left(\frac{\log H_j}{\log T_j}\right)^{\delta_1} = \infty$ for some $\delta_1 > \delta$ implies that \mathcal{A} is a Behrend sequence, while $\sum_j \left(\frac{\log H_j}{\log T_j}\right)^{\delta_2} < \infty$ for some $\delta_2 < \delta$ implies that \mathcal{A} is not a Behrend sequence.

We refer the reader to chapter 1 of [42] for further results and comments on Behrend sequences. Once more, we see how fertile Erdős’ problems and conjectures revealed themselves along the years.

Erdős follows with refined questions concerning the set of multiples of an interval. I slightly alter the notation in order to match subsequent works.

Denote by $\varepsilon(y, z)$ the density of integers having a divisor d satisfying $y < d \leq z$ and by $\varepsilon_1(y, z)$ the density of integers having precisely one divisor d , $y < d \leq z$. Besicovitch proved $\liminf \varepsilon(y, 2y) = 0$ and I proved that if $(\log z)/\log y \rightarrow 1$, then $\lim \varepsilon(y, z) = 0$ [40] (chapter V). It is easy to see that this result is best possible, i.e. $\lim \varepsilon(y, z) = 0$ implies $(\log z)/\log y \rightarrow 1$.

Further I can prove that $\varepsilon_1(y, z) < c/(\log y)^\alpha$ for a certain $0 < \alpha < 1$. Perhaps $\varepsilon_1(y, z)$ is unimodal for $z > y + 1$, but I know nothing about this. I do not know where $\varepsilon_1(y, z)$ assumes its maximum.

I am sure that $\varepsilon_1(y, z)/\varepsilon(y, z) \rightarrow 0$ for $z = 2y$. If $z - y$ is small then clearly $\varepsilon_1(y, z)/\varepsilon(y, z) \rightarrow 1$ and I do not know where the transition occurs.

Some time ago the following question occurred to me: let k be given and $n > n_0(k)$. Is there an absolute constant α so that for every $n < m \leq n^k$ there is a t , $0 < t \leq (\log n)^\alpha$, so that $m + t$ has a divisor in $]n, 2n]$? More generally: if $n + 1 = a_1 < a_2 < \dots$ is the sequence of integers which have a divisor d , $n < d \leq 2n$, determine or estimate $\max_{a_i < n^k} (a_{i+1} - a_i)$.

Nearly all these questions are now essentially settled. In [70], I proved that if $z - y \rightarrow \infty$ and $z \leq y \left\{ 1 + (\log y)^{1 - \log 4} e^{-\xi \sqrt{\log_2 y}} \right\}$ with $\xi \rightarrow \infty$, then $\varrho_1(y, z) := \varepsilon_1(y, z)/\varepsilon(y, z) \rightarrow 1$, while $\varrho_1(y, z) \geq e^{-c\sqrt{\log y \log_2 y}}$ when

$$z_0(y) := y \left\{ 1 + (\log y)^{1 - \log 4} \right\} < z \leq 2y.$$

On seeing this, Erdős changed his mind concerning the asymptotic behaviour of $\varrho_1(y, z)$ and conjectured that this quantity should tend to a positive limit for $z = 2y$. Ford [38] then proved that $\varrho_1(y, z) \asymp 1$ when $y + 1 \leq z \ll y$. Thus, the transition imagined by Erdős should ideally be seen as a frontier between the cases when $\varrho_1(y, z)$ tends to 1 or to a constant less than 1. We still do not know whether $\varrho_1(y, z)$ tends to a limit when $z_0(y) < z \ll y$ but it follows from Ford's estimates in [38] that $\varrho_1(y, z) \rightarrow 0$ if $z/y \rightarrow \infty$. I conjecture that $\varrho_1(y, z) \rightarrow 1$ when y, z tend to infinity in such a way that $z > y \left\{ 1 + (\log y)^{1 - \log 4 + \varepsilon} \right\}$.

To my knowledge, the question of the unimodality of $\varepsilon_1(y, z)$ as a function of z is still open.

The last problem seems difficult and represents a deep open question. Let $M_n(x)$ denote the counting function of $\mathcal{M}(]n, 2n])$ and set

$$M_n(x) = \varepsilon_n x + R_n(x) \quad (x \geq 1).$$

Then $a_{i+1} - a_i = \{1 - R_n(a_{i+1}) + R_n(a_i)\}/\varepsilon_n$. Since

$$1/\varepsilon_n \asymp (\log n)^\delta (\log_2 n)^{3/2},$$

the first question amounts to asking whether

$$\max_{a_i \leq n^k} |R_n(a_{i+1}) - R_n(a_i)| \ll_k (\log n)^\beta$$

for some β independent of k .

Note that Hall [42] studied the quadratic mean of $R_n(x)$. His lower bound implies that $\sup_x |R_n(x)| \gg n^c$ with

$$c := \frac{1}{2} - \log(\pi^2/6)/\log 4 \approx 0.14098.$$

However, he recently observed [43] that the results obtained in [42] imply much more, namely

$$\sup_x |R_n(x)| > 2^{\{1+o(1)\}n/(2 \log n)}.$$

This follows on noticing that $]n, 2n] = \mathcal{A} \cup \mathcal{B}$ where \mathcal{A} comprises all primes in the interval and \mathcal{B} includes all remaining, composite integers. Then $(a, b) = 1$ whenever $a \in \mathcal{A}$, $b \in \mathcal{B}$. It only remains to apply equations (3.26), (3.10) and (3.20) from [42].¹⁰ Although this does not contradict Erdős' conjecture, it shows that it must be delicate.

I conclude this survey of posterity and descendants of Erdős' paper [24] by quoting a problem that was for him a constant concern even though he thought it might be intractable by any technique at our disposal. Here again, I slightly alter some notations and correct a confusion.

Finally I state an old problem of mine which is probably very difficult and which seems to be unattackable by the methods of probabilistic number theory: denote by $P^+(n)$ the greatest prime factor of n . Is it true that the density of integers n satisfying $P^+(n+1) > P^+(n)$ is $\frac{1}{2}$? Is it true that the density of integers for which

$$(31) \quad P^+(n+1) > P^+(n)n^\alpha$$

exists for every α ? Pomerance and I proved [32] that the upper density of the integers satisfying $n^{-\varepsilon} < P^+(n+1)/P^+(n) < n^\varepsilon$ tends to 0 with ε .

Let $E := \{n \geq 1 : P^+(n) > P^+(n+1)\}$. The conjecture that E has asymptotic density $\frac{1}{2}$ stems from the general hypothesis that n and $n+1$ should be multiplicatively independent. It lies in the same class of problems than the famous *abc*-conjecture.

A general theorem of Hildebrand [49] implies that E has positive lower asymptotic density, but I did not check the numerical value that can be derived from this result. In [32] it is shown that if N is large, then for at least $0.0099N$ values of $n \leq N$ we have $P^+(n) > P^+(n+1)$, and for at

¹⁰The author takes pleasure in thanking Richard R. Hall for letting him include this proof here.

least $0.0099N$ values of $n \leq N$ we have $P^+(n) < P^+(n + 1)$. It follows from theorem 1.2 of [6] that each inequality occurs on a set of integers n of lower asymptotic density

$$\log \left(\frac{1}{1 - c} \right) - 2 \int_0^c \log \left(\frac{1 - v}{1 - v - 2c} \right) \frac{dv}{1 - v}$$

provided $0 < c < 1/5$. The maximum of this expression is greater than 0.05544, which improves the result from [32].

In [32] it is shown that $P^+(n) < P^+(n + 1) < P^+(n + 2)$ holds infinitely often, and it is conjectured that too $P^+(n) > P^+(n + 1) > P^+(n + 2)$ holds infinitely often. This conjecture was proved by Balog [2].

Among several, two further very interesting problems are described in Erdős' seminal article. I chose not to discuss them in detail since they lie somewhat aside of the main stream of the paper, concentrated on the distribution of divisors and typical multiplicative structure of integers.

Thus, I only mention (too) briefly the questions of the number $\Phi(x)$ of distinct values of Euler's totient $\varphi(n)$ in $[1, x]$ and that of an infinite sequence $\{p_j\}_{j=1}^\infty$ of primes such that $p_{j+1} \equiv 1 \pmod{p_j}$ ($j \geq 1$).

On the first problem, a crucial and impressive progress was made by Ford [37]. Improving on results by Pillai [58], Erdős [17], [19], Erdős-Hall [25], [26], Pomerance [60] and Maier-Pomerance [54], he could show that, for large x , we have

$$\Phi(x) \asymp \frac{x}{\log x} e^{C(\log_3 x - \log_4 x)^2} (\log_2 x)^D (\log_3 x)^E,$$

where C and D are positive, explicit constants and $E = D - 2C + \frac{1}{2}$.

On the second problem, Erdős asks whether $\lim p_j^{1/j} = \infty$ necessarily holds and expresses the belief that $p_j < \exp\{j(\log j)^{1+o(1)}\}$ is possible. To my knowledge, both questions are still open. However, Ford, Konyagin and Luca made significant progress in [39].

In conclusion and in the spirit described in the introduction of this article, I hope that this paper will meet two goals. The first is, as for any survey paper, to set records straight, isolate problems and stimulate further research.

Intimately linked to the personality of this so special and so moving (in every sense) man Paul Erdős was, the second goal consists in modestly helping to maintain a fair picture of his offering to mathematics. His problems

have too often been considered as tricky, disconnected questions. All those who worked with him for some time will agree that, even unformulated, he had in mind the bases of many theories and of even more links between these theories. Now that he can read in the Great Book all answers to his innumerable questions, and indeed select the most elegant ones, no doubt he grins once in a while, realizing how close he has been and pondering how many clues he left for us, even if we still cannot understand them all.

Acknowledgements. The author takes pleasure in expressing here warm thanks to R. Balasubramanian, N. Bingham, R. de la Bretèche, C. Dartyge, I. Z. Ruzsa and T. Stoll for their help during the preparation of this paper.

REFERENCES

- [1] R. Ahlswede & L.H. Khachatrian, Density inequalities for sets of multiples, *J. Number Theory* **55** n° 2 (1995), 170–180.
- [2] A. Balog, On triplets with descending largest prime factors, *Studia Sci. Math. Hungar.* **38** (2001), 45–50.
- [3] F. A. Behrend, Three reviews; of papers by Chowla, Davenport and Erdős, *Jahrbuch über die Fortschritte der Mathematik* **60** (1935), 146–149.
- [4] F. A. Behrend, Generalizations of an inequality of Heilbronn and Rohrbach, *Bull. Amer. Math. Soc.* **54** (1948), 681–684.
- [5] A. S. Besicovitch, On the density of certain sequences, *Math. Annalen* **110** (1934), 336–341.
- [6] R. de la Bretèche, C. Pomerance & G. Tenenbaum, On products of ratios of consecutive integers, *Ramanujan J.* **9** (2005), 131–138.
- [7] R. de la Bretèche & G. Tenenbaum, Sur les lois locales de la répartition du k -ième diviseur d'un entier, *Proc. London Math. Soc.* (3) **84** (2002), 289–323.
- [8] R. de la Bretèche & G. Tenenbaum, Oscillations localisées sur les diviseurs, *J. London Math. Soc.* (2) **85** (2012), 669–693.
- [9] R. de la Bretèche & G. Tenenbaum, Moyennes de fonctions arithmétiques de formes binaires, *Mathematika* **58** (2012), 290–304.
- [10] R. de la Bretèche & G. Tenenbaum, Conjecture de Manin pour certaines surfaces de Châtelet, *Inst. Math. Jussieu* (2013), 1–61.
- [11] S. Chowla, On abundant numbers, *J. Indian Math. Soc.* (2) **1** (1934), 41–44.
- [12] H. Davenport, Über numeri abundantes, *Sitzungsber. Preuß. Akad. Wiss., Phys.-Math. Kl.* **26–29** (1933), 830–837.
- [13] H. Davenport & P. Erdős, On sequences of positive integers, *Acta Arith.* **2** (1937), 147–151.
- [14] H. Davenport & P. Erdős, On sequences of positive integers, *J. Indian Math. Soc.* **15** (1951), 19–24.

- [15] J.-M. De Koninck & G. Tenenbaum, Sur la loi de répartition du k -ième facteur premier d'un entier, *Math. Proc. Camb. Phil. Soc.* **133** (2002), 191–204.
- [16] P. Erdős, On the density of the abundant numbers, *J. London Math. Soc.* **9**, n° 4 (1934), 278–282.
- [17] P. Erdős, On the normal number of prime factors of $p - 1$ and some related problems concerning Euler's φ -function, *Quart. J. Math. (Oxford)* **6** (1935), 205–213.
- [18] P. Erdős, A generalization of a theorem of Besicovitch, *J. London Math. Soc.* **11** (1936), 92–98.
- [19] P. Erdős, Some remarks on Euler's φ -function and some related problems, *Bull. Amer. Math. Soc.* **51** (1945), 540–544.
- [20] P. Erdős, On the density of some sequences of integers, *Bull. Amer. Math. Soc.* **54** (1948), 685–692.
- [21] P. Erdős, On some applications of probability to analysis and number theory, *J. London Math. Soc.* **39** (1964), 692–696.
- [22] P. Erdős, On the distribution of prime divisors, *Aequationes Math.* **2**, (1969), 177–183.
- [23] P. Erdős, Problem 218 and solution, *Canad. Math. Bull.* **17** (1974), 621–622.
- [24] P. Erdős, Some unconventional problems in number theory, *Astérisque* **61** (1979), 73–82, Soc. math. France.
- [25] P. Erdős & R. R. Hall, On the values of Euler's φ -function, *Acta Arith.* **22** (1973), 201–206.
- [26] P. Erdős and R. R. Hall, Distinct values of Euler's φ -function, *Mathematika* **23** (1976), 1–3.
- [27] P. Erdős & R. R. Hall, The propinquity of divisors, *Bull. London Math. Soc.* **11** (1979), 304–307.
- [28] P. Erdős & M. Kac, The Gaussian law of errors in the theory of additive number theoretic functions, *Amer. J. Math.* **62** (1940), 738–742.
- [29] P. Erdős, R. R. Hall & G. Tenenbaum, On the densities of sets of multiples, *J. reine angew. Math.* **454** (1994), 119–141.
- [30] P. Erdős & J.-L. Nicolas, Répartition des nombres superabondants, *Bull. Soc. Math. France* **103** (1975), 65–90.
- [31] P. Erdős & J.-L. Nicolas, Méthodes probabilistes et combinatoires en théorie des nombres, *Bull. sc. math.* (2) **100** (1976), 301–320.
- [32] P. Erdős & C. Pomerance, On the largest prime factors of n and $n + 1$, *Aequationes Math.* **17** (1978), no. 2–3, 311–321.
- [33] P. Erdős & G. Tenenbaum, Sur la structure de la suite des diviseurs d'un entier, *Ann. Inst. Fourier (Grenoble)* **31**, 1 (1981), 17–37.
- [34] P. Erdős & G. Tenenbaum, Sur les diviseurs consécutifs d'un entier, *Bull. Soc. Math. de France* **111** (1983), 125–145.
- [35] P. Erdős & G. Tenenbaum, Sur les densités de certaines suites d'entiers, *Proc. London Math. Soc.*, (3) **59** (1989), 417–438.
- [36] P. Erdős & G. Tenenbaum, Sur les fonctions arithmétiques liées aux diviseurs consécutifs, *J. Number Theory*, **31** (1989), 285–311.

- [37] K. Ford, The distribution of totients, *Ramanujan J.* **2** (1998), n^{os} 1–2, 67–151.
- [38] K. Ford, The distribution of integers with a divisor in a given interval. *Ann. of Math. (2)* **168**, n^o 2 (2008), 367–433.
- [39] K. Ford, S. V. Konyagin & F. Luca, Prime chains and Pratt trees, *Geom. Funct. Anal.* **20** n^o 5 (2010), 1231–1258.
- [40] H. Halberstam & H.-E. Richert, *Sieve Methods*, Academic Press, London, New York, San Francisco, 1974.
- [41] R. R. Hall, Sets of multiples and Behrend sequences, *A tribute to Paul Erdős*, 249–258, Cambridge Univ. Press, Cambridge, 1990.
- [42] R. R. Hall, *Sets of multiples*, Cambridge Tracts in Mathematics, 118, Cambridge University Press, Cambridge, 1996.
- [43] R. R. Hall, Private communication, November 11, 2012.
- [44] R. R. Hall & G. Tenenbaum, On the average and normal orders of Hooley's Δ -function, *J. London Math. Soc. (2)* **25** (1982), 392–406.
- [45] R. R. Hall & G. Tenenbaum, Les ensembles de multiples et la densité divisorielle, *J. Number Theory* **22** (1986), 308–333.
- [46] R. R. Hall & G. Tenenbaum, *Divisors*, Cambridge tracts in mathematics 90, Cambridge University Press (1988, paperback ed. 2008).
- [47] R. R. Hall & G. Tenenbaum, On Behrend sequences, *Math. Proc. Camb. Phil. Soc.* **112** (1992), 467–482.
- [48] K. Henriot, Nair–Tenenbaum bounds uniform with respect to the discriminant, *Math. Proc. Camb. Phil. Soc.* **152** (2012), n^o 3, 405–424.
- [49] A. Hildebrand, On a conjecture of Balog, *Proc. Amer. Math. Soc.* **95**, n^o 4 (1985), 517–523.
- [50] C. Hooley, A new technique and its applications to the theory of numbers, *Proc. London Math. Soc. (3)* **38** (1979), 115–151.
- [51] S. Kerner & G. Tenenbaum, Sur la répartition divisorielle normale de $\theta d \pmod{1}$, *Math. Proc. Camb. Phil. Soc.* **137** (2004), 255–272.
- [52] P. Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris (1937, 2nd ed. 1954).
- [53] H. Maier, On the Möbius function, *Trans. Amer. Math. Soc.* **301**, n^o 2 (1987), 649–664.
- [54] H. Maier & C. Pomerance, On the number of distinct values of Euler's φ -function, *Acta Arith.* **49** (1988), 263–275.
- [55] H. Maier & G. Tenenbaum, On the set of divisors of an integer, *Invent. Math.* **76** (1984), 121–128.
- [56] H. Maier & G. Tenenbaum, On the normal concentration of divisors, 2, *Math. Proc. Camb. Phil. Soc.* **147** n^o 3 (2009), 593–614.
- [57] M. Nair & G. Tenenbaum, Short sums of certain arithmetic functions, *Acta Math.* **180**, (1998), 119–144.
- [58] S. Pillai, On some functions connected with $\varphi(n)$, *Bull. Amer. Math. Soc.* **35** (1929), 832–836.
- [59] G. Pólya & G. Szegő, *Problems and theorems in analysis, II*, Classics in Mathematics, Springer-Verlag, Berlin, 1998.

- [60] C. Pomerance, On the distribution of the values of Euler's function, *Acta Arith.* **47** (1986), 63–70.
- [61] A. Raouj, Sur la densité de certains ensembles de multiples, I, II, *Acta Arith.* **69**, n° 2 (1995), 121–152, 171–188.
- [62] A. Raouj, A. Stef & G. Tenenbaum, Mesures quadratiques de la proximité des diviseurs, *Math. Proc. Camb. Phil. Soc.* **150** (2011), 73–96.
- [63] O. Robert, Sur le nombre des entiers représentables comme somme de trois puissances, *Acta Arith.* **149** n° 1 (2011), 1–21.
- [64] I.Z. Ruzsa & G. Tenenbaum, A note on Behrend sequences, *Acta Math. Hung.* **72**, n° 4 (1996), 327–337.
- [65] A. Stef, *L'ensemble exceptionnel dans la conjecture d'Erdős concernant la proximité des diviseurs*, Thèse de doctorat de l'Université Nancy 1, UFR STMIA, juin 1992.
- [66] G. Tenenbaum, Sur la densité divisorielle d'une suite d'entiers, *J. Number Theory* **15**, n° 3 (1982), 331–346.
- [67] G. Tenenbaum, Sur la probabilité qu'un entier possède un diviseur dans un intervalle donné, *Compositio Math.* **51** (1984), 243–263.
- [68] G. Tenenbaum, Sur la concentration moyenne des diviseurs, *Comment. Math. Helvetici* **60** (1985), 411–428.
- [69] G. Tenenbaum, Fonctions Δ de Hooley et applications, *Séminaire de Théorie des nombres, Paris 1984-85, Prog. Math.* **63** (1986), 225–239.
- [70] G. Tenenbaum, Un problème de probabilité conditionnelle en Arithmétique, *Acta Arith.* **49** (1987), 165–187.
- [71] G. Tenenbaum, Sur une question d'Erdős et Schinzel, II, *Inventiones Math.* **99** (1990), 215–224.
- [72] G. Tenenbaum, On block Behrend sequences, *Math. Proc. Camb. Phil. Soc.* **120** (1996), 355–367.
- [73] G. Tenenbaum, Uniform distribution on divisors and Behrend sequences, *L'Enseignement Mathématique* **42** (1996), 153–197.
- [74] G. Tenenbaum, *Introduction à la théorie analytique et probabiliste des nombres*, third ed., coll. Échelles, Berlin, 2008, 592 pp.
- [75] R. C. Vaughan, Sur le problème de Waring pour les cubes, *C.R. Acad. Sci. Paris Sér. I Math.* **301**, n° 6 (1985), 253–255.

Gérald Tenenbaum

Institut Élie Cartan,
Université de Lorraine,
BP 70239,
F-54506 Vandœuvre-lès-Nancy Cedex,
France

e-mail:

gerald.tenenbaum@univ-lorraine.fr

ERDŐS ON POLYNOMIALS

VILMOS TOTIK*

Some results of Erdős on polynomials and some later developments are reviewed. The topics that this survey covers are: discrepancy estimates for zero distribution, orthogonal polynomials, distribution and spacing of their zeros and critical points of polynomials.

1. INTRODUCTION

Polynomials were Paul Erdős' favorite objects in analysis. He devoted many works to them, and in his problem lectures and papers he repeatedly returned to their theory. His major interest concerning them can be roughly divided into the following areas:

- 1) interpolation,
- 2) discrepancy theorems for zeros,
- 3) inequalities,
- 4) size and growth of polynomials,
- 5) geometric problems for lemniscates,
- 6) orthogonal polynomials,
- 7) spacing of zeros,
- 8) geometry of zeros of derivatives,
- 9) polynomials with integer coefficients.

He wrote most papers on interpolation. Several surveys have been devoted to Erdős' work on interpolation, see e.g. D. S. Lubinsky's and P. Vértesi's surveys [22] and [34] in the Erdős memorial volume and Vértesi's survey [35] in this volume. For inequalities, particularly for inequalities on the size of the derivatives of polynomials see T. Erdélyi's papers [5], [6]. We shall not touch topic 4) (questions like how small the norm of a polynomial with ± 1

*Supported by European Research Council Advanced Grant No. 267055

coefficients on the unit circle can be, or if polynomials of degree at most $(1 + \varepsilon)n$ interpolate in n points, then does their minimal norm necessarily tend to infinity—for some interpolation data—when $\varepsilon \rightarrow 0$?) or topic 5) (questions like the minimal length of lemniscates or largest possible area for lemniscate domains) because there has not been a real breakthrough in those questions; see the papers [5] and [16]. Also, 9) (including questions on cyclotomic polynomials) has been adequately reviewed in [3].

Therefore, this survey will be devoted to some recent developments concerning

- 2) discrepancy theorems for zeros,
- 6) orthogonal polynomials,
- 7) spacing of zeros,
- 8) geometry of zeros of derivatives.

In the areas 2), 6) and 7) most of Erdős' earlier papers were with Paul Turán, his lifelong friend. In their works in these directions interpolation has always been in the background. By now more powerful tools have been developed, but the impact of the Erdős–Turán papers has been enormous, and lasts until today.

2. DISCREPANCY THEOREMS

We start with a problem of P. L. Chebyshev. In connection with a question in mechanics he was lead to replacing x^4 on $[-1, 1]$ by a combination of smaller powers. He answered the general question: how well x^n can be approximated by linear combination of smaller powers, i.e. he determined the quantity

$$t_n = \inf_{P_n(x)=x^n+\dots} \|P_n\|_{[-1,1]},$$

where $\|\cdot\|_K$ denotes the supremum norm on a set K :

$$\|P_n\|_K = \sup_{z \in K} |P_n(z)|.$$

He found that

$$(1) \quad t_n = \frac{2}{2^n},$$

the extremal polynomials being the so called Chebyshev polynomials

$$T_n(z) = \frac{1}{2^{n-1}} \cos(n \arccos z).$$

The Chebyshev polynomials have uniformly distributed zeros in the sense that if we project the zeros (all lying in $(-1, 1)$) vertically onto the unit circle to get $2n$ points, then the points so obtained are uniformly distributed there in the sense that they divide the circle into $2n$ equal arcs, see Figure 1.

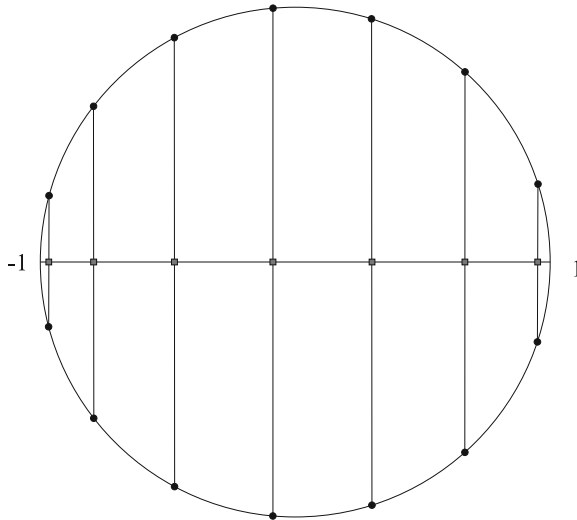


Fig. 1. Uniform distribution of the Chebyshev zeros

What Erdős and Turán observed in [13] is that if the norm of a monic polynomial $P_n(z) = z^n + \dots$ with all its zeros on $[-1, 1]$ is not much larger than the minimal norm t_n , then the zeros of P_n are almost like the zeros of the optimal polynomial T_n , i.e. in a sense the zeros (more precisely their projection on the unit circle) are uniformly distributed.

Theorem 2.1 (Erdős–Turán, 1940). *If $P_n(z) = z^n + \dots$ has all its zeros $\{x_j\}$ in $[-1, 1]$ and*

$$(2) \qquad \|P_n\|_{[-1,1]} \leq \frac{A_n}{2^n},$$

then for any $-1 \leq a < b \leq 1$

$$\left| \frac{\#\{x_j \in (a, b)\}}{n} - \frac{\arcsin b - \arcsin a}{\pi} \right| \leq \frac{8}{\log 3} \sqrt{\frac{\log A_n}{n}}.$$

In particular, if

$$\limsup_{n \rightarrow \infty} \|P_n\|_{[-1,1]}^{1/n} = \frac{1}{2},$$

then the distribution of the zeros is the arcsine distribution (note that, by Chebyshev's theorem, necessarily

$$\liminf_{n \rightarrow \infty} \|P_n\|_{[-1,1]}^{1/n} \geq \frac{1}{2}.$$

In other words, the zeros of asymptotically minimal polynomials have arcsine distribution.

Let us sketch the original argument of Erdős and Turán from [13]; a different approach will be given in the next section. First of all it is enough to prove the upper estimate

$$(3) \quad \#\{x_j \in (a, b)\} - \frac{n(\arcsin b - \arcsin a)}{\pi} \leq \frac{4}{\log 3} \sqrt{n \log A_n},$$

since the matching lower bound (with $4/\log 3$ replaced by $8/\log 3$) follows if we apply (3) to the two complementary intervals $[-1, a]$ and $[b, 1]$. Let $a = \cos \beta$, $b = \cos \alpha$, $\alpha, \beta \in [0, \pi]$, let $k = [n(\beta - \alpha)/\pi]$, and assume that there are at least $k + 2l$ zeros of P_n in $[a, b]$. Consider the following modification of Chebyshev's problem: minimize the supremum norm of monic polynomials $Q_n(z) = z^n + \dots$ on $[-1, 1]$ under the constraint that the polynomial has $k + 2l$ zeros in $[a, b]$. There is an extremal polynomial Q_n , and by a simple variational argument $|Q_n|$ takes its maximal value (with respect to $[-1, 1]$) in between any two of its consecutive zeros lying in (a, b) . According to a lemma of M. Riesz if a trigonometric polynomial of degree n takes its maximum absolute value at a ζ , then it has no zero in the interval $(\zeta - \pi/2n, \zeta + \pi/2n)$. Hence, the trigonometric polynomial $Q_n(\cos \theta)$ cannot have more than $[n(\beta - \alpha)/\pi] = k$ zeros in the interior of (α, β) . Thus, to have $k + 2l$ zeros in $[\alpha, \beta]$ it must have $2l$ zeros at α and β , so in at least one of the endpoints of $[\alpha, \beta]$ it has at least l zeros. Therefore, by assumption,

$$\frac{A_n}{2^n} \geq \min_{\psi_n} \|\psi_n\|_{[-1,1]},$$

where ψ_n is a polynomial which has a zero of multiplicity l somewhere in $[-1, 1]$. As a consequence,

$$\frac{A_n}{2^n} \geq \min_{\psi_n} \left(\frac{1}{\pi} \int_{-1}^1 \frac{|\psi_n(\xi)|^2}{\sqrt{1 - \xi^2}} d\xi \right)^{1/2}.$$

If $I_n(x_0)$ is the minimum value of the norm on the right for all ψ_n which has a zero at x_0 of multiplicity l , then clearly

$$\frac{A_n}{2^n} \geq \min_{x_0 \in [-1,1]} I_n(x_0).$$

On applying the Zhoukowskii transformation $\zeta = \frac{1}{2}(z + \frac{1}{z})$, it follows after multiplication by z^n that $I_n(x_0)^2$ is the minimum

$$\frac{1}{\pi 2^{2n+1}} \int_{-\pi}^{\pi} |\Psi_{2n}|^2,$$

where the minimum is taken for all algebraic polynomials $\Psi_{2n} = z^{2n} + \dots$ which have a zero of multiplicity l at both $e^{\pm i\theta_0}$, where $\cos \theta_0 = x_0$. Reduce the assumption to have a single zero of multiplicity l , which then can be moved to any point on the unit circle by rotation, hence (by moving it to -1)

$$(4) \quad I_n(x_0)^2 \geq \frac{1}{\pi 2^{2n+1}} \min_{\Phi_{2n-l}} \int_{|z|=1} |\Phi_{2n-l}(z)|^2 |1+z|^{2l},$$

the minimum being taken for all polynomials $\Phi_{2n-l}(z) = z^{2n-l} + \dots$ of degree $2n-l$.

Regard here $|1+z|^{2l}$ as a weight function w on the unit circle. It is well known (easily follows from orthogonality) that the minimum in (4) is attained for the $(2n-l)$ -th monic orthogonal polynomial with respect to w . Erdős and Turán figured out the explicit form

$$l \binom{2n+l}{l} (1+z)^{-2l} \int_{-1}^z (z-t)^{l-1} (1+t)^l t^{2n-l} dt$$

for this orthogonal polynomial (once this form is given, one can rather easily check that it is a polynomial of degree $(2n-l)$ with leading coefficient 1 and that it is orthogonal to every smaller power). In other words, the minimum in (4) is attained for this function, and the minimum value for the right-hand side in (4) can then be explicitly calculated and it is

$$\frac{1}{2^{2n}} \binom{2n+l}{l} \binom{2n}{l}^{-1}.$$

Now Stirling's formula easily yields the lower bound

$$\frac{1}{2^n} \exp \left[\left(\frac{\log 3}{4} \right)^2 \frac{l^2}{n} \right]$$

for $I_n(x_0)$. Thus,

$$A_n \geq \exp \left[\left(\frac{\log 3}{4} \right)^2 \frac{l^2}{n} \right],$$

from which (3) immediately follows. ■

Indeed, this is a marvellous argument that gives a sharp estimate. However, it is also clear that it would be difficult to carry it over to Jordan curves or to disconnected sets. We shall see an alternative approach in the next section suitable in such situations.

3. SOME LOGARITHMIC POTENTIAL THEORY

Theorem 2.1 has been used in a number of situations, and has been extended to various directions. Erdős himself proved in [7] that if, besides (2) with $A_n = O(1)$, the maximum of $|P_n|$ in between any two consecutive zeros is $\geq c/2^n$, then

$$\left| \frac{\#\{x_j \in (a, b)\}}{n} - \frac{\arcsin b - \arcsin a}{\pi} \right| \leq C \frac{\log n}{n}.$$

To have a basis for generalization and to understand what is behind Theorem 2.1 (in particular, why the number $1/2$ and the arcsine distribution play such a prominent role) we need to consider what happens if the norm is taken on two intervals or on an even more general set. To do that we shall need to introduce some concepts from potential theory.

First of all, if K is any compact set on the complex plane then we can form Chebyshev's problem on K : what is the minimal norm $\|P_n\|_K$ of monic polynomials $P_n(z) = z^n + \dots$ for a given n ? Call this minimal norm $t_n(K)$. We assume that K has infinitely many points (otherwise $t_n(K) = 0$ for all large n). It is immediate from the definition that $t_{n+m}(K) \leq t_n(K)t_m(K)$, i.e. $\log t_{n+m}(K) \leq \log t_n(K) + \log t_m(K)$, and then it is an easy exercise about sequences that the limit $(\log t_n(K))/n$ exists (it is actually, equal to the infimum of all the numbers $\{(\log t_n(K))/n\}_{n=1}^\infty$). In other words, the limit

$$(5) \quad t(K) = \lim_{n \rightarrow \infty} t_n(K)^{1/n}$$

exists. This $t(K)$ is called the Chebyshev constant of K .

A related quantity is the so called logarithmic capacity that can be obtained via the equilibrium measure of K . If μ is a unit Borel-measure on E , then its logarithmic energy is

$$I(\mu) = \int \int \log \frac{1}{|z - t|} d\mu(z) d\mu(t).$$

If this is finite for some μ , then there is a unique minimizing measure μ_E , called the equilibrium measure of E . Examples:

- the equilibrium measure of $[-1, 1]$ is

$$d\mu_{[-1,1]}(t) = \frac{1}{\pi\sqrt{1-t^2}} dt,$$

which is called the Chebyshev (or arcsine) distribution,

- if C_1 is the unit circle, then

$$d\mu_{C_1}(e^{it}) = \frac{1}{2\pi} dt$$

is the normalized arc measure on C_1 .

Now with the minimal energy $I(K) = \inf_{\mu} I(\mu)$ the logarithmic capacity $\text{cap}(K)$ of K is defined as

$$(6) \quad \text{cap}(K) = e^{-I(K)}.$$

Naturally, if all energies $I(\mu)$ are infinite (in which case there is no equilibrium measure), then $\text{cap}(K) = 0$.

Examples:

- if K is a disk/circle of radius r then $\text{cap}(K) = r$,
- $\text{cap}([-1, 1]) = 1/2$.

It is a simple fact (a consequence of the maximum principle for subharmonic functions) that if $P_n(z) = z^n + \dots$, then

$$(7) \quad \|P_n\|_K \geq \text{cap}(K)^n.$$

Now in the original Chebyshev problem and in Theorem 2.1 the constant $1/2$ appears because it is the logarithmic capacity of $[-1, 1]$: $\text{cap}([-1, 1]) = 1/2$. We can also see that Chebyshev's theorem $t_n \geq 2 \cdot (1/2)^n$ (see (1)) is just a sharper form of (7).

There is yet another related quantity introduced by M. Fekete, the transfinite diameter of K . For a given natural number n we consider n points on K that maximize the product of their distances, i.e. for which the supremum

$$\delta_n(K) := \sup_{z_1, \dots, z_n \in K} \prod_{i \neq j} |z_i - z_j|$$

is achieved. They may not be unique, the points in any maximizing system are called (n -th) Fekete points on K . It is not difficult to show that the limit

$$(8) \quad \delta(K) = \lim_{n \rightarrow \infty} \delta_n^{\frac{1}{n(n-1)}}(K)$$

exists, and this limit is called the transfinite diameter of K .

It is a theorem due (different parts) to Fekete, A. Zygmund and G. Szegő (see e.g. [28, Theorem 5.5.2, Corollary 5.5.5]) that the three quantities: the Chebyshev constant (see (5)), the logarithmic capacity (see (6)) and the transfinite diameter (see (8)) are the same:

$$(9) \quad \text{cap}(K) = \delta(K) = t(K).$$

In modern mathematics mostly the logarithmic capacity is used. Of course, Erdős knew (9), but he never used logarithmic capacity – he was always talking about the transfinite diameter of a set (after all he must have heard it from Fekete himself).

After these preparations let us return to the Erdős–Turán discrepancy Theorem 2.1. It can be formulated as: for any $-1 \leq a < b \leq 1$

$$(10) \quad \left| \frac{\#\{x_j \in (a, b)\}}{n} - \int_a^b \frac{1}{\pi\sqrt{1-x^2}} dx \right| \leq \frac{8}{\log 3} \sqrt{\frac{\log A_n}{n}}.$$

Let δ_x be the “Dirac delta” at x , i.e. the point mass 1 placed to x . If we introduce the normalized zero distribution

$$\nu_n = \frac{1}{n} \sum_j \delta_{x_j}$$

associated with the zeros of P_n , then an equivalent form is: with the Chebyshev distribution

$$d\mu_{[-1,1]}(x) = \frac{1}{\pi\sqrt{1-x^2}} dx$$

for any interval $I \subset [-1, 1]$

$$|\nu_n(I) - \mu_{[-1,1]}(I)| \leq \frac{8}{\log 3} \sqrt{\frac{\log A_n}{n}}.$$

Note that here $\mu_{[-1,1]}$ is the equilibrium measure of $[-1, 1]$, and this is the appropriate form for generalizations.

Let K be a finite union of smooth Jordan arcs (homeomorphic images of $[0, 1]$), and let $P_n(z) = z^n + \dots$ be a monic polynomial. Recall from (7) that we necessarily have $\|P_n\|_K \geq \text{cap}(K)^n$, so asymptotically minimal polynomials on K satisfy

$$(11) \quad \lim_{n \rightarrow \infty} \|P_n\|_K^{1/n} = \text{cap}(K).$$

Erdős and Turán repeatedly mentioned (see e.g. [12, p. 165]) a theorem of Fekete that was communicated to them verbally which claimed that if all zeros of P_n are on single Jordan curve K then (11) is true if and only if the zeros are distributed uniformly with respect to the conformal map Φ of $\mathbf{C} \setminus K$ onto the exterior of the unit disk (i.e. the Φ -image of the zeros is uniformly distributed on the unit circle). This seems to be the first extension of the Erdős–Turán discrepancy theorem from an interval to a general curve. Note that the equilibrium measure μ_K is the Φ -pull-back of the normalized arc-measure on the unit circle: $\mu_K(E) = |\Phi(E)|/2\pi$, so Fekete’s theorem can be rephrased saying that (11) is true if and only if the asymptotic distribution of the zeros is the equilibrium distribution.

The most general form of the Erdős–Turán discrepancy theorem is due to V. V. Andrievskii and H-P. Blatt [1, Theorem 2.4.2]. It involves the quantity A_n for which

$$(12) \quad \|P_n\|_K \leq A_n \operatorname{cap}(K)^n,$$

and neighborhoods J^* of subarcs $J \subset K$ depicted in Figure 2.

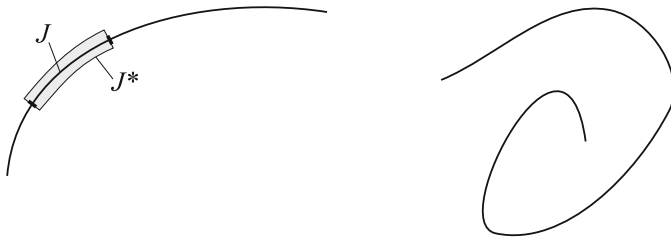


Fig. 2. A neighborhood J^* of a J of K

Theorem 3.1 (Andrievskii–Blatt, 1995–2000). *Let K be a finite union of disjoint smooth Jordan arcs; let ν_n be the normalized zero distribution of a monic polynomial P_n of degree n and let A_n be defined by (12). Then for any subarc $J \subset K$ we have*

$$(13) \quad |\nu_n(J^*) - \mu_K(J^*)| \leq C \sqrt{\frac{\log A_n}{n}},$$

where C depends only on K .

In particular, if $\|P_n\|_K^{1/n} \rightarrow \operatorname{cap}(K)$, then $\nu_n \rightarrow \mu_K$, as was claimed by Fekete for one arc.

When K consists of piecewise smooth arcs, then the square root on the right of (13) must be replaced by a different power that depends on the angles in between the different smooth arcs of K .

To give a flavor of a potential-theoretic argument, in closing this section we give a “modern” approach to the discrepancy Theorem 2.1 of Erdős and Turán (modulo the exact constant).

Let $\mu = \mu_{[-1,1]}$ be the equilibrium measure of $[-1, 1]$ (the arcsine measure). It is again enough to prove an appropriate upper bound for $(\nu_n - \mu)([-1, a])$, $a \in (-1, 1)$. For simplicity assume that $a \in [-2/3, 2/3]$. For a $\delta > 0$ consider the pair of intervals $I^+ := [-1, a]$, $I^- := [a + \delta, 1]$ (a so called condenser). All the constants c_i below depend on δ , but the important c_2 , c_3 , c_5 , c_6 and c_7 lie in between two fixed constants independent of δ . The following are rather simple facts from potential theory. There is a signed measure $\sigma = \sigma^+ - \sigma^-$ (the so called condenser equilibrium measure) such that σ^\pm are positive probability measures, σ^\pm is supported on I^\pm , the logarithmic potential

$$U^\sigma(z) = \int \log \frac{1}{|z - t|} d\sigma(t)$$

of σ equals a constant c_1 on I^- and another constant $c_1 + c_2/\log(1/\delta)$ on I^+ , and everywhere else it lies in between these two constants. It is also true that if $I = I^+ \cup I^-$, then the equilibrium measure μ_I of I majorizes $\sigma^+ + \sigma^-$: $\sigma^+ + \sigma^- \leq (c_3/\delta \log(1/\delta))\mu_I$, furthermore (by Frostman’s theorem [28, Theorem 3.3.4]) the equilibrium potential U^{μ_I} is constant $c_4 (= \log 1/\text{cap}(I))$ on I , it is everywhere else less than c_4 , but on the interval $[a, a + \delta]$ it is bigger than $c_4 - c_5\delta$.

Using these, we obtain from Fubini’s theorem

$$- \int U^\sigma d(\mu - \nu_n) = - \int U^{\mu - \nu_n} d\sigma.$$

Here, since by Frostman’s theorem [28, Theorem 3.3.4] the equilibrium potential U^μ is identically equal to $\log 1/\text{cap}([-1, 1]) = \log 2$ on $[-1, 1]$, we have

$$U^{\mu - \nu_n}(z) = \log 2 + \log |P_n(z)|^{1/n} \leq \frac{\log A_n}{n}, \quad z \in [-1, 1],$$

by the definition of the constant A_n in (2). Hence, since $\sigma(\mathbf{C}) = 0$, we can continue the preceding line as

$$\begin{aligned} - \int U^{\mu - \nu_n} d\sigma &= \int \left(\frac{\log A_n}{n} - U^{\mu - \nu_n} \right) d\sigma \\ &\leq \int \left(\frac{\log A_n}{n} - U^{\mu - \nu_n} \right) d(\sigma^+ + \sigma^-). \end{aligned}$$

Replace on the right $\sigma^+ + \sigma^-$ by the larger $(c_3/\delta \log(1/\delta))\mu_I$, and apply again Fubini's theorem to conclude the following bound for the right-hand side

$$\frac{c_3}{\delta \log(1/\delta)} \frac{\log A_n}{n} - \frac{c_3}{\delta \log(1/\delta)} \int U^{\mu_I} d(\mu - \nu_n).$$

In the last integral U^{μ_I} can be replaced by $U^{\mu_I} - c_4$ (the total mass of $\mu - \nu_n$ is 0), and since $U^{\mu_I} - c_4 = 0$ on I , the integral becomes

$$\int_a^{a+\delta} (U^{\mu_I} - c_4) d(\mu - \nu_n).$$

Since $U^{\mu_I} - c_4 \leq 0$, if we omit here the measure $-\nu_n$ then we decrease the integral. Finally, from $\mu([a, a + \delta]) \leq c_6\delta$ and from $U^{\mu_I} - c_4 \geq -c_5\delta$ on $[a, a + \delta]$ we can conclude

$$-\int U^\sigma d(\mu - \nu_n) \leq \frac{c_3}{\delta \log(1/\delta)} \frac{\log A_n}{n} + \frac{c_3}{\delta \log(1/\delta)} c_5 c_6 \delta^2.$$

On the left we can replace U^σ by $U^\sigma - c_1$, and then the left-hand side becomes

$$-\frac{c_2}{\log(1/\delta)} (\mu - \nu_n)([-1, a]) - \int_a^{a+\delta} (U^\sigma - c_1) d(\mu - \nu_n).$$

Since the last signed integral is at least

$$-\int_a^{a+\delta} (U^\sigma - c_1) d\mu \geq -\frac{c_2}{\log(1/\delta)} \mu([a, a + \delta]) \geq -\frac{c_2 c_6 \delta}{\log(1/\delta)},$$

we finally obtain

$$-\frac{c_2}{\log(1/\delta)} (\mu - \nu_n)([-1, a]) - \frac{c_2 c_6 \delta}{\log(1/\delta)} \leq \frac{c_3}{\delta \log(1/\delta)} \frac{\log A_n}{n} + c_3 c_5 c_6 \frac{\delta}{\log(1/\delta)},$$

i.e.

$$(\nu_n - \mu)([-1, a]) \leq c_7 \left(\frac{\log A_n}{n\delta} + \delta \right).$$

Now the $\delta = \sqrt{\frac{\log A_n}{n}}$ choice gives the desired

$$(\nu_n - \mu)([-1, a]) \leq 2c_7 \sqrt{\frac{\log A_n}{n}}. \quad \blacksquare$$

It is clear from this proof that with appropriate modifications it can be given on smooth Jordan curves, or even on unions of such curves.

4. A SECOND DISCREPANCY THEOREM

Erdős and Turán had a second discrepancy theorem for the zeros of polynomials which had equally important consequences.

Note first of all, that the results from the preceding sections have no direct analogues for Jordan curves (homeomorphic images of the unit circle). Consider e.g. the polynomials z^n on the unit circle C_1 . These have norm 1, which is the (n -th power of the) capacity of C_1 , and still all their zeros lie far from C_1 , which carries the equilibrium distribution. In this section we shall discuss how to get discrepancy theorems for the zeros on Jordan curves.

Let us start with a theorem of R. Jentzsch from 1918: if the radius of convergence of a power series $\sum_{j=0}^\infty a_j z^j$ is 1, then the zeros of (all) the partial sums $\sum_0^n a_j z^j$, $n = 1, 2, \dots$ are dense at ever point of the unit circle. Szegő made a refinement in 1922: there is a sequence $n_1 < n_2 < \dots$ such that if $z_{j,n} = r_{j,n} e^{i\theta_{j,n}}$, $1 \leq j \leq n$ are the zeros of $\sum_0^n a_j z^j$, then $\{\theta_{j,n_k}\}_1^{n_k}$ is asymptotically uniformly distributed (and $r_{j,n_k} \approx 1$ for most j , i.e. for every $\varepsilon > 0$ there are only $o(n_k)$ zeros outside the ring $1 - \varepsilon < |z| < 1 + \varepsilon$).

In connection with these Erdős and Turán proved in [15] the following. Let $P_n(z) = a_n z^n + \dots + a_0$ be a polynomial with zeros $z_j = r_j e^{i\theta_j}$, $1 \leq j \leq n$, and let $C_1 = \{|z| = 1\}$ be the unit circle.

Theorem 4.1 (Erdős–Turán, 1950). *For any interval $J \subset [-\pi, \pi]$*

$$(14) \quad \left| \frac{\#\{\theta_j \in J\}}{n} - \frac{|J|}{2\pi} \right| \leq 16 \sqrt{\frac{\log(\|P_n\|_{C_1} / \sqrt{|a_0 a_n|})}{n}}.$$

Note that

$$\|P_n\|_{C_1} \leq \sum_j |a_j|,$$

so we can replace $\|P_n\|_{C_1}$ on the right of (14) by $\sum_j |a_j|$:

$$(15) \quad \left| \frac{\#\{\theta_j \in J\}}{n} - \frac{|J|}{2\pi} \right| \leq 16 \sqrt{\frac{\log(\sum_j |a_j| / \sqrt{|a_0 a_n|})}{n}}.$$

An immediate consequence is that P_n has at most

$$32 \sqrt{n \log \left(\sum_j |a_j| / \sqrt{|a_0 a_n|} \right)}$$

real zeros (just apply the inequality (15) to the degenerate intervals $J = \{0\}$ and $J = \{\pi\}$). This is better than previous estimates of B. Bloch, G. Pólya and E. Schmidt on the number of real zeros of polynomials, and recaptures a theorem (modulo a constant) of I. Schur.

Next, consider Szegő's theorem mentioned before. In considering $\sum_{j=0}^\infty a_j z^j$ we may assume $a_0 \neq 0$. Now the radius of convergence of this power series is 1 precisely if

$$\limsup_n |a_n|^{1/n} = 1,$$

and this easily implies the existence of a subsequence $\{n_k\}$ with

$$C_{n_k} := \left(\frac{\sum_{j=0}^{n_k} |a_j|}{\sqrt{|a_0 a_{n_k}|}} \right)^{1/n_k} \rightarrow 1.$$

If $z_{j,n} = r_{j,n} e^{i\theta_{j,n}}$ are the zeros of $\sum_{j=0}^n a_j z^j$, then, by (15), we have

$$\left| \frac{\#\{\theta_{j,n_k} \in J\}}{n_k} - \frac{|J|}{2\pi} \right| \leq 16\sqrt{\log C_{n_k}} \rightarrow 0,$$

which shows the uniform distribution of the arguments of the zeros. A relatively simple argument gives that the number of zeros outside any ring $1 - \varepsilon < |z| < 1 + \varepsilon$ tends to zero. Thus, one can easily get both the Jentzsch and the Szegő theorem mentioned before from the Erdős-Turán inequality (14).

This second discrepancy theorem of Erdős and Turán has also been extended in various directions, see e.g. [2], [17]. We only mention the following generalization due to Andrievskii and Blatt [1, Theorem 2.4.5]. Note first of all that if $a_n = 1$, then $\sqrt{|a_0 a_n|}$ in (14)–(15) is just $\sqrt{|P_n(0)|}$, so the following statement is a direct generalization.

Theorem 4.2 (Andrievskii–Blatt, 1995-2000). *If Γ is a smooth Jordan curve, z_0 a fixed point inside Γ , $P_n(z) = z^n + \dots$ and*

$$B_n := \frac{\|P_n\|_\Gamma}{\sqrt{\text{cap}(\Gamma)^n |P_n(z_0)|}},$$

then for all arc $J \subset \Gamma$

$$\left| \frac{\#\{z_j \in J^*\}}{n} - \mu_\Gamma(|J|) \right| \leq C_0 \sqrt{\frac{\log B_n}{n}}.$$

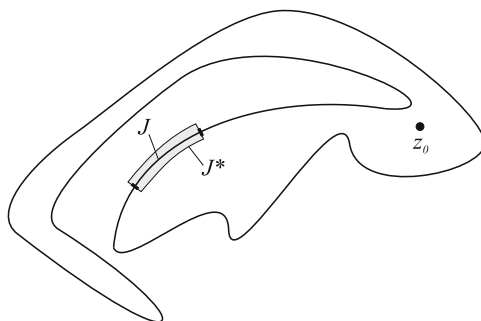


Fig. 3. The position of z_0 and the neighborhood J^* of an arc J

See Figure 3 for the position of z_0 and J^* . Recall that μ_Γ is the equilibrium measure of Γ .

In some form the theorem is actually true for a family of Jordan curves.

The Erdős–Turán discrepancy theorems have motivated many later works; eventually a deep theory of discrepancy of signed measures have evolved, see e.g. the book [1].

5. ORTHOGONAL POLYNOMIALS ON A FINITE INTERVAL

Let ρ be a positive Borel-measure with compact support on the complex plane. The orthonormal polynomials $p_n(z) = \gamma_n z^n + \dots$, $n = 0, 1, \dots$, with respect to ρ are the unique polynomials with $\gamma_n > 0$ and

$$\int p_n \overline{p_m} d\rho = \begin{cases} 0 & \text{if } n \neq m \\ 1 & \text{if } n = m. \end{cases}$$

If S is the support of ρ , then for the leading coefficients γ_n it is always true (see [30, Corollary 1.1.7]) that

$$(16) \quad \frac{1}{\text{cap}(S)} \leq \liminf_{n \rightarrow \infty} \gamma_n^{1/n}.$$

Earlier results on orthogonal polynomials had mostly been about the classical Hermite, Jacobi and Laguerre polynomials. Erdős and Turán were among the first (along with T. J. Stieltjes, S. N. Bernstein and Szegő) who got general results for rather general measures. However, they were always considering the case when the support is $[-1, 1]$.

Theorem 5.1 (Erdős–Turán, 1940). *If the support of ρ is $[-1, 1]$, $d\rho(x) = w(x)dx$ and $w > 0$ almost everywhere, then*

- (a) *the asymptotic zero distribution of p_n is the Chebyshev distribution,*
- (b)

$$|p_n(z)|^{1/n} \rightarrow |z + \sqrt{z^2 - 1}|, \quad z \notin [-1, 1].$$

In the latter limit the convergence is uniform on compact subsets of $\overline{\mathbb{C}} \setminus [-1, 1]$. In particular, it also follows from this theorem that

$$\lim_{n \rightarrow \infty} \gamma_n^{1/n} = 2$$

(c.f. (16) and note that $\text{cap}([-1, 1]) = 1/2$).

Since the classical Jacobi polynomials have also this behavior, one could say that the condition “ $w > 0$ almost everywhere on $[-1, 1]$ ” implies classical behavior. This innocently looking condition turns out to be quite crucial, e.g. we shall see that the behavior of p_n and their zeros is totally different if ρ vanishes on a subinterval of $[-1, 1]$.

It took about 40 years for sharper results, when, in 1977-82, E. A. Rakhmanov [26]–[27] proved that not just (b) is true, but also the stronger

$$(17) \quad \frac{p_{n+1}(z)}{p_n(z)} \rightarrow z + \sqrt{z^2 - 1}, \quad z \notin [-1, 1].$$

H. Widom showed in 1967 that no ratio asymptotics as in (17) is possible if the support is not connected. Thus, in that case one should settle with an analogue of (b) in Theorem 5.1. To state this analogue we need the concept of Green’s function. Let Ω be the unbounded connected component of $\overline{\mathbb{C}} \setminus S$ (where S is the support of ρ), and we assume that S has positive logarithmic capacity, so it has equilibrium measure μ_S (see Section 3). With this equilibrium measure the Green’s function $g_{\overline{\mathbb{C}} \setminus S}(z)$ of $\overline{\mathbb{C}} \setminus S$ with pole at infinity is the function

$$g_{\overline{\mathbb{C}} \setminus S}(z) = \int \log |z - t| d\mu_S(t) - \log \text{cap}(K), \quad z \in \Omega$$

(it is customary to set $g_{\overline{\mathbb{C}} \setminus S}$ to be zero outside Ω). An alternative definition is that $g_{\overline{\mathbb{C}} \setminus S}$ is the unique nonnegative harmonic function in Ω which behaves at infinity as $\log |z| + \text{const}$ and at “almost all points” of $\partial\Omega$ (“almost all” with respect to logarithmic capacity) has zero limit.

Examples:

- if C_R is the circle about the origin of radius R , then

$$g_{\overline{C} \setminus C_R}(z) = \log(|z|/R),$$

-

$$g_{\overline{C} \setminus [-1,1]}(z) = \log |z + \sqrt{z^2 - 1}|.$$

Thus, the function $|z + \sqrt{z^2 - 1}|$ appearing in (b) in Theorem 5.1 can be recognized as the exponential of the Green’s function of $\overline{C} \setminus [-1, 1]$, while the Chebyshev distribution in part (a) is the equilibrium distribution. These guide us to a general formulation.

In discussing the general form of Theorem 5.1 for simplicity assume that $S = \text{supp}(\mu)$ has connected complement and empty interior (e.g. $S \subset \mathbf{R}$), and S is regular in the sense that $g_{\overline{C} \setminus S}(z) \rightarrow 0$ as $z \rightarrow \zeta \in \partial\Omega$, $z \in \Omega$, for all $\zeta \in \partial\Omega$. This latter condition is a mild one, most sets that naturally appear satisfy it. We also assume that there is no zero capacity set that carries the measure ρ .

The next result has evolved through the works of J. Ullman, Erdős, Turán, Widom, H. Stahl and W. Van Assche; in the presented form it is taken from the monograph [30].

Theorem 5.2. *The following are pairwise equivalent.*

- (i) *The asymptotic zero distribution of the orthogonal polynomials p_n is the equilibrium distribution μ_S of the support S of ρ ,*
- (ii)

$$\gamma_n^{1/n} \rightarrow \frac{1}{\text{cap}(S)} \quad \text{as } n \rightarrow \infty,$$

- (iii)

$$|p_n(z)|^{1/n} \rightarrow e^{g_{\overline{C} \setminus S}(z)}, \quad z \notin \text{Con}(S),$$

- (iv) *for all (or one) $0 < q < \infty$*

$$\sup_{P_n} \frac{\|P_n\|_S^{1/n}}{\|P_n\|_{L^q(\rho)}^{1/n}} \rightarrow 1.$$

If either of these properties holds then we say that ρ belongs to the **Reg** class. $\rho \in \mathbf{Reg}$ is a very weak regularity assumption on the measure. $\rho \in \mathbf{Reg}$, i.e. regular behavior means roughly that the measure is not too thin on any part of its support, and in terms of the orthogonal polynomials it means that the orthogonal polynomials behave non-pathologically.

Condition (ii) claims that the leading coefficients are asymptotically minimal (see (16)), while property (iv) says that in n -th root sense the integral norms of polynomials with respect to ρ are about the same (of the same order) as their supremum norm on the support S of ρ (note that $\|P_n\|_{L^q(\mu)} \leq \mu(\mathbf{C})^{1/q} \|P_n\|_S$).

If S has nonzero interior or $\overline{\mathbf{C}} \setminus S$ is not connected, then the equivalence of (ii)–(iv) is still true; but the asymptotic zero distribution may not be μ_S . Consider e.g. the arc-measure on the unit circle or the area-measure on the unit disk as ρ . In these cases the n -th orthonormal polynomial is a constant multiple of z^n , which has all its zeros at the origin, while the equilibrium measure is the normalized arc-measure on the unit circle.

With the **Reg** class we can see that Theorem 5.1 claims nothing else than $S = [-1, 1]$ and $d\rho(x) = w(x)dx$ with $w > 0$ almost everywhere on $[-1, 1]$ imply $\rho \in \mathbf{Reg}$. The condition “ $w > 0$ almost everywhere” is called the (original) Erdős–Turán criterion. In the monograph [30] we called

$$(18) \quad \frac{d\rho(z)}{d\mu_S} > 0 \quad \mu_S - \text{almost everywhere}$$

the general Erdős–Turán criterion. On the left the derivative is the Radon–Nikodym derivative of ρ with respect to the equilibrium measure μ_S of $S = \text{supp}(\rho)$. (When $S = [-1, 1]$ then we have $d\mu_S(x) = (\pi\sqrt{1-x^2})^{-1}dx$ and then clearly (18) is true if and only if

$$\frac{d\rho(x)}{dx} > 0 \quad \text{almost everywhere on } [-1, 1],$$

so (18) is, indeed, a generalization of the original Erdős–Turán criterion.) In the general case we have (see [30, Theorem 4.1.1])

Theorem 5.3 (Stahl–Totik, 1990). *The Erdős–Turán criterion (18) implies $\rho \in \mathbf{Reg}$.*

By now there are many weaker (more powerful) criteria for regularity, see [30, Ch. 4], but no necessary and sufficient condition is known. The only necessary condition (in terms of the size of the measure ρ) is the following: if the support of ρ is $[-1, 1]$ and $\rho \in \mathbf{Reg}$, then for all $\eta > 0$ the capacity of the set

$$E_{\eta,n} := \left\{ x \in [-1, 1] \mid \rho \left(\left[x - \frac{1}{n}, x + \frac{1}{n} \right] \right) > e^{-\eta n} \right\}$$

tends to 1/2 (the capacity of $[-1, 1]$) as $n \rightarrow \infty$.

A closest sufficient condition is

Criterion λ^* : the support of ρ is $[-1, 1]$ and for every $\eta > 0$ the measure of the set $E_{\eta, n}$ tends to 2 (as $n \rightarrow \infty$).

Thus, criterion λ^* implies $\rho \in \mathbf{Reg}$. An analogous criterion for general sets using capacity is

Criterion Λ^* : there is an L such that the capacity of the set

$$(19) \quad \{z \in S \mid \rho(\Delta_{1/n}(z)) > n^{-L}\}$$

tends, as $n \rightarrow \infty$, to the capacity $\text{cap}(S)$ of the support S of ρ (here $\Delta_{1/n}(z)$ denotes the disk of radius $1/n$ with center at z).

In [12] Erdős claimed to have proven a necessary and sufficient condition for $\rho \in \mathbf{Reg}$, but he did not state the condition and he had never published it. He periodically returned to the following statement conjectured by him which, according to [8], he had never been able to fully prove: if $S = [-1, 1]$ and $d\rho(x) = w(x)dx$ with a bounded w , then $\rho \in \mathbf{Reg}$ if and only if $\text{cap}(E_\varepsilon) \rightarrow 1/2$ as $\varepsilon \rightarrow 0$, where E_ε is any set obtained from $\{x \mid w(x) > 0\}$ by removing a subset of measure $< \varepsilon$.

This seems to be still open, though the sufficiency easily follows from Criterion Λ^* in (19).

Regularity plays an important role in the general theory of orthogonal polynomials. It gives a weak global condition under which many properties of orthogonal polynomials can be localized. We shall see examples in the next section.

6. SPACING OF ZEROS OF ORTHOGONAL POLYNOMIALS

Let again $d\rho(x) = w(x)dx$ be a measure on $[-1, 1]$, p_n the orthonormal polynomials with respect to ρ and let $x_j = x_{j,n} = \cos\theta_{j,n} = \cos\theta_j$, $\theta_j \in [0, \pi]$, be the zeros of p_n in increasing order. In this case all zeros lie in $(-1, 1)$, and in the 1930's and 1940's Erdős and Turán had many results on the spacing of these zeros. For the following discussion we speak of rough spacing when

$$(20) \quad \theta_{j-1} - \theta_j \sim \frac{1}{n}, \quad \text{i.e.} \quad \frac{c_1}{n} \leq \theta_{j-1} - \theta_j \leq \frac{c_2}{n}.$$

Fine zero spacing means

$$(21) \quad \theta_{j-1} - \theta_j \approx \frac{\pi}{n}, \quad \text{i.e.} \quad n(\theta_{j-1} - \theta_j) \rightarrow \frac{\pi}{n}.$$

For example, classical (Jacobi) polynomials obey fine spacing inside $(-1, 1)$.

As a first result on rough spacing we mention [12, Theorem VIII] which was the first general result on local rough spacing.

Theorem 6.1 (Erdős–Turán, 1940). *If the support of ρ is $[-1, 1]$, $d\rho(x) = w(x)dx$ with a w that lies in between two positive constants on some interval $[a, b]$, then inside any interval $[a + \varepsilon, b - \varepsilon]$ the zeros of the orthogonal polynomials obey rough spacing.*

By now it has become clear that rough spacing of zeros is basically equivalent to ρ being a doubling measure:

$$\rho(2I) \leq C\rho(I) \quad \text{for all intervals } I \subset [-1, 1].$$

Here $2I$ is the interval I enlarged twice from its center. More precisely, the following was proved in [24, Theorem 1].

Theorem 6.2 (Mastroianni–Totik, 2010). *If ρ is doubling on $[-1, 1]$, then p_n obey rough zero spacing (on the whole interval $[-1, 1]$).*

This includes all previous result on rough spacing of zeros. Furthermore, if ρ is doubling then for the so called Cotes numbers

$$\frac{1}{\lambda_{n,j}} = \sum_{k=0}^n p_k(z_{n,j})^2$$

(these appear in Gaussian quadrature) we have

$$(22) \quad 0 < c \leq \frac{\lambda_{n,j+1}}{\lambda_{n,j}} \leq C$$

uniformly in n and j . Now this uniform boundedness and rough zero spacing is actually equivalent to the doubling condition, see [24, Theorem 3]. It is an open problem if rough spacing alone is equivalent to ρ being doubling (in other words, if rough spacing (20) implies (22)).

These results also have a local version, see [32] and [33].

Fine zero spacing requires more smoothness on the weight. It follows from some deep results of Szegő and Bernstein that if $w \geq c > 0$ (with $d\rho(x) = w(x)dx$) on $[-1, 1]$ and w is twice differentiable on an interval, then inside this interval there is a strong asymptotic formula for the orthogonal polynomials which easily implies fine zero spacing. Erdős and Turán found this approach too restrictive (too “big gun” is used), and they gave the following beautiful theorem.

Theorem 6.3 (Erdős–Turán, 1940). *If $d\rho(x) = w(x)dx$ where $w > 0$ is continuous on $[-1, 1]$, then p_n obeys fine zero spacing for the zeros lying in any subinterval $(-1 + \varepsilon, 1 - \varepsilon)$, i.e.*

$$(23) \quad \theta_{j-1} - \theta_j \approx \frac{\pi}{n}$$

there.

This is no longer true if w is allowed to vanish somewhere on $[-1, 1]$, and it is a delicate question what properties of w imply fine zero spacing. It has turned out that this question is related to some universality problems in random matrix theory, namely to a well defined and “universal” (i.e. independent of ρ) behavior of the kernel function

$$\sum_{k=0}^n p_k(z + a/n)p_k(z + b/n) \quad a, b \in \mathbf{C}.$$

D. S. Lubinsky [21] proved in 2009 this universality under the $\rho \in \mathbf{Reg}$ global condition and under local continuity and positivity of w . The following is a consequence from [19]:

Theorem 6.4 (Levin–Lubinsky, 2008). *If $\rho \in \mathbf{Reg}$ and w is continuous and positive at $z_0 \in (-1, 1)$, then (23) is true for the zeros x_j that lie close to x_0 : $x_j - z_0 = O(1/n)$.*

Now what happens if ρ vanishes on some subinterval of $[-1, 1]$, or more generally, if $d\rho(x) = w(x)dx$ is supported on some general compact set S of the real line? Then the equilibrium measure μ_S of S enters into zero spacing. More precisely, we need the density of that equilibrium measure: if $I \subset S$ is an interval, then μ_S is absolutely continuous on I with respect to Lebesgue-measure: $d\mu_S(t) = \omega_S(t)dt$, and its density ω_S is a C^∞ function there.

Examples:

- for the unit circle/disk the equilibrium density is the identically $1/2\pi$ function on the unit circle,
-

$$\omega_{[-1,1]}(t) = \frac{1}{\pi\sqrt{1-t^2}}, \quad t \in (-1, 1).$$

The following general fine zero spacing theorem was proved by B. Simon [29] and by the author [31] (recall that ω_S is the equilibrium density of the support S of ρ).

Theorem 6.5 (Simon, Totik, 2008-2009). *If $\rho \in \mathbf{Reg}$ and $w(t) := d\rho(t)/dt$ is continuous and positive at a $z_0 \in \text{Int}(S)$, then*

$$(24) \quad \lim_{n \rightarrow \infty} n\omega_S(z_0)(x_{j+1,n} - x_{j,n}) = 1, \quad x_{j,n} - z_0 = O(1/n).$$

Furthermore, if $w > 0$ is continuous on an interval (a, b) , then

$$(25) \quad \lim_{n \rightarrow \infty} n\omega_S(x_j)(x_{j+1,n} - x_{j,n}) = 1$$

uniformly for $x_j \in [a + \varepsilon, b - \varepsilon]$.

It is quite remarkable that local spacing $x_{j+1} - x_j$ of zeros not only reflect (via $\omega_S(x)$) the (global) support of the measure, but also the position of the zero x_j inside that support.

It is also true that if $\log w \in L^1(I)$ on some interval I , then (24) is true at almost all $z_0 \in I$, see [31]. It is an open problem if (24) is true (say on $[-1, 1]$) almost everywhere if, instead of $\log w \in L^1(I)$, we assume only the Erdős–Turán condition $w > 0$ a.e.

7. ERDŐS WEIGHTS

Besides orthogonal polynomials with respect to measures with compact support, orthogonal polynomials associated with weights on the whole real line have important applications. The prototypes are the Hermite polynomials associated with the weight function $w(x) = \exp(-x^2)$. If $d\rho(x) = w(x)dx$ is supported on the whole real line, then the zeros $z_{j,n}$ of the n -th orthogonal polynomials spread out: the largest zero $x_{n,n}$ tends to ∞ and the smallest zero $x_{1,n}$ tends to $-\infty$ as $n \rightarrow \infty$. So in this case one cannot speak of classical zero distribution. One rather considers so called contracted zeros that are obtained by transforming the interval $[x_{1,n}, x_{n,n}]$ linearly onto $[-1, 1]$, and considering the zeros under this linear transformation. Note that this contraction brings all the zeros to $[-1, 1]$, and if these contracted zeros have an asymptotic distribution σ , then σ is called the contracted distribution of the zeros.

In the paper [8] Erdős proved the following.

Theorem 7.1 (Erdős, 1969). *Let $0 < w(x) < C$ on the real line, and assume that to every $\varepsilon > 0$ there is an x_ε such that for every $|x| > x_\varepsilon$ if y is of the same sign as x and $|y| \geq (1 + \varepsilon)|x|$, then*

$$(26) \quad w(y) < w(x)^2$$

holds. Then the contracted zero distribution of the corresponding orthogonal polynomials is the Chebyshev (arcsine) distribution.

It is easy to see that the condition (26) implies

$$(27) \quad w(x) = o(e^{-|x|^\alpha}), \quad |x| \rightarrow \infty$$

for all α . In that same paper Erdős conjectured that (27) alone is sufficient for arcsine contracted zero distribution, but without further regularity this may not be true (a note by Lubinsky). However, under some regularity of

the weight (like monotonicity around infinity) the results of [18, Theorem 14.2] and [20, Theorem 12.2] imply the conjecture (a note by Lubinsky), but those conditions are not as simple as (26).

Why do the conditions (26) and (27) appear in this respect? Already Erdős noticed that if $w(x) = \exp(-|x|^\alpha)$ with some $\alpha > 0$, then the contracted zero distribution is not the Chebyshev distribution (since then it has been calculated that it is

$$\frac{\alpha}{\pi} \int_{|t|}^1 \frac{u^{\alpha-1}}{\sqrt{u^2 - t^2}} du, \quad t \in [-1, 1],$$

so one needs faster decrease to get arcsine distribution. Today weights satisfying (27) are called Erdős weights. The theory (orthogonal polynomials, approximation theory, polynomial inequalities) of Erdős weights has been developed by Lubinsky and Levin (and coauthors) in a series of papers and in the monographs [20] and [18]. There is an analogue on a finite interval: there those weights are called Erdős weights that vanish at the endpoints faster than any power of x ; typical examples $\exp(-1/(1-x^2)^\alpha)$, $\exp(-\exp(1/(1-x^2)^\alpha))$.

8. CRITICAL POINTS OF POLYNOMIALS

Let P_n be a polynomial of degree n , let z_1, \dots, z_n be its zeros and ξ_1, \dots, ξ_{n-1} the zeros of P'_n .

The classical Gauss–Lucas theorem from the mid 1800's claims that every ξ_j is in the convex hull of $\{z_1, \dots, z_n\}$.

Erdős and I. Niven simultaneously with N. G. de Bruijn and T. A. Springer proved in 1947–48 that

$$\frac{1}{n-1} \sum_{j=1}^{n-1} |\Im \xi_j| \leq \frac{1}{n} \sum_{k=1}^n |\Im z_k|,$$

which implies (the reader is asked to do it!)

$$\frac{1}{n-1} \sum_{j=1}^{n-1} |\xi_j| \leq \frac{1}{n} \sum_{k=1}^n |z_k|.$$

This latter theorem lead to a fascinating area about the location of critical points ξ_j . First of all, it was extended by de Bruijn and Springer [4]: for all positive integer m

$$\frac{1}{n-1} \sum_{j=1}^{n-1} |\xi_j|^m \leq \frac{1}{n} \sum_{k=1}^n |z_k|^m.$$

They also conjectured that if $\varphi : \mathbf{C} \rightarrow \mathbf{R}_+$ is convex (in the classical sense that $\varphi(\alpha z + (1-\alpha)w) \leq \alpha\varphi(z) + (1-\alpha)\varphi(w)$ for all z, w and $0 < \alpha < 1$), then

$$\frac{1}{n-1} \sum_{j=1}^{n-1} \varphi(\xi_j) \leq \frac{1}{n} \sum_{k=1}^n \varphi(z_k).$$

Now this has known to be a very strong property through the works in the theory of majorization by Weyl, Birkhoff and Hardy-Littlewood-Pólya. This conjecture of de Bruijn and Springer remained open for more than half a century, and there were several related conjectures (see e.g. [23] and [25]) about the relationship between the zeros ξ_j and z_k .

Many of these conjectures have been resolved by S. M. Malamud [23] and R. Pereira [25] in two simultaneous and independent papers in 2003. To state their theorem let us recall that an $(n-1) \times n$ size $\mathcal{A} = (a_{ij})$ matrix is doubly stochastic if

- $a_{ij} \geq 0$,
- each row-sum equals 1, and
- each column-sum equals $(n-1)/n$.

Let

$$\mathbf{Z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \quad \mathbf{\Xi} = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_{n-1} \end{pmatrix}$$

With these the key property is

Theorem 8.1 (Malamud, Pereira, 2003). *There is a doubly stochastic matrix \mathcal{A} such that $\mathbf{\Xi} = \mathcal{A}\mathbf{Z}$.*

The Gauss–Lucas theorem, the de Bruijn–Springer conjecture etc. are all immediate consequences. Indeed, we have

$$\xi_j = \sum_{k=1}^n a_{jk} z_k,$$

so if φ is convex then

$$\begin{aligned} \frac{1}{n-1} \sum_{j=1}^{n-1} \varphi(\xi_j) &\leq \frac{1}{n-1} \sum_{j=1}^{n-1} \sum_{k=1}^n a_{jk} \varphi(z_k) \\ &= \frac{1}{n-1} \sum_{k=1}^n \varphi(z_k) \sum_{j=1}^{n-1} a_{jk} = \frac{1}{n} \sum_{k=1}^n \varphi(z_k). \end{aligned}$$

Other examples:

$$1) \quad \frac{1}{n-1} \sum_{j=1}^{n-1} |\Re \xi_j|^m \leq \frac{1}{n} \sum_{k=1}^n |\Re z_k|^m, \quad m \geq 1.$$

2) If all zeros lie in the upper-half plane, then

$$\left(\prod_{k=1}^n \Im z_k \right)^{1/n} \leq \left(\prod_{j=1}^{n-1} \Im \xi_j \right)^{1/(n-1)}.$$

Erdős would have loved these results particularly that their proof is quite simple. Malamud and Pereira developed related theories of matrix operations (inverse spectral theorems for normal matrices resp. differentiators), and they obtained Theorem 8.1 as a consequence. But if one only wants to prove Theorem 8.1, then the Malamud-Pereira argument is rather simple (we present Pereira's proof without differentiators). Indeed, we may assume P_n to have leading coefficient 1. Let $\mathbf{E}_1, \dots, \mathbf{E}_n$ be the standard orthonormal basis in \mathbf{C}^n , \mathbf{A} the diagonal matrix/operator with diagonal entries z_1, \dots, z_n , and let $\mathbf{v} = (1, 1, \dots, 1)^T / \sqrt{n}$. With this

$$\mathbf{v}^T (x\mathbf{I}_n - \mathbf{A})^{-1} \mathbf{v} = \frac{1}{n} \sum_{j=1}^n (x - z_j)^{-1} = \frac{1}{n} \frac{P'_n(x)}{P_n(x)}.$$

Let $\mathbf{e}_n = \mathbf{v}$, \mathbf{e}_n^\perp its orthogonal complement and \mathbf{P} the orthogonal projection onto \mathbf{e}_n^\perp . Choose an orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$ in \mathbf{e}_n^\perp in which $\mathbf{P}\mathbf{A}|_{\mathbf{e}_n^\perp}$ has a triangular matrix $\tilde{\mathbf{B}}$. Then $\mathbf{e}_1, \dots, \mathbf{e}_n$ is an orthonormal basis in \mathbf{C}^n and $\tilde{\mathbf{B}}$ is the $(n-1) \times (n-1)$ principal minor of the matrix $\tilde{\mathbf{A}}$ of the operator

\mathbf{A} in that basis. Now if $\tilde{\mathbf{v}} = (0, \dots, 0, 1)^T$ is the representation of $\mathbf{v} = \mathbf{e}_n$ in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$, then

$$\tilde{\mathbf{v}}^T (x\mathbf{I}_n - \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{v}} = ((x\mathbf{I}_n - \mathbf{A})^{-1} \mathbf{e}_n, \mathbf{e}_n) = \mathbf{v}^T (x\mathbf{I}_n - \mathbf{A})^{-1} \mathbf{v} = \frac{1}{n} \frac{P'_n(x)}{P_n(x)}$$

and

$$\tilde{\mathbf{v}}^T (x\mathbf{I}_n - \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{v}} = \det(x\mathbf{I}_{n-1} - \tilde{\mathbf{B}}) / \det(x\mathbf{I}_n - \tilde{\mathbf{A}})$$

because both sides give the (n, n) element of the matrix $(x\mathbf{I}_n - \tilde{\mathbf{A}})^{-1}$. Since the denominator on the right is the characteristic polynomial of $\tilde{\mathbf{A}}$, which is the same as the characteristic polynomial of \mathbf{A} i.e. $P_n(x)$, we get that $P'_n(x)/n = \det(x\mathbf{I}_{n-1} - \tilde{\mathbf{B}})$. Therefore, the diagonal elements in $\tilde{\mathbf{B}}$ (the eigenvalues of $\tilde{\mathbf{B}}$) are ξ_1, \dots, ξ_{n-1} , the zeros of P'_n . With $\mathbf{e}_j = \sum_{k=1}^n (\mathbf{e}_j, \mathbf{E}_k) \mathbf{E}_k$, $j = 1, \dots, n-1$, we have then for $1 \leq j \leq n-1$, $\tilde{\mathbf{e}}_j = (0, \dots, 0, 1, 0, \dots, 0)^T$ (with the 1 in the j -th position)

$$\xi_j = \tilde{\mathbf{e}}_j^T \tilde{\mathbf{B}} \tilde{\mathbf{e}}_j = \tilde{\mathbf{e}}_j^T \tilde{\mathbf{A}} \tilde{\mathbf{e}}_j = (\mathbf{A} \mathbf{e}_j, \mathbf{e}_j) = \sum_{k=1}^n z_k |(\mathbf{e}_j, \mathbf{E}_k)|^2.$$

Now this is the required representation, since $\sum_{k=1}^n |(\mathbf{e}_j, \mathbf{E}_k)|^2 = \|\mathbf{e}_j\|^2 = 1$ and $\sum_{j=1}^{n-1} |(\mathbf{e}_j, \mathbf{E}_k)|^2 = (n-1)/n$ because $|(\mathbf{e}_n, \mathbf{E}_k)|^2 + \sum_{j=1}^{n-1} |(\mathbf{e}_j, \mathbf{E}_k)|^2 = \|\mathbf{E}_k\|^2 = 1$ and $|(\mathbf{e}_n, \mathbf{E}_k)|^2 = |(\mathbf{v}, \mathbf{E}_k)|^2 = 1/n$. ■

The author thanks L. Kérchy, D. S. Lubinsky and the referee for their valuable suggestions concerning the presentation.

REFERENCES

- [1] V. V. Andrievskii and H-P. Blatt, *Discrepancy of signed measures and polynomial approximation*, Springer Monographs in Mathematics. Springer-Verlag, New York, 2002.
- [2] V. V. Andrievskii and H-P. Blatt, Erdős–Turán type theorems on quasiconformal curves and arcs, *J. Approx. Theory*, **97**(1999), 334–365.
- [3] P. Borwein, Paul Erdős and polynomials, *Paul Erdős and his mathematics*, I (Budapest, 1999), 161–174, Bolyai Soc. Math. Stud., **11**, János Bolyai Math. Soc., Budapest, 2002.
- [4] N. G. de Bruijn and T. A. Springer, On the zeros of a polynomial and of its derivative, II, *Nederl. Akad. Wetensch.*, **50**(1947) 264–270, = *Indagationes Math.*, **9**(1947), 458–464.

- [5] T. Erdélyi, Markov-Bernstein type inequalities for polynomials under Erdős type constraints, *Paul Erdős and his mathematics*, I (Budapest, 1999), 219–239, Bolyai Soc. Math. Stud., **11**, János Bolyai Math. Soc., Budapest, 2002.
- [6] T. Erdélyi, Polynomials with Littlewood-type coefficient constraints, *Approximation theory, X* (St. Louis, MO, 2001), 153–196, Innov. Appl. Math., Vanderbilt Univ. Press, Nashville, TN, 2002.
- [7] P. Erdős, On the uniform distribution of the roots of certain polynomials, *Ann. Math.*, **43**(1942), 59–64.
- [8] P. Erdős, On the distribution of the roots of orthogonal polynomials, *Proceedings of the Conference on the Constructive Theory of Functions (Approximation Theory)* (Budapest, 1969), pp. 145–150. Akadémiai Kiadó, Budapest, 1972.
- [9] P. Erdős and I. Niven, On the roots of a polynomial and its derivative, *Bull. Amer. Math. Soc.*, **54**(1948), 184–190.
- [10] P. Erdős and P. Turán, On interpolation I, *Ann. Math.*, **38**(1937), 142–155.
- [11] P. Erdős and P. Turán, On interpolation II, On the distribution of the fundamental points of Lagrange interpolation, *Ann. Math.*, **39**(1938), 703–724.
- [12] P. Erdős and P. Turán, On interpolation III, Interpolatory theory of polynomials, *Ann. Math.*, **41**(1940), 510–553.
- [13] P. Erdős and P. Turán, On the distribution of certain sequences of points, *Ann. Math.*, **41**(1940), 162–173.
- [14] P. Erdős and P. Turán, On a problem in the theory of uniform distribution, I and II, *Indag. Math.*, **10**(1948), 370–378, 406–413.
- [15] P. Erdős and P. Turán, On the distribution of roots of polynomials, *Ann. Math.*, **51**(1950), 105–119.
- [16] A. E. Eremenko and W. Hayman, On the length of lemniscates, *Paul Erdős and his mathematics*, I (Budapest, 1999), 241–242, Bolyai Soc. Math. Stud., **11**, János Bolyai Math. Soc., Budapest, 2002.
- [17] T. Ganelius, Sequences of analytic functions and their zeros, *Ark. Mat.*, **3**(1953), 1–50.
- [18] E. Levin and D. S. Lubinsky, *Orthogonal polynomials for exponential weights*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, **4**. Springer-Verlag, New York, 2001.
- [19] A. L. Levin, and D. S. Lubinsky, Applications of universality limits to zeros and reproducing kernels of orthogonal polynomials, *J. Approx. Theory.*, **150**(2008), 69–95.
- [20] D. S. Lubinsky, *Strong asymptotics for extremal errors and polynomials associated with Erdős-type weights*, Pitman Research Notes in Mathematics Series, **202**. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1989.
- [21] D. S. Lubinsky, A new approach to universality limits involving orthogonal polynomials, *Ann. of Math.*, **170**(2009), 915–939.
- [22] D. S. Lubinsky, A taste of Erdős on interpolation, *Paul Erdős and his mathematics*, I (Budapest, 1999), 423–454, Bolyai Soc. Math. Stud., **11**, János Bolyai Math. Soc., Budapest, 2002.

- [23] S. M. Malamud, Inverse spectral problem for normal matrices and the Gauss-Lucas theorem, *Trans. Amer. Math. Soc.*, **357**(2005), 4043–4064.
- [24] G. Mastroianni and V. Totik, Uniform spacing of zeros of orthogonal polynomials, *Constr. Approx.*, **32**(2010), 181–192.
- [25] R. Pereira, Differentiators and the geometry of polynomials, *J. Math. Anal. Appl.*, **285**(2003), 336–348.
- [26] E. A. Rakhmanov, On the asymptotics of the ratio of orthogonal polynomials, *Math. USSR Sb.*, **32** (1977), 199–213.
- [27] E. A. Rakhmanov, The asymptotic behavior of the ratio of orthogonal polynomials. II. (Russian) *Mat. Sb. (N.S.)* **118(160)**(1982), 104–117.
- [28] T. Ransford, *Potential Theory in the Complex plane*, Cambridge University Press, Cambridge, 1995.
- [29] B. Simon, Two extensions of Lubinsky’s universality theorem, *J. D’Analyse Math.*, **105**(2008), 345–362.
- [30] H. Stahl and V. Totik, *General Orthogonal Polynomials* Encyclopedia of Mathematics, **43**, Cambridge University Press, New York 1992
- [31] V. Totik, Universality and fine zero spacing on general sets, *Arkiv för Math.*, **47**(2009), 361–391.
- [32] V. Totik and with T. Varga, Nonsymmetric fast decreasing polynomials and applications, *J. Math. Anal. Appl.*, **394**(2012), 378–390.
- [33] T. Varga, Uniform spacing of zeros of orthogonal polynomials for locally doubling measures, *Analysis*, (to appear).
- [34] P. Vértesi, On the Lebesgue function and Lebesgue constant: A tribute to Paul Erdős, *Paul Erdős and his mathematics*, I (Budapest, 1999), 705–728, Bolyai Soc. Math. Stud., **11**, János Bolyai Math. Soc., Budapest, 2002.
- [35] P. Vértesi, Paul Erdős and interpolation: problems, results and new developments, *Erdős Centennial* (Budapest, 2013), 711–730, Bolyai Soc. Math. Stud., **25**, János Bolyai Math. Soc., Budapest, 2013.

Vilmos Totik

*Bolyai Institute,
Analysis Research Group of the
Hungarian Academy of Sciences,
University of Szeged,
Szeged,
Aradi v. tere 1, 6720,
Hungary*

and

*Department of Mathematics and
Statistics,
University of South Florida,
4202 E. Fowler Ave, CMC342,
Tampa, FL 33620-5700,
USA*

e-mail: totik@mail.usf.edu

PAUL ERDŐS AND INTERPOLATION: PROBLEMS, RESULTS, NEW DEVELOPMENTS

PÉTER VÉRTESI

1. INTRODUCTION

Pál (Paul) Erdős was born 100 years ago (March 26, 1913 in Budapest). He died on September 20, 1996 in Warsaw, when he attended a conference. He wrote about 1500 papers mainly with coauthors including those more than 80 works which are closely connected with approximation theory (interpolation, mean convergence, orthogonal polynomials, a.s.o.).

The present paper tries to give a short summary of some significant results proved by *Erdős* (and his coauthors) and their new developments in approximation theory, primarily in interpolation; in a way it is an updated version of my previous work [47] from 1990.

We use several survey papers written by *Erdős* himself and on his works (cf. [1], [2] and [3] and their references); sometimes we quote them without mentioning the actual work explicitly. Moreover, we use and quote the corresponding part of the book “A Panorama of Hungarian Mathematics in the Twentieth Century” [1].

2. INTERPOLATION, LAGRANGE INTERPOLATION, LEBESGUE FUNCTION, LEBESGUE CONSTANT, OPTIMAL LEBESGUE CONSTANT

What is interpolation? “Perhaps it would be interesting to dig to the roots of the theory and to indicate its historical origin. Newton, who wanted to draw conclusions from the observed location of comets at equidistant times as to their location at arbitrary times arrived at the problem of determining a ‘geometric’ curve passing through arbitrarily many given points. He solved

this problem by the interpolation polynomial bearing his name ”(Pál Turán [1, p. 23].)

Interpolation theory has been one of the favorite subjects of the twentieth century’s Hungarian approximators. The backbone (mainly of classical interpolation) is the theory developed by *Lipót Fejér*, *Ervin Feldheim*, *Géza Grünwald*, *Pál Turán* and, of course, by *Pál Erdős*.

2.1.

Let us begin with some definitions and notation. Let $C = C(I)$ denote the space of continuous functions on the interval $I := [-1, 1]$, and let \mathcal{P}_n denote the set of algebraic polynomials of degree at most n . $\|\cdot\|$ stands for the usual maximum norm on C . Let X be an *interpolatory matrix (array)*, i.e.,

$$X = \{ x_{kn} = \cos \vartheta_{kn}; \quad k = 1, \dots, n; \quad n = 0, 1, 2, \dots \},$$

with

$$(2.1) \quad -1 \leq x_{nn} < x_{n-1,n} < \dots < x_{2n} < x_{1n} \leq 1,$$

$0 \leq \vartheta_{kn} \leq \pi$, and consider the corresponding *Lagrange interpolation polynomial*

$$(2.2) \quad L_n(f, X, x) := \sum_{k=1}^n f(x_{kn}) \ell_{kn}(X, x), \quad n \in \mathbb{N}.$$

Here, for $n \in \mathbb{N}$,

$$\ell_{kn}(X, x) := \frac{\omega_n(X, x)}{\omega'_n(X, x_{kn})(x - x_{kn})}, \quad 1 \leq k \leq n,$$

with

$$\omega_n(X, x) := \prod_{k=1}^n (x - x_{kn}),$$

are polynomials of exact degree $n - 1$. They are called the *fundamental polynomials* associated with the *nodes* $\{x_{kn}, k = 1, \dots, n\}$ obeying the relations $\ell_{kn}(X, x_{jn}) = \delta_{kj}$, $1 \leq k, j \leq n$.

The main question is: For what choices of the interpolation array X we can expect that (uniformly, pointwise, etc.) $L_n(f, X) \rightarrow f$ ($n \rightarrow \infty$)?

By the classical Lebesgue estimate,

(2.3)

$$\begin{aligned} |L_n(f, X, x) - f(x)| &\leq |L_n(f, X, x) - P_{n-1}(f, x)| + |P_{n-1}(f, x) - f(x)| \\ &\leq \left(\sum_{k=1}^n |\ell_{k,n}(X, x)| + 1 \right) E_{n-1}(f), \end{aligned}$$

therefore, with the notations

$$(2.4) \quad \lambda_n(X, x) := \sum_{k=1}^n |\ell_{kn}(X, x)|, \quad n \in \mathbb{N},$$

$$(2.5) \quad \Lambda_n(X) := \|\lambda_n(X, x)\|, \quad n \in \mathbb{N},$$

(Lebesgue function and Lebesgue constant (of Lagrange interpolation), respectively,) we have for $n \in \mathbb{N}$

$$(2.6) \quad |L_n(f, X, x) - f(x)| \leq \{\lambda_n(X, x) + 1\} E_{n-1}(f)$$

and

$$(2.7) \quad \|L_n(f, X) - f\| \leq \{\Lambda_n(X) + 1\} E_{n-1}(f).$$

Above, as usual

$$E_{n-1}(f) := \min_{P \in \mathcal{P}_{n-1}} \|f - P\|.$$

In 1914 Georg Faber proved the then rather surprising lower bound

$$(2.8) \quad \Lambda_n(X) \geq \frac{1}{12} \log n, \quad n \geq 1,$$

for any interpolation array X . Based on this result he obtained

Theorem 2.1. *For any fixed interpolation array X there exists a function $f \in C$ for which*

$$(2.9) \quad \overline{\lim}_{n \rightarrow \infty} \|L_n(f, X)\| = \infty.$$

2.2.

The preceding estimates underline the importance of the Lebesgue function, $\lambda_n(X, x)$, and the Lebesgue constant, $\Lambda_n(X)$.

To go further, first we state the counterpart of (2.8). Namely, using an estimate of *L. Fejér*

$$\Lambda_n(T) = \frac{2}{\pi} \log n + O(1),$$

one can see that the order $\log n$ in (2.8) is best possible (here T is the Chebyshev matrix, i.e. $x_{kn} = \cos \frac{2k-1}{2n}\pi$).

A very natural problem, raised and answered in 1958 by *Erdős*, says that $\lambda_n(X, x)$ is “big” on a “large” set.

Theorem 2.2 (*Erdős* [4]). *For any fixed interpolation matrix $X \subset [-1, 1]$, real $\varepsilon > 0$, and $A > 0$, there exists $n_0 = n_0(A, \varepsilon)$ so that the set*

$$\{x \in \mathbb{R}, \lambda_n(X, x) \leq A \text{ for all } n \geq n_0(A, \varepsilon)\}$$

has measure less than ε .

The proof of Theorem 2.2 is based on the following simple looking statement (cf. [4, Lemma 3]).

Let y_1, y_2, \dots, y_t be any t ($t > t_0$) distinct numbers in $[-1, 1]$ not necessarily in increasing order. Then, for at least one j ($1 \leq j \leq t$),

$$\sum_{k=1}^{j-1} \frac{1}{|y_k - y_j|} > \frac{t \log t}{8}.$$

(The half-page proof is based on the inequality between the arithmetic and harmonic means.)

Let us mention a nice, relatively new, generalization of this statement. In his paper [31] *Ying Guang Shi* proved as follows:

Let, for a fixed p , $0 < p < \infty$,

$$f_j(p, \mathbf{y}) := \sum_{k=1}^{j-1} \frac{1}{|y_k - y_j|^p}, \quad j = 1, 2, \dots, t; \quad t \geq 2.$$

Then

$$\frac{1}{t} \sum_{j=1}^{t-1} f_j(p, \mathbf{y}) \geq \begin{cases} \frac{t-1}{2^{1+p}}, & 0 < p < 1, \\ \frac{(t-1) \log t}{4}, & p = 1, \\ \frac{(t-1)^{1+p}}{2^p t}, & p > 1. \end{cases}$$

Moreover, the order is the best possible and it is attained by the equidistant nodes.

The next statement, the more or less complete pointwise estimation, is due to P. Erdős and P. Vértesi [5] from 1981.

Theorem 2.3. *Let $\varepsilon > 0$ be given. Then, for any fixed interpolation matrix $X \subset [-1, 1]$ there exist sets $H_n = H_n(\varepsilon, X)$ of measure $\leq \varepsilon$ and a number $\eta = \eta(\varepsilon) > 0$ such that*

$$(2.10) \quad \lambda_n(X, x) > \eta \log n$$

if $x \in [-1, 1] \setminus H_n$ and $n \geq 1$.

Closer investigation shows that (instead of the original $\eta = c\varepsilon^3$) $\eta = c\varepsilon$ can be attained. The behaviour of the Chebyshev matrix, T , shows that (2.10) is the best possible regarding the order $\log n$.

2.3.

Let us say some words about the *optimal Lebesgue constant*. In 1961, P. Erdős, improving a previous result of P. Turán and himself (see [6]), proved that

$$(2.11) \quad \left| \Lambda_n^* - \frac{2}{\pi} \log n \right| \leq c,$$

where

$$\Lambda_n^* := \min_{X \subset I} \Lambda_n(X), \quad n \geq 1,$$

is the *optimal Lebesgue constant*. As a consequence of this result, the closer investigation of Λ_n^* attracted the attention of many mathematicians.

In 1978, Ted Kilgore, Carl de Boor and Alan Pinkus proved the so-called Bernstein-Erdős conjectures concerning the optimal interpolation array X (cf. [7] and [8]).

To formulate the conjecture and the result, let X be *canonical* if $x_{1n} = -x_{nn} = 1$. An elementary argument shows that to obtain the value Λ_n^* it is enough to consider the *canonical* matrices only. Moreover, if

$$\mu_{kn}(X) = \max_{x_{kn} \leq x \leq x_{k-1,n}} \lambda_n(X, x), \quad 2 \leq k \leq n, \quad n \geq 3,$$

denote the $n - 1$ unique local maximum values of $\lambda_n(X, x)$,¹ then we state

¹It is easy to see that for *arbitrary* interpolatory X , $\lambda_n(X, x)$ is a piecewise polynomial with $\lambda_n(X, x) \geq 1$ and $\lambda_n(X, x) = 1$ iff $x = x_{kn}$, $1 \leq k \leq n$. Between the consecutive nodes $\lambda_n(X, x)$ has a *single* maximum, and in $(-1, x_{nn})$ and $(x_{1n}, 1)$ it is convex and monotone (see [46, p. 95]).

Theorem 2.4. *Let $n \geq 3$. We have*

- (i) *there exists a unique optimal canonical X^* with*
- (ii) $\mu_{kn}(X^*) = \mu_{\ell n}(X^*) \quad 2 \leq k, \ell \leq n$.

Moreover, for arbitrary interpolatory X

- (iii) $\min_{2 \leq k \leq n} \mu_{kn}(X) \leq \Lambda_n^* \leq \max_{2 \leq k \leq n} \mu_{kn}(X)$.

Using this result, (2.11) can be considerably improved. Namely,

$$(2.12) \quad \Lambda_n^* = \frac{2}{\pi} \log n + \chi + o(1), \quad n \rightarrow \infty,$$

where $\chi = \frac{2}{\pi} \left(\gamma + \log \frac{4}{\pi} \right) = 0.521251\dots$ and $\gamma = 0.577215\dots$ is the Euler constant (cf. *P. Vértesi* [9]).

2.4.

One of the most talented approximators, the Hungarian *Géza Grünwald*, was a holocaust victim; he was killed in 1942 at the age 32. He was about 25 when, in two fundamental papers, he proved that the Lagrange interpolation can be very bad even for the good matrix $T = \left\{ \cos \frac{2k-1}{2n} \pi \right\}$.

Theorem 2.5 (Grünwald–Marcinkiewicz²). *There exists a function $f \in C$ for which*

$$\overline{\lim}_{n \rightarrow \infty} |L_n(f, T, x)| = \infty$$

for every $x \in [-1, 1]$.

In their third joint paper, [10] *Erdős* and *Grünwald* claimed to prove the existence of an $f \in C$ for which

$$(2.13) \quad \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \left| \sum_{k=1}^n L_k(f, T, x) \right| = \infty,$$

²At the same time the same statement was proved by the Polish mathematician, *Józef Marcinkiewicz*. We must note some other similarities between them. Both were born in 1910; both included the above theorem into their PhD dissertations; they were submitted in 1935; moreover *Marcinkiewicz* was also a victim of the WWII: as his teacher *Antoni Zygmund* writes: “On September 2 [1939], the second day of the war I came across him accidentally in the street in Wilno [Vilnius], already in military uniform. . . A few months later came the news that he was a prisoner of war and was asking for mathematical books”. It seems that this was the last news about *Marcinkiewicz*.

for all $x \in [-1, 1]$. However, as it was discovered later by *Erdős* himself, there had been an oversight in the proof and the method only gives the result with the modulus sign *inside* the summation.

Only in [11], where *Erdős* and *Gábor Halász* (who was born four years after the Erdős–Grünwald paper) were able to complete the proof and obtained the following.

Theorem 2.6. *Given a positive sequence $\{\varepsilon_n\}$ converging to zero however slowly, one can construct a function $f \in C$ such that for almost all $x \in [-1, 1]$*

$$(2.14) \quad \frac{1}{n} \left| \sum_{k=1}^n L_k(f, T, x) \right| \geq \varepsilon_n \log \log n$$

for infinitely many n .

The right-hand side is optimal, for in the paper [12] *Erdős* proved

Theorem 2.7.

$$\frac{1}{n} \left| \sum_{k=1}^n L_k(f, T, x) \right| = o(\log \log n)$$

for almost all x , whenever $f \in C$.

The proof of Theorem 2.7 is an ingenious combination of ideas from number theory, probability and interpolation; it is not by chance that the authors are *Erdős* and *Halász*!

2.5.

After the result of *Grünwald* and *Marcinkiewicz* a natural problem was to obtain an analogous result for an *arbitrary* array X . In [4, p. 384], *Erdős* wrote: “In a subsequent paper I hope to prove the following result:

Let $X \subset [-1, 1]$ be any point group [interpolatory array]. Then there exists a continuous function $f(x)$ so that for almost all x

$$\overline{\lim}_{n \rightarrow \infty} |L_n(f, X, x)| = \infty.”$$

After 4 years of work, *Erdős* and *P. Vértesi* proved the above result ([14]–[15]). *Erdős* writes in [13]: “[Here we prove the above] statement in full detail. The detailed proof turns out to be quite complicated and several unexpected difficulties had to be overcome.”³

³In a personal letter *Erdős* wrote about the *main idea of the proof*: [First] “we should prove that for every fixed A and $\eta > 0$ there exists an M ($M = M(A, \eta)$) such that if we

2.6.

Another significant contribution of the Hungarian approximators to interpolation is the so called “fine and rough theory” (the name was coined by *Erdős* and *Turán* in their basic joint paper [16] dedicated to *L. Fejér* on his 75th birthday in 1955).

In the class $\text{Lip } \alpha$ ($0 < \alpha < 1$; we use the natural setting) a natural error estimate for Lagrange interpolation is

$$\|L_n(f, X) - f\| \leq cn^{-\alpha} \Lambda_n(X)$$

(cf. (2.7)). *Erdős* and *Turán* raised the obvious question: *How sharp is this estimate in terms of the order of the Lebesgue constant as $n \rightarrow \infty$? They themselves considered interpolatory arrays X where*

$$\Lambda_n(X) \sim n^\beta \quad (\beta > 0).$$

In the above paper [16] they prove essentially

Theorem 2.8. *Let X be as above. If $\alpha > \beta$, then we have uniform convergence in $\text{Lip } \alpha$. If $\alpha \leq \beta/(\beta + 2)$, then for some $f \in \text{Lip } \alpha$, Lagrange interpolation is divergent.*

These two cases comprise what is called the “rough theory”, since *solely on the basis of the order of $\Lambda_n(X)$ one can decide the convergence-divergence behavior. However,*

Theorem 2.9. *If $\beta/(\beta + 2) < \alpha \leq \beta$ then anything can happen. That is, there is an interpolatory array Y_1 with $\Lambda_n(Y_1) \sim n^\beta$ and a function $f_1 \in \text{Lip } \alpha$ such that $\overline{\lim}_{n \rightarrow \infty} \|L_n(f_1, Y_1)\| = \infty$, and another interpolation array Y_2 with $\Lambda_n(Y_2) \sim n^\beta$, such that $\lim_{n \rightarrow \infty} \|L_n(f, Y_2) - f\| = 0$ for every $f \in \text{Lip } \alpha$.*

That is, to decide the convergence-divergence behavior *we need more information than just the order of the Lebesgue constant.* The corresponding situation is called “fine theory”.

This paper of *Erdős* and *Turán* has been very influential. It left open a number of problems and attracted the attention not only of the Hungarian school of interpolation (*Géza Freud, Ottó Kis, Melánia Sallay, József Szabados, P. Vértesi*), but also of others (including *R. J. Nessel, W. Dickmeis, E. van Wickeren*).

divide the interval $[-1, 1]$ into M equal parts I_1, \dots, I_M then

$$\sum'_k |\ell_{k,n}(X, x)| > A, \quad x \in I_r,$$

apart from a set of measure $\leq \eta$. Here \sum' means that k takes those values for which $x \notin I_r$ ”.

2.7.

The Faber-theorem (2.9) is a special case of a general statement proved by *S. M. Losinskii* and *F. I. Harsiladze* on (linear) projection operators (p.o.). (That means $\mathcal{L}_n : C \rightarrow \mathcal{P}_{n-1}$ is a linear bounded operator and $\mathcal{L}_n(f) \equiv f$ iff $f \in \mathcal{P}_{n-1}$). Namely, they proved that if

$$\|\mathcal{L}_n\| := \sup_{\|f\| \leq 1} \|\mathcal{L}_n(f, x)\|, \quad f \in C,$$

then

$$(2.15) \quad \|\mathcal{L}_n\| \geq \frac{\log n}{8\sqrt{\pi}}$$

(\mathcal{L}_n is a p.o.). If $\mathcal{L}_n = L_n(X)$ (Lagrange interpolation), then, obviously $\Lambda_n(X) = \|\mathcal{L}_n\|$.

In his paper [17], *G. Halász* formulated some results on

$$\mathcal{L}_n(x) := \sup_{\|f\| \leq 1} |\mathcal{L}_n(f, x)|, \quad f \in C$$

(it generalizes the Lebesgue function $\lambda_n(X, x)$). Among others he states

Theorem 2.10. For any sequence of projections \mathcal{L}_n

- (i) $\overline{\lim}_{n \rightarrow \infty} \mathcal{L}_n(x) = \infty$ on a set of positive measure in $[-1, 1]$;
- (ii) $\lim_{n \rightarrow \infty} \int_{-1}^1 h(\log \mathcal{L}_n(x)) \log \mathcal{L}_n(x) dx = \infty$ whenever

$$I := \int_2^\infty \frac{h(x)}{x \log x} dx = \infty.$$

- (iii) If $I < \infty$ then there exists a sequence \mathcal{L}_n such that

$$\sup_n \int_{-1}^1 h(\log \mathcal{L}_n(x)) \log \mathcal{L}_n(x) dx < \infty.$$

2.8.

Here we mention some recent developments of the previous results. First, let us see the multidimensional analogon of the estimation (2.15).

Let \mathbb{R}^d (direct product) be the Euclidean d -dimensional space ($d \geq 1$, fixed) and let $\mathbb{T}^d = \mathbb{R}^d \pmod{2\pi\mathbb{Z}^d}$ denote the d -dimensional torus, where $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$.

Further, let $C(\mathbb{T}^d)$ denote the space of (complex valued) continuous functions on \mathbb{T}^d . By definition they are 2π -periodic in each variable.

For $g \in C(\mathbb{T}^d)$ we define its Fourier series by

$$g(\boldsymbol{\vartheta}) \sim \sum_{\mathbf{k}} \hat{g}(\mathbf{k})e^{i\mathbf{k}\cdot\boldsymbol{\vartheta}}, \quad \hat{g}(\mathbf{k}) = \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} g(\mathbf{t})e^{-i\mathbf{k}\cdot\mathbf{t}} d\mathbf{t},$$

where $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2, \dots, \vartheta_d) \in \mathbb{T}^d$, $\mathbf{k} = (k_1, k_1, \dots, k_d) \in \mathbb{Z}^d$ and $\mathbf{k} \cdot \boldsymbol{\vartheta} = \sum_{l=1}^d k_l \vartheta_l$ (scalar product).

The *rectangular* n -th partial sum of the Fourier series is defined by

$$S_{nd}^{[r]}(g, \boldsymbol{\vartheta}) := \sum_{|\mathbf{k}|_\infty \leq n} \hat{g}(\mathbf{k})e^{i\mathbf{k}\cdot\boldsymbol{\vartheta}} \quad (n \in \mathbb{N}_0 = \{0, 1, 2, \dots\});$$

the *triangular* one is

$$S_{nd}(g, \boldsymbol{\vartheta}) := \sum_{|\mathbf{k}|_1 \leq n} \hat{g}(\mathbf{k})e^{i\mathbf{k}\cdot\boldsymbol{\vartheta}} \quad (n \in \mathbb{N}_0).$$

Above, $|\mathbf{k}|_\infty = \max_{1 \leq l \leq d} |k_l|$ and $|\mathbf{k}|_1 = \sum_{k=1}^d |k_l|$ (they are the l_p norms of the multiindex \mathbf{k} for $p = \infty$ and $p = 1$). The names “rectangular” and “triangular” refer to the shape of the corresponding indices of terms when $d = 2$ and $0 \leq k_1, k_2, |\mathbf{k}|_\infty \leq n, |\mathbf{k}|_1 \leq n$ respectively.

In a way the investigation of the $S_{nd}^{[r]}$ is apparent: in many cases in essence it is a one variable problem (see [42] and [41]).

However there are only relatively few works dealing with the triangular (or l_1) summability (cf. [43] and [44]).

Introducing the notations

$$D_{nd}(\boldsymbol{\vartheta}) = \sum_{|\mathbf{k}|_1 \leq n} e^{i\mathbf{k}\cdot\boldsymbol{\vartheta}} \quad (n \geq 1),$$

where $\mathbf{k} \in \mathbb{Z}^d$, one can see that

$$\begin{aligned} S_{nd}(g, \boldsymbol{\vartheta}) &= (g * D_{nd})(\boldsymbol{\vartheta}) := \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} g(\boldsymbol{\vartheta} - \mathbf{t}) D_{nd}(\mathbf{t}) \, d\mathbf{t} = \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} g(\boldsymbol{\vartheta} + \mathbf{t}) D_{nd}(\mathbf{t}) \, d\mathbf{t}, \end{aligned}$$

where as before, $g \in C(\mathbb{T}^d)$, $\boldsymbol{\vartheta}, \mathbf{t} \in \mathbb{T}^d$.

Let $\|g\| := \max_{\boldsymbol{\vartheta} \in \mathbb{T}^d} |g(\boldsymbol{\vartheta})|$,

$$\|S_{nd}\| := \max_{\substack{g \in C(\mathbb{T}^d) \\ \|g\| \leq 1}} \|S_{nd}(g, \boldsymbol{\vartheta})\| \quad (n \geq 1)$$

and

$$\|g\|_p := \left(\int_{\mathbb{T}^d} |g(\boldsymbol{\vartheta})|^p \, d\boldsymbol{\vartheta} \right)^{1/p}$$

if $g \in L^p := \{\text{the set of all measurable } 2\pi \text{ periodic (in each variable) functions on } \mathbb{T}^d\}$, $1 \leq p < \infty$.

We state

Theorem 2.11. *We have, for any fixed $d \geq 1$,*

$$(2.16) \quad \|D_{nd}\|_1 = \|S_{nd}\| \sim (\log n)^d \quad (n \geq 2).^4$$

One of the most characteristic properties of the Fourier series in one dimension is the so called Faber–Marcinkiewicz–Berman theorem, namely that the operator S_n has the smallest norm among all projection operators (cf. [45, p. 281] for other details). This part extends the above statement for S_{nd} , $d \geq 1$.

Let \mathcal{T}_{nd} be the space of trigonometric polynomials of form

$$\sum_{|\mathbf{k}|_1 \leq n} (a_{\mathbf{k}} \cos(\mathbf{k} \cdot \boldsymbol{\vartheta}) + b_{\mathbf{k}} \sin(\mathbf{k} \cdot \boldsymbol{\vartheta})),$$

where $\mathbf{k} = (k_1, k_2, \dots, k_d)$ and $k_1, \dots, k_d \geq 0$, arbitrary real numbers. Moreover, let T_{nd} be a linear trigonometric projection operator on $C(\mathbb{T}^d)$, i.e. $T_{nd}(g, \boldsymbol{\vartheta}) = g(\boldsymbol{\vartheta})$ for $g \in \mathcal{T}_{nd}$ and $T_{nd}(g, \boldsymbol{\vartheta}) \in \mathcal{T}_{nd}$ for other $g \in C(\mathbb{T}^d)$.

⁴Here and later $a_n \sim b_n$ means that $0 < c_1 \leq a_n b_n^{-1} \leq c_2$ where c, c_1, c_2, \dots are positive constants, not depending on n ; they may denote different values in different formulae.

Theorem 2.12. For any linear trigonometric projection operator T_{nd} , one has

$$\frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} T_{nd}(g_{\mathbf{t}}, \boldsymbol{\vartheta} - \mathbf{t}) dt = S_{nd}(g, \boldsymbol{\vartheta}) \quad (g \in C(\mathbb{T}^d)),$$

$$\|T_{nd}\| \geq \|S_{nd}\|,$$

where $g_{\mathbf{t}}(\boldsymbol{\vartheta}) = g(\boldsymbol{\vartheta} + \mathbf{t})$ is the \mathbf{t} -translation operator.

Now we formulate a generalization of (2.15).

Theorem 2.13. If \mathcal{L}_{nd} is a projection of $C(I^d)$ onto \mathcal{P}_{nd} then

$$\|\mathcal{L}_{nd}\| \geq \frac{1}{2} \|S_{nd}\|.$$

Above, \mathcal{L}_{nd} is a projection of $C(I^d)$ ($:=$ the set of continuous functions of d -variables on $I^d = [-1, 1]^d$) onto \mathcal{P}_{nd} iff it is linear, $\mathcal{L}_{nd}(p) = p$ if $p \in \mathcal{P}_{nd}$ and $\mathcal{L}_{nd}(f) \in \mathcal{P}_{nd}$ for any $f \in C(I^d)$.

Proofs, further statements, references and some historical remarks about Part 2.8 are in the paper *László Szili* and *P. Vértesi* [19].

3. MEAN CONVERGENCE OF INTERPOLATION

3.1.

As it has turned out the estimation of the Lebesgue function

$$\lambda_n(X, x) = \sum_{k=1}^n |\ell_{kn}(X, x)|$$

is fundamental in getting “negative” (divergence)-type results for the Lagrange interpolation using the uniform (or maximum) norm.

These facts resulted that the attention turned to the *mean convergence* of interpolation. The first such result is due to *P. Erdős* and *P. Turán* [20] from 1937.

Theorem 3.1. For an arbitrary weight w and $f \in C$,

$$\lim_{n \rightarrow \infty} \int_{-1}^1 \{L_n(f, w, x) - f(x)\}^2 w(x) dx = 0.$$

Here and later w is a weight if $w \geq 0$ and $0 < \int_{-1}^1 w < \infty$; $L_n(f, w)$ is the Lagrange interpolation with nodes at on the roots of the corresponding orthonormal polynomials (ONP) $p_n(w)$.

During the years 1936–1939, *P. Erdős* and *P. Turán* wrote 3 fundamental papers “On interpolation I, II, III” ([20], [32], [33]; they appeared in 1937, 1938 and 1940). This survey will quote many problems and theorems of them. We strongly suggest to read these papers to the interested readers.

Using the Chebyshev roots, *P. Erdős* and *Ervin Feldheim* proved much more [21]:

Theorem 3.2. *Let $f \in C$ and $p > 1$. Then*

$$\lim_{n \rightarrow \infty} \int_{-1}^1 |f(x) - L_n(f, T, x)|^p \frac{1}{\sqrt{1-x^2}} dx = 0.$$

3.2.

Theorem 3.1 is a reasonable motivation of the problem (cf. *P. Erdős, Géza Freud, P. Turán* [23, Problem VIII], [24], [25]).

Does there exists a weight w and $f \in C$ such that

$$\overline{\lim}_{n \rightarrow \infty} \|f - L_n(f, w)\|_{p,w} = \infty$$

for every $p > 2$?

(Above $\|g\|_{p,w}$ stands for $\|gw^{1/p}\|_p$.)

After a lot of results proved by *Richard Askey, Paul Nevai* and others *Y. G. Shi* came to a new general idea where the nodes x_{kn} are *not necessarily* the roots of an orthogonal system $p_n(w)$. Namely he realized the surprising fact that for the mean convergence the expressions

$$(3.1) \quad \gamma_1(X, x) := \sum_{k=1}^n |x - x_{kn}(X, x)| |\ell_{kn}(X, x)|, \quad n \geq 1,$$

are fundamentals (instead of $\lambda_n(X, x) = \sum_{k=1}^n |\ell_{kn}(X, x)|$). Using many basic ideas of the proof of Theorem 2.3, he proves (among others)

Theorem 3.3. *Let $\varepsilon > 0$ be given. Then for any fixed interpolatory matrix $X \subset [-1, 1]$, there exists sets $H_n = H_n(\varepsilon, X)$ of measure $\leq \varepsilon$ such that*

$$(3.2) \quad \gamma_1(X, x) \geq \frac{\varepsilon}{24}, \quad n \geq 1,$$

whenever $x \in [-1, 1] \setminus H_n$. (cf. Theorem 2.3).

The above statement is a special case of [26, Theorem 1], the latter one uses the proof of the generalization of Theorem 2.3 (cf. *P. Vértesi* [27]).

Now, by [26, Theorem 1], *Y. G. Shi* ([26, Corollary 14]) obtains.

Theorem 3.4. *Let u and w be weights. If with a fixed $p_0 \geq 2$*

$$\left\| \frac{1}{\sqrt{w\sqrt{1-x^2}}} \right\|_{p,u} = \infty \quad \text{for every } p > p_0,$$

then there exists an $f \in C$ satisfying

$$\overline{\lim}_{n \rightarrow \infty} \|L_n(f, w)\|_{p,u} = \infty \quad \text{whenever } p > p_0.$$

This theorem obviously answers the (generalization of the) question raised at the beginning of Part 3.2. For other similar problems the reader may consult with [26].

4. CONVERGENCE BY RAISING THE DEGREE

4.1.

Motivated by *Lipót Fejér's* classical results (i.e, if the degree of the interpolation polynomial is about two times bigger than the number of interpolation points, then we can get convergence (cf. [28, Theorem XI])), *Erdős* raised the following question. Given $\varepsilon > 0$, suppose we interpolate at n nodes, but allow polynomials of degree at most $n(1 + \varepsilon)$. Under what conditions will they converge for all continuous function?

The first answer was given by himself in [29]. Namely, he proved:

Theorem 4.1. *If the absolute values of the fundamental polynomials $\ell_{kn}(X, x)$ are uniformly bounded in $x \in [-1, 1]$, k ($1 \leq k \leq n$) and $n \in \mathbb{N}$, then for every $\varepsilon > 0$ and $f \in C$ there exists a sequence of polynomials $\varphi_n = \varphi_n(x) = \varphi_n(f, \varepsilon, x)$ with*

- (i) $\deg \varphi_n \leq n(1 + \varepsilon)$,
- (ii) $\varphi_n(x_{kn}) = f(x_{kn})$, $1 \leq k \leq n$, $n \in \mathbb{N}$,
- (iii) $\lim_{n \rightarrow \infty} \|\varphi_n - f\| = 0$.

The complete answer for a more general system is in the paper of *Erdős, András Kroó* and *Szabados* [30].

Theorem 4.2. For every $f \in C$ and $\varepsilon > 0$, there exists a sequence of polynomials $p_n(f)$ of degree at most $n(1 + \varepsilon)$ such that

$$p_n(f, x_{k,n}) = f(x_{k,n}), \quad 1 \leq k \leq n,$$

and that

$$\|f - p_n(f)\| \leq cE_{[n(1+\varepsilon)]}(f)$$

holds for some $c > 0$, if and only if

$$(4.1) \quad \limsup_{n \rightarrow \infty} \frac{N_n(I_n)}{n|I_n|} \leq \frac{1}{\pi}$$

whenever I_n is a sequence of subintervals of I such that $\lim_{n \rightarrow \infty} n|I_n| = \infty$ and

$$(4.2) \quad \liminf_{n \rightarrow \infty} \left(n \min_{1 \leq k \leq n-1} (\vartheta_{k+1,n} - \vartheta_{n,k}) \right) > 0.$$

Here $N_n(I_n)$ is the number of the $\vartheta_{k,n}$ in $I_n \subset I$. Condition (4.1) ensures that the nodes are not too dense, and condition (4.2) says that adjacent nodes should not be too close.

5. WEIGHTED LAGRANGE INTERPOLATION, WEIGHTED LEBESGUE FUNCTION, WEIGHTED LEBESGUE CONSTANT

5.1.

Let f be a continuous function. If, instead of the interval $[-1, 1]$, we try to approximate it on \mathbb{R} , we have to deal with the obvious fact that polynomials (of degree ≥ 1) tend to infinity if $|x| \rightarrow \infty$. So to get a suitable approximation tool, we may try to moderate their growth applying proper weights.

If the weight $w(x) = e^{-Q(x)}$, $x \in \mathbb{R}$, satisfies

$$\lim_{|x| \rightarrow \infty} \frac{Q(x)}{\log |x|} = \infty,$$

as well as some other mild restrictions and the Akhiezer–Babenko–Carleson–Dzrbasjan relation

$$\int_{-\infty}^{\infty} \frac{Q(x)}{1+x^2} dx = \infty,$$

then for $f \in C(w, \mathbb{R})$, where

$$C(w, \mathbb{R}) := \{f; f \text{ is continuous on } \mathbb{R} \text{ and } \lim_{|x| \rightarrow \infty} f(x)w(x) = 0\},$$

we have, if $\|\cdot\|$ denotes now the supnorm on \mathbb{R} ,

$$E_n(f, w) := \inf_{p \in \mathcal{P}_n} \|(f - p)w\| \equiv \inf_{p \in \mathcal{P}_n} \|fw - pw\| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

So, instead of approximating $f \in C$ by $L_n(f, X)$ on $[-1, 1]$, we may estimate $\{f(x)w(x) - L_n(f, w, X, x)\}$ on the real line \mathbb{R} for $f \in C(w, \mathbb{R})$. Here $X \subset \mathbb{R}$,

$$t_k(x) := t_{kn}(w, X, x) := \frac{w(x)\omega_n(X, x)}{w(x_k)\omega'_n(X, x_k)(x - x_k)}, \quad 1 \leq k \leq n,$$

and

$$L_n(f, w, X, x) := \sum_{k=1}^n \{f(x_k)w(x_k)\} t_k(x), \quad n \in \mathbb{N}.$$

The Lebesgue estimate now has the form

$$(5.1) \quad |L_n(f, w, X, x) - f(x)w(x)| \leq \{\lambda_n(w, X, x) + 1\} E_{n-1}(f, w)$$

where the (*weighted*) *Lebesgue function* is defined by

$$(5.2) \quad \lambda_n(w, X, x) := \sum_{k=1}^n |t_k(w, X, x)|, \quad x \in \mathbb{R}, \quad n \in \mathbb{N};$$

the existence of $r_{n-1}(f, w)$ for which $E_{n-1}(f, w) = \|(f - r_{n-1})w\|$ is well-known.

Formula (5.2) implies the natural definition of the (*weighted*) *Lebesgue constant*

$$(5.3) \quad \Lambda_n(w, X) := \|\lambda_n(w, X, x)\|, \quad n \in \mathbb{N}.$$

Estimation (5.1) and its immediate consequence

$$\|L_n(f, w, X) - fw\| \leq \{\Lambda_n(w, X) + 1\} E_{n-1}(f, w), \quad n \in \mathbb{N},$$

show that, analogously to the classical case, the investigation of $\lambda_n(w, X, x)$ and $\Lambda_n(w, X)$ is of fundamental importance to get convergence-divergence results for the weighted Lagrange interpolation (cf. Part 2.1).

To expect reasonable estimations, as it turns out, we need a considerable knowledge about the weight $w(x)$ and on the behaviour of the ONP $p_n(w^2, x)$ corresponding to the weight w^2 .

5.2.

As *P. Nevai* writes in his instructive monograph [34, Part 4.15], about 40 years ago there was a great amount of information on orthogonal polynomials on infinite intervals, however as *Géza Freud* realized in the sixties, there had been a complete lack of *systematic treatment* of the general theory; the results were of mostly ad hoc nature. And *G. Freud*, in the last 10 years of his life, laid down the basic tools of the systematic investigation.

During the years a great number from the approximators and/or orthogonalists joined *G. Freud* and his work, including many Hungarians. As a result, today our knowledge is more comprehensive and more solid than before.

Now we introduce the so called Mhaskar–Rakmanov–Saff number, denoted by $a_n(w)$. $a_n(w)$ is a generalization of the number $q_n(w)$ defined by *G. Freud*. Instead of the definition we show a useful property of $a_n(w)$ and give an example (cf. [35]).

$$(5.4) \quad \begin{cases} \|r_n w\| = \max_{|x| \leq a_n(w)} |r_n(x)w(x)|, \\ \|r_n w\| > |r_n(x)w(x)| \quad \text{for } |x| > a_n(w) \end{cases}$$

if $r_n \in \mathcal{P}_n$ ($r_n \not\equiv 0$; $\|\cdot\|$ is the supnorm on \mathbb{R}) and that asymptotically (as $n \rightarrow \infty$) $a_n(w)$ is the smallest such number. Relation (5.4) may be formulated such that $r_n w$ “lives” on $[-a_n, a_n]$.

As an example, let $Q(x) = |x|^\alpha$. Then

$$q_n(w) \sim n^{1/\alpha} \quad \text{and} \quad a_n(w) = c(\alpha) n^{1/\alpha}, \quad \alpha > 1.$$

In 1972, *P. Erdős* defined (as today called) the Erdős weights. The prototype of $w \in \mathcal{E}$ (\mathcal{E} is the collection of the Erdős weights) is the case when $Q(x) = Q_{k,\alpha} = \exp_k(|x|^\alpha)$ ($k \geq 1$, $\alpha > 1$, $\exp_k := \exp(\exp(\dots))$, the k th iterated exponential); for other details on \mathcal{E} see [36], [37] and [35]. As an interesting and maybe surprising fact that generalizing the method and ideas of our common paper with *Erdős*, one can prove a statement on the *weighted* Lebesgue function $\lambda_n(w, X, x)$ (see *P. Vértesi* [38]).

Theorem 5.1. *Let $w \in \mathcal{E}$. If $\varepsilon > 0$ is an arbitrary fixed number, then for any interpolatory matrix $X \subset \mathbb{R}$ there exist sets $H_n = H_n(w, \varepsilon, X)$ with $|H_n| \leq 2a_n(w)\varepsilon$ such that*

$$\lambda_n(w, X, x) \geq \frac{\varepsilon}{3840} \log n$$

if $x \in [-a_n(w), a_n(w)] \setminus H_n$, $n \geq n_1(\varepsilon)$.

This statement is a complete analogue of Theorem 2.4. Roughly speaking, it says that the weighted Lebesgue function is at least $c \log n$ on a “big part” of $[-a_n, a_n]$ for arbitrary fixed $X \subset (-\infty, \infty)$ and $w \in \mathcal{E}$.

Without going into the details we remark that the previous consideration and statement can be developed for other weights (cf. [38]).

To finish this survey we quote another theorem on weighted approximation which corresponds to the result of Erdős from 1943 (see Theorem 4.1). Namely we have

Theorem 5.2. *Let $w \in \mathcal{E}$. If $|t_{kn}(w, X, x)| \leq A$ uniformly in $x \in \mathbb{R}$, k and n , then for every $\varepsilon > 0$ and to every $f \in C(w^{1+\varepsilon}, \mathbb{R})$, there exists a sequence of polynomials $\varphi_\Delta(x) = \varphi_\Delta(f, \varepsilon, x) \in \mathcal{P}_\Delta$ such that*

- (i) $\Delta \leq n(1 + \varepsilon + c \varepsilon n^{-2/3})$,
- (ii) $\varphi_\Delta(x_{kn}) = f(x_{kn})$, $1 \leq k \leq n$, $n \in \mathbb{N}$,
- (iii) $\|w^{1+\varepsilon}(f - \varphi_\Delta)\| \leq cE_\Delta(f, w^{1+\varepsilon})$.

The proof and similar results using other exponents $Q(x)$ are in *L. Szili* and *P. Vértesi* [39] and [40].

REFERENCES

- [1] P. Vértesi, *Classical (unweighted) and weighted interpolation*, in the book “A Panorama of Hungarian Mathematics in the Twentieth Century. I”, Bolyai Society Mathematical Studies **14** (2005), 71–117.
- [2] T. Erdélyi and P. Vértesi, In memoriam Paul Erdős, (1913–1986), *J. Approx. Theory* **94** (1998), 1–41.
- [3] L. Babai, C. Pommerance and P. Vértesi, The Mathematics of Paul Erdős, Notices of the AMS **45**(1) (1998), 19–31.
- [4] P. Erdős, Problems and results on the theory of interpolation, I, *Acta Math. Acad. Sci. Hungar.*, **9** (1958), 381–388.
- [5] P. Erdős and P. Vértesi, On the Lebesgue function of interpolation, in: *Functional Analysis and Approximation* (P. L. Butzer, B. Sz.-Nagy, E. Görlich, eds.), ISNM, **60**, Birkhäuser (1981), 299–309.
- [6] P. Erdős, Problems and results on the theory of interpolation, II, *Acta Math. Acad. Sci. Hungar.*, **12** (1961), 235–244.
- [7] C. de Boor and A. Pinkus, Proof of the conjectures of Bernstein and Erdős concerning the optimal nodes for polynomial interpolation, *J. Approx. Theory*, **24** (1978), 289–303.
- [8] T. A. Kilgore, A characterization of the Lagrange interpolating projection with minimal Tchebycheff norm, *J. Approx. Theory*, **24** (1978), 273–288.
- [9] P. Vértesi, Optimal Lebesgue constant for Lagrange interpolation, *SIAM J. Numer. Anal.*, **27** (1990), 1322–1331.

- [10] P. Erdős and G. Grünwald, Über die arithmetischen Mittelwerte der Lagrangeschen Interpolationspolynome, *Studia Math.*, **7** (1938), 82–95.
- [11] P. Erdős and G. Halász, On the arithmetic means of Lagrange interpolation, in: *Approximation Theory* (J. Szabados, K. Tandori, eds.), Colloq. Math. Soc. J. Bolyai, **58** (1991), pp. 263–274.
- [12] P. Erdős, Some theorems and remarks on interpolation, *Acta Sci. Math. (Szeged)*, **12** (1950), 11–17.
- [13] P. Erdős and P. Vértesi, On the almost everywhere divergence of Lagrange interpolation of polynomials for arbitrary systems of nodes, *Acta Math. Acad. Sci. Hungar.*, **36** (1980), 71–89.
- [14] P. Erdős and P. Vértesi, Correction of some misprints in our paper: “On the almost everywhere divergence of Lagrange interpolation polynomials for arbitrary systems of nodes” [*Acta Math. Acad. Sci. Hungar.*, **36**, no. 1–2 (1980), 71–89], *Acta Math. Acad. Sci. Hungar.* **38** (1981), 263.
- [15] P. Erdős and P. Vértesi, On the almost everywhere divergence of Lagrange interpolation, in: *Approximation and Function Spaces* (Gdańsk, 1979), North-Holland (Amsterdam–New York, 1981), pp. 270–278.
- [16] P. Erdős and P. Turán, On the role of the Lebesgue function the theory of Lagrange interpolation, *Acta Math. Acad. Sci. Hungar.*, **6** (1955), 47–66.
- [17] G. Halász, On projections into the space of trigonometric polynomials, *Acta Sci. Math. (Szeged)*, **57** (1993), 353–366.
- [18] R.A. DeVore and G.G. Lorentz, *Constructive Approximation*, Springer-Verlag, Berlin (1993).
- [19] L. Szili and P. Vértesi, On multivariate projection operators, *Journal of Approximation Theory*, **159** (2009), 154–164.
- [20] P. Erdős, Problems and results on the theory of interpolation, I, *Acta Math. Acad. Sci. Hungar.*, **9** (1958), 381–388.
- [21] P. Erdős and E. Feldheim, Sur le mode de convergence pour l’interpolation de Lagrange, *C. R. Acad. Sci. Paris*, **203** (1936), 913–915.
- [22] E. Feldheim, Sur le mode de convergence dans l’interpolation de Lagrange, *Dokl. Nauk. USSR*, **14** (1937), 327–331.
- [23] P. Turán, On some open problems of approximation theory, *J. Approx. Theory*, **29** (1980), 23–85.
- [24] P. Erdős and P. Turán, On interpolation, I. Quadrature and mean convergence in the Lagrange interpolation, *Ann. of Math.*, **38** (1937), 142–155.
- [25] L. Fejér, Sur les fonctions bornées et intégrables, *C. R. Acad. Sci. Paris*, **131** (1900), 984–987.
- [26] Y. G. Shi, Bounds and inequalities for general orthogonal polynomials on finite intervals, *J. Approx. Theory*, **73** (1993), 303–319.
- [27] P. Vértesi, Lebesgue function type sums of Hermite interpolations, *Acta Math. Acad. Sci. Hungar.*, **64** (1994), 341–349.
- [28] L. Fejér, Über Interpolation, *Göttinger Nachrichten* (1916), 66–91.
- [29] P. Erdős, On some convergence properties in the interpolation polynomials, *Ann. of Math.*, **44** (1943), 330–337.

- [30] P. Erdős, A. Kroó and J. Szabados, On convergent interpolatory polynomials, *J. Approx. Theory*, **58** (1989), 232–241.
- [31] Y.G. Shi, On an extremal problem concerning interpolation, *Acta Sci. Math. (Szeged)*, **65** (1999), 567–575.
- [32] P. Erdős and P. Turán, On interpolation, II., *Ann. of Math.*, **39** (1938), 703–724.
- [33] P. Erdős and P. Turán, On interpolation, III., *Ann. of Math.*, **41** (1940), 510–553.
- [34] P. Nevai, Géza Freud, orthogonal polynomials and Christoffel functions: A case study, *J. Approx. Th.*, **48** (1986), 3–167.
- [35] A.L. Levin and D.S. Lubinsky, *Orthogonal Polynomials for Exponential Weights*, Springer, New York, 2001.
- [36] P. Erdős, On the distribution of roots of orthogonal polynomials, in: *Proceedings of the Conference on the Constructive Theory of Functions*, (eds. G. Alexits et al.). Akadémiai Kiadó, Budapest, 1972, pp.145–150.
- [37] P. Erdős and G. Freud, On orthogonal polynomials with regularly distributed zeros, *Proc. London Math. Soc.* **29** (1974), 521–537.
- [38] P. Vértesi, On the Lebesgue function of weighted Lagrange interpolation, II., *J. Austral Math. Soc. (Series A)*, **65** (1998), 145–162.
- [39] P. Vértesi, An Erdős type convergence process in weighted interpolation. I. (Freud type weights), *Acta Math. Hungar.*, **91** (2000), 195–215.
- [40] L. Szili and P. Vértesi, An Erdős type convergence process in weighted interpolation, II, *Acta Math. Acad. Sci. Hungar.*, **98** (2003), 129–162.
- [41] E.M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, Princeton (1971).
- [42] A. Zygmund, *Trigonometric Series*, Vol. I and Vol. II, Cambridge University Press, Cambridge (1959).
- [43] J. G. Herriot, Nörlund summability of multiple Fourier series, *Duke Math. J.*, **11** (1944), 735–754.
- [44] H. Berens and Y. Xu, Fejér means for multivariate Fourier series, *Mat. Z.* **221** (1996), 449–465.
- [45] R. A. DeVore and G. G. Lorentz, *Constructive Approximation*, Springer-Verlag, Berlin, 1993.
- [46] J. Szabados and P. Vértesi, *Interpolation of Functions*, World Science Publisher, Singapore, New Jersey, London, Hong Kong, 1990.
- [47] P. Vértesi, *Some recent results on interpolation*, in: A tribute to Paul Erdős, Cambridge Univ. Press, 1990, 451–457.

Péter Vértesi

*Alfréd Rényi Institute of Mathematics,
Hungarian Academy of Sciences,
Budapest, Reáltanoda u. 13–15, H-1053,
Hungary*

e-mail: vertesi.peter@renyi.mta.hu