COMBINATORICA
Bolyai Society – Springer-Verlag

COMBINATORICA **20** (2) (2000) 203–218

TILING TURÁN THEOREMS

JÁNOS KOMLÓS

Received June 2, 1998

We discuss degree conditions for finding many disjoint copies of a fixed graph H in a large graph G.

Notation

$V(G)$ and $E(G)$ denote the vertex-set and the edge-set of the graph G, and we write $v(G)=|V(G)|$ (order of G) and $e(G)=|E(G)|$ (size of G). G_n denotes n-graphs, that is, graphs of order n. $N(v)$ is the set of neighbours of $v\in V$. Hence $|N(v)|=deg(v)=deg_G(v)$ is the degree of v. $\delta(G)$ stands for the minimum, and $\Delta(G)$ for the maximum degree in G. When A and B are disjoint subsets of $V(G)$, we denote by $e(A,B)$ the number of edges of G with one endpoint in A and the other in B. We write $\chi(G)$ for the chromatic number of G. For graphs G and H, $H\subset G$ means that G has a subgraph isomorphic to H. ex (for 'extremal') stands for the Turán function: $\mathrm{ex}(n,H)=\max\{e(G):v(G)=n, H\not\subset G\}$. Given two graphs, H and G, an H-matching in G (or a tiling of G with H) is a subgraph of G consisting of vertex-disjoint copies of H. An H-factor in G is a (complete) tiling of G with $\lfloor v(G)/v(H)\rfloor$ copies of H.

Mathematics Subject Classification (1991): 15A52, 11L03; 15A36, 11C20
Research supported in part by the NSF grant DMS-9801396.

0209–9683/100/$6.00 ©2000 János Bolyai Mathematical Society

1. Introduction

Perhaps the two most important theorems in extremal graph theory are Turán's theorem and the Erdős-Stone-Simonovits theorem (see below). They determine tight conditions on the existence of a fixed graph H (or one in a fixed family of graphs) as subgraphs in a large host-graph G.

Theorem 1 **(Turán 1941 [23]).** *Let G be a graph on n vertices and r a positive integer. If*

$$e(G) > \left(1 - \frac{1}{r-1}\right) \frac{n^2}{2}$$

then $K_r \subset G$.

(In fact, this is a useful but somewhat weakened form of the original theorem of Turán.)

Theorem 2 **(Erdős-Stone-Simonovits 1946/1966 [15, 14]).** *For every graph H and $\varepsilon > 0$ there is a threshold $n_0 = n_0(H, \varepsilon)$ such that the following holds for all $n \geq n_0$ and all n-graphs G. If*

$$e(G) > \left(1 - \frac{1}{\chi(H) - 1} + \varepsilon\right) \frac{n^2}{2}$$

then $H \subset G$.

This leads to the following limit theorem.

Theorem 3 **(Fundamental Theorem of Extremal Graph Theory).**

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n,H)}{\binom{n}{2}} = 1 - \frac{1}{\chi(H)-1} \tag{1}$$

The important message here is that the chromatic number of H is the quantity that matters most in extremal graph theory (as opposed to random graph theory, where the average degree of H – or more precisely $MAD(H)$ (maximum average degree) – is the relevant characteristic).

In this note we investigate the tiling problem, in which we wish to find many vertex-disjoint copies of H in G, or even a complete tiling of G with H. For tiling problems a (reasonably) large number of edges in G is not sufficient any more, the natural conditions set lower bounds on all degrees. Dirac's theorem on Hamilton paths [10] solves the 1-factor problem ($H = K_2$), the Corrádi-Hajnal theorem [8] handles triangle-factors ($H = K_3$), and finally the (very hard) Hajnal-Szemerédi theorem settles the K_r-factor problem for all r.

Theorem 4 (Hajnal-Szemerédi 1970 [16]). *Let G be an n-graph. If*

$$\delta(G) \geq \left(1 - \frac{1}{r}\right) n$$

then G has a K_r-factor.

Note that $1/(r-1)$ in Turán's theorem changed now to $1/r$; it is obviously harder to get a whole K_r-factor in G than just a single copy of K_r. Now what should r in this limit be replaced by for a general graph H ? The natural guess is still the chromatic number of H. This is discussed in the next section.

1.1. The Alon-Yuster theorems

In the 90's Noga Alon and Raphael Yuster extended the Hajnal-Szemerédi theorem to arbitrary H in various ways.

Theorem 5 (Alon-Yuster 1992 [3]). *For every graph H and $\varepsilon > 0$ there is a threshold $n_0 = n_0(H, \varepsilon)$ such that, if $n \geq n_0$ and a graph G_n satisfies the degree condition*

$$\delta(G_n) \geq \left(1 - \frac{1}{\chi(H)} + \varepsilon\right) n, \tag{2}$$

then G_n contains an H-matching that covers at least $(1-\varepsilon)n$ vertices.

Remark. There are two slacks εn in this theorem, the extra εn in the degree condition (2), and the relaxed requirement that as many as εn vertices of G_n may not be covered by the H-matching. These two slacks are very different. One can simply set $\varepsilon = 0$ in (2), but it is much harder to get rid of the slack in the conclusion. This was done in the following theorem.

Theorem 6 (Alon-Yuster 1996 [4]). *If G_n satisfies (2) and $n \geq n_0(H, \varepsilon)$, then G_n has an H-factor.*

Several examples show that both slacks cannot be eliminated simultaneously, but Alon and Yuster conjectured that either one can be set to zero and the other one to a constant.

Conjecture 1 (Alon-Yuster 1992 [3]). For every graph H there is a constant K such that, if G_n satisfies

$$\delta(G_n) \geq \left(1 - \frac{1}{\chi(H)}\right) n, \tag{3}$$

then it has an H-matching that covers all but at most K vertices.

Conjecture 2 (Alon-Yuster 1996 [4]). For every graph H there is a constant K such that, if G_n satisfies

$$\delta(G_n) \geq \left(1 - \frac{1}{\chi(H)}\right) n + K, \tag{4}$$

then it has an H-factor.

(The conjecture has been proved by Komlós, Sárközy and Szemerédi.)

Some authors use the notation $T(n,H) = \text{ex}(n,H)+1 = \min\{m : (v(G) = n, e(G) \geq m) \to (H \subset G)\}$ for the Turán numbers. We define the following analogues and call them **Tiling Turán numbers:**

$$TT(n,H) = \min\{m : \delta(G_n) \geq m \text{ implies that } G_n \text{ has an } H\text{-factor}\}.$$

Thus Theorem 6 says that

$$TT(n,H) \leq \left(1 - \frac{1}{\chi(H)}\right) n + o(n), \tag{5}$$

and Conjecture 2 can be restated as

$$TT(n,H) \leq \left(1 - \frac{1}{\chi(H)}\right) n + O(1). \tag{6}$$

Alon and Yuster also remark that these bounds are essentially best possible. Thus, the asymptotic behavior of $TT(n,H)$ seems to be completely understood: the limit of $TT(n,H)/n$ seems to be $1-1/\chi(H)$. This suggests the same message as Theorem 3 did for the Turán problem: the chromatic number of H probably is the relevant quantity for tiling problems as well. We discuss this remark a little later.

Of course, it is still possible to fine-tune the known estimates for $TT(n,H)$, e.g. by characterizing those graphs H for which the error term K in Conjecture 2 is actually 0, that is, $TT(n,H) \leq (1-1/\chi(H))n$. Dirac's theorem and the Hajnal-Szemerédi theorem show that paths and cliques are like that. Another example is given in a conjecture of Erdős and Faudree [13], which says that C_4 also has error term 0. In fact, for cliques and C_4 the inequality is an equality. But a full characterization of these graphs is probably hard.

However, this fine-tuning is all one-sided (upper bounds) and we should not neglect the need for matching lower bounds.

1.2. Limit behavior

Let us now return to the less subtle question of determining the limit of $TT(n,H)/n$. It is indeed equal to $1/2$ for all bipartite graphs H, as shown by Theorem 6 and the following two lower bounds:

- **Example 1.** Let $v(H) \geq 3$, let n be divisible by $v(H)$, and let G_n consists of two cliques of orders k and $n-k$ where $k = \lfloor (n-1)/2 \rfloor$. Then $\delta(G_n) = k-1$, but G_n does not contain an H-factor, since k is not divisible by $v(H)$. Hence $TT(n,H) \geq k = \lfloor (n-1)/2 \rfloor$.

- **Example 2.** Let $H = K_r$, and let G_n be a complete r-partite graph with the following color-class sizes: if $n = qr + R$ where $0 \leq R < r$, then the smallest color-class has $q-1$ vertices, and the rest are as evenly distributed as possible. Then G_n does not contain a K_r-factor, and hence $TT(n,K_r) \geq \delta(G_n) + 1 = n + 1 -$ the largest class size $\geq ((r-1)n-1)/r$.

Several similar examples (see [3,4]) with various chromatic numbers may also suggest that $TT(n,H)$ may be very close to (even within a constant of) $(1-1/\chi(H))n$ for any graph H.

The first crack on this seemingly perfect picture is given by the El-Zahar conjecture.

Conjecture 3 (El-Zahar 1984 [11]). Let $n_1 + n_2 + \ldots + n_k = n$ and let $\delta(G_n) \geq \sum \lceil n_i/2 \rceil$. Then G_n contains k vertex-disjoint cycles of orders $n_1, \ldots, n_k$.

This fascinating conjecture was proved recently by Sarmad Abbasi [1]. In particular, let n be divisible by ℓ and let $H = C_\ell$, the cycle on ℓ vertices. If ℓ is even then $TT(n,H) = n/2$ as expected. But if ℓ is odd then $TT(n,H) = n(l+1)/(2l)$, which is much less than $(1-1/\chi(H))n = 2n/3$ if $\ell > 3$. Hence the upper bounds (5) and (6) may be way off for some graphs with chromatic number greater than 2.

An even more disturbing phenomenon is the discrepancy, even for some bipartite graphs, between perfect tiling and almost perfect tiling. The ordinary Turán function ex is "continuous" in that the degrees (or total number of edges) needed to guarantee a copy of H in G_n is not much different from those needed for 100 copies or even $o(n)$ copies of H. However, this is not the case with tiling. There is a big difference between the two Alon-Yuster theorems, Theorems 5 and 6. While the latter one is best possible for quite a few graphs, including all bipartite graphs, the first one is definitely not sharp for many of them, e.g. for bipartite graphs with unequal color-classes.

An illuminating example is $H=P_3$, a path of order 3 (length 2). (The covering problem with copies of P_3 was settled by Enomoto, Kaneko and Tuza in [12].) If we drop the condition that 3 divides n in Example 1 above, then $TT(n,P_3)\geq n/2-1$ is not necessarily true any more. Example 1 works when $n=6k$ or $n=6k+4$, but not when $n=6k+2$. (The reason is that $0\equiv 1+2 \pmod 3$ and $1\equiv 2+2 \pmod 3$, but there are no x and y such that $x+y\equiv 2 \pmod 3$, $0\leq x,y<3$, and $x+y>2$.)

While $TT(n,P_3)=n/2-1$ for $n=6k$, for numbers of the form $n=6k+2$ we have the much smaller $TT(n,P_3)=2k=(n-2)/3$. This sudden drop caused by number-theoretic rather than graph-theoretic reasons, is unsettling.

The erratic behavior for $TT(n,H)$ as a function of n was created by the particular convention we made: when $n=qv(H)+R$, $0\leq R<v(H)$, then we tolerate R leftover vertices but not a few more. The flexibility we showed in extending the definition of TT from values of n divisible by $v(H)$ to other values, should be stretched a little further by allowing for a few additional leftover vertices than what divisibility demands. While for some n a perfect covering with P_3-s does indeed need $\delta\sim n/2$, a minimum degree $\delta\sim n/3$ is sufficient for covering $n-4$ vertices with P_3-s regardless of the particular form of n. **To cover a graph by copies of H almost perfectly can be much easier than to cover it perfectly.**

Let us embed these two extreme questions into a continuous range of tiling problems. This will lead to a more consistent overall picture. We will see that Example 1 above is more or less inconsequential, and the extremal graphs are similar to Example 2.

We define $TT(n,H,M)$ to be the minimum number m such that, if G_n is an n-graph with minimum degree $\delta(G_n)\geq m$, then there is an H-matching covering at least M vertices in G_n. For example, Theorem 5 says that $TT(n,H,n-o(n))\leq(1-1/\chi(H))n+o(n)$.

Given a real number x, $0<x<1$, let us study $TT(n,H,xn)$, the minimum degree needed to guarantee that at least an x proportion of the vertices of G_n are covered by (vertex-disjoint) copies of H. It turns out that **the function**

$$f_H(x)=\lim_{n\to\infty}\frac{1}{n}TT(n,H,xn)$$

is linear in x (this is quite trivial for $H=K_r$, but not at all easy for a general H), and it is natural to define

$$f_H(0)=\lim_{x\downarrow 0}f_H(x) \quad\text{and}\quad f_H(1)=\lim_{x\uparrow 1}f_H(x).$$

Obviously,

$$f_H(0) = 1 - \frac{1}{\chi(H)-1} \quad \text{(the Turán density)},$$

and Theorem 5 shows that

$$f_H(1) \le 1 - \frac{1}{\chi(H)}.$$

In the next section we give lower bounds. In Section 3 we will explicitely determine the function $f_H(x)$ for every H, and see that it is related to coloring properties of H.

2. Lower bounds

What quantity can replace $\chi = \chi(H)$ in (2) (or (3)(5)(6)). (Below we will write r for $\chi(H)$.) An obvious obstruction is that one cannot embed a graph H into a graph G_n of lower chromatic number. That is, we certainly cannot replace χ with anything less than $\chi - 1$ as the counterexample $G_n = K_{r-1}(n/(r-1))$ shows. But there may be a room for improvement between χ and $\chi - 1$.

To get a more subtle obstruction, let us try to tile complete r-partite graphs G_n with copies of H. This is a natural way to get good lower bounds, because it is reasonable to expect that the extremal graphs for the tiling problem, just as in the case of the Turán problem, are close to complete r-partite graphs. Indeed, all lower bounds in this section are obtained by using such graphs.

First we have one more notation: we write P^r for the set of monotone probability vectors of dimension r, that is, $P^r = \{\underline{\alpha} \in [0,1]^r : \alpha_1 \le \alpha_2 \le \ldots \le \alpha_r,\ \sum \alpha_i = 1\}$. For $\underline{\alpha}, \underline{\beta} \in P^r$, we write $\underline{\alpha} \prec \underline{\beta}$ (or $\underline{\beta} \succ \underline{\alpha}$) if $\underline{\beta}$ dominates $\underline{\alpha}$, that is, if

$$\sum_{i=1}^{k} \alpha_i \le \sum_{i=1}^{k} \beta_i, \quad k = 1, 2, \ldots, r-1. \tag{7}$$

(Note that this is different from the standard notation used in [20]: $\underline{\alpha} \prec \underline{\beta}$ here is $\underline{\beta} \prec \underline{\alpha}$ there.) Given a coloring of a graph G_n using $r = \chi(G_n)$ colors with color-class sizes $n_1 \le \ldots \le n_r$, the color-vector of the coloring is the vector $\underline{\alpha} \in P^r$ defined by $\alpha_i = n_i/n$.

Now, let G_n be a complete r-partite graph with color-class sizes $\beta_i n$, where $\underline{\beta} \in P^r$. We want to see under what conditions on $\underline{\beta}$ can G_n have an H-factor, or at least an almost H-factor. It is quite clear that the necessary and sufficient condition for this is that $\underline{\beta}$ is in the convex hull of the color-vectors $\underline{\alpha}$ of H and their permuted copies (the vectors whose coordinates are those of $\underline{\alpha}$ permuted). When H has only one color-vector $\underline{\alpha}$ then this turns out to be equivalent to $\underline{\beta} \succ \underline{\alpha}$ (see the proof of Lemma 12), but there is no such simple characterization in the case when H has many essentially different r-colorings.

Here is a condition that is necessary for any H: If β_1 is smaller than even the smallest possible α_1 in any color-vector $\underline{\alpha}$ of H, then G_n certainly cannot have an H-factor. (Similarly, in order to have an H-factor, $\beta_1 + \beta_2$ must be at least as large as the smallest possible $\alpha_1 + \alpha_2$, and so on, but these conditions turn out to be less relevant.) Now $\delta(G_n) = (1-\beta_r)n$, and so we get the lower bound

$$TT(n,H) \geq \sup\{(1-\beta_r)n : \underline{\beta} \in P^r, \text{ and } \beta_1 < \alpha_1 \text{ for all color-vectors } \underline{\alpha} \text{ of } H\}. \tag{8}$$

It is easy to see that the right-hand side here is equal to $1-(1-\min \alpha_1)/(r-1)$. Thus, we introduce the following quantity which measures in a way how color-critical H is.

Definition 1. For an r-chromatic graph H on h vertices we write $\sigma = \sigma(H)$ for the smallest possible color-class size in any r-coloring of H. The **critical chromatic number** of H is the number

$$\chi_{cr}(H) = (r-1)h/(h-\sigma). \tag{9}$$

It is easy to see that $\chi - 1 < \chi_{cr} \leq \chi$, and $\chi_{cr} = \chi = r$ if and only if every r-coloring of H has equal color-class sizes.

Examples. $\chi_{cr}(K_r) = r = \chi(K_r)$, $\chi_{cr}(C_{2k}) = 2 = \chi(C_{2k})$, $\chi_{cr}(C_{2k+1}) = 2+1/k$ (see the El-Zahar conjecture), for the Petersen graph P we have $\chi_{cr}(P) = 2+6/7$, and for a complete bipartite graph B with color-class sizes $a \leq b$ we have $\chi_{cr}(B) = 1 + a/b$.

Using χ_{cr}, (8) can be restated as follows:

$$TT(n,H) \geq \left(1 - \frac{1}{\chi_{cr}(H)}\right)n.$$

The following is a more general bound.

Theorem 7 (General Lower Bound). *Let H have parameters $\chi = \chi(H)$ and $\chi_{cr} = \chi_{cr}(H)$. Then, for all $0 < M \leq n$,*

$$TT(n,H,M) \geq M\left(1 - \frac{1}{\chi_{cr}}\right) + (n-M)\left(1 - \frac{1}{\chi - 1}\right) \tag{10}$$

whence

$$f_H(x) \geq x\left(1-\frac{1}{\chi_{cr}}\right)+(1-x)\left(1-\frac{1}{\chi-1}\right) \quad \textit{for} \ \ 0<x<1,$$

and

$$f_H(1) \geq 1-\frac{1}{\chi_{cr}}$$

Proof. Let us write $h=v(H)$, $r=\chi(H)$, let m be the smallest integer strictly greater than

$$\frac{n-M\sigma/h}{r-1}=\frac{M}{\chi_{cr}}+\frac{n-M}{r-1}$$

and let G_n be the complete r-partite graph with $r-1$ color-classes of size m and a leftover color-class. The relations $M\leq n$ and $\sigma\leq h/r$ imply that this leftover class is the smallest one. Since it has size $n-(r-1)m<M\sigma/h$, G_n cannot have an H-matching that covers at least M vertices. Thus,

$$\begin{aligned} TT(n,H,M) &\geq \delta(G_n)+1=n-(m-1)\geq n-\left(\frac{M}{\chi_{cr}}+\frac{n-M}{r-1}\right) \\ &= M\left(1-\frac{1}{\chi_{cr}}\right)+(n-M)\left(1-\frac{1}{r-1}\right) \end{aligned}$$

∎

In the next section we present the main theorem which says that the lower bound (10) is in fact an asymptotic equality.

3. Almost-perfect matchings

Before stating the central theorem of the paper, we recall (Theorem 3) that

$$f_H(0)=1-\frac{1}{\chi(H)-1}$$

Theorem 8 (Main Tiling Theorem). *Given H with chromatic number r and critical chromatic number χ_{cr}, we write*

$$g(x)=x\left(1-\frac{1}{\chi_{cr}}\right)+(1-x)\left(1-\frac{1}{r-1}\right) \quad \textit{for} \ \ x\in(0,1).$$

Then, for all $x\in(0,1)$,

$$f_H(x):=\lim_{n\to\infty}\frac{1}{n}TT(n,H,xn)=g(x).$$

In particular,

$$f_H(1) = \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}TT(n,H,(1-\varepsilon)n) = 1-\frac{1}{\chi_{cr}(H)}$$

that is, for every graph H and $\varepsilon>0$ there is a threshold $n_0=n_0(H,\varepsilon)$ such that, if $n\geq n_0$ and a graph G_n satisfies the degree condition

$$\delta(G_n)\geq\left(1-\frac{1}{\chi_{cr}(H)}\right)n, \tag{11}$$

then G_n contains an H-matching that covers all but at most εn vertices.

The following bound is an improved form of Conjecture 1 (but not of Conjecture 2!).

Conjecture 4 (see [17]). For every graph H there is a constant K such that, if G_n is a graph satisfying (11), then G_n contains an H-matching that covers all but at most K vertices.

This is best possible for every H. Hence,

$$TT(n,H,n-K)=\left(1-\frac{1}{\chi_{cr}(H)}\right)n+O(1).$$

An 'El-Zahar form' of the conjecture would say that if H_i are graphs with $\sum v(H_i)\leq n$, and G_n satisfies

$$\delta(G_n)\geq\sum_i v(H_i)\left(1-\frac{1}{\chi_{cr}(H_i)}\right)$$

then G_n contains, as a subgraph, the vertex-disjoint union of the H_i. While this is probably true for the union of many small graphs H_i, it is false for one single, large, expanding bipartite graph H, even if we replace χ_{cr} by χ.

4. The proof

4.1. The Regularity Lemma

In this section, we collect all the information we need here about Szemerédi's Regularity Lemma. For more, see the surveys [19, 17].

In a bipartite graph $G=(A,B,E)$ (A and B are the color classes), the *density* is defined as

$$d(A,B)=\frac{e(A,B)}{|A|\cdot|B|}$$

We say that $G=(A,B,E)$ is an ε-*regular* pair (or more often we say that (A,B) is an ε-regular pair) if

$$X\subset A, |X|>\varepsilon|A|,\ Y\subset B, |Y|>\varepsilon|B| \quad \text{imply} \quad |d(X,Y)-d(A,B)|<\varepsilon.$$

We say that $G=(A,B,E)$ is an (ε,δ)-*super-regular* pair if

$$X\subset A, |X|>\varepsilon|A|,\ Y\subset B, |Y|>\varepsilon|B| \quad \text{imply} \quad e(X,Y)>\delta|X||Y|;$$

furthermore, deg $(a)>\delta|B|$ for all $a\in A$, and deg $(b)>\delta|A|$ for all $b\in B$. We will use the following consequence of Szemerédi's Regularity lemma [22].

Theorem 9 (Regularity Lemma: degree form). *For every $\varepsilon>0$ there is an $M=M(\varepsilon)$ such that, if $G=(V,E)$ is any graph and $d\in[0,1]$ is any real number, then there is a partition of the vertex-set V into $k+1$ 'clusters' $V_0,V_1,\dots,V_k$, and there is a subgraph $G'\subset G$ with the following properties:*

- $k\le M$
- $|V_0|\le\varepsilon|V|$
- *all clusters V_i, $i\ge 1$, are of the same size $N\le\lceil\varepsilon|V|\rceil$*
- $deg_{G'}(v)>deg_G(v)-(d+\varepsilon)|V|$ *for all* $v\in V$
- *for each $i\ge1$, $G'|_{V_i}$ is empty*
- all *pairs $G'|_{V_i\times V_j}$ $(1\le i<j\le k)$ are ε-regular, each with a density either 0 or exceeding d.*

Given such an ε-regular partition, we define the *reduced graph* R on $\{V_1,V_2,\dots,V_k\}$ by connecting V_i and V_j if $G'|_{V_i\times V_j}$ (G' restricted to (V_i,V_j)) has positive density (and hence a density greater than d).

The next property, stating that large subgraphs of a regular pair are regular, is easy to see from the definition of regularity.

Fact 10 (Slicing Lemma). Let (A,B) be an ε-regular pair with density d, and, for some $\alpha>\varepsilon$, let $A'\subset A$, $|A'|\ge\alpha|A|$, $B'\subset B$, $|B'|\ge\alpha|B|$. Then (A',B') is an ε'-regular pair with $\varepsilon'=\max\{\varepsilon/\alpha,2\varepsilon\}$, and for its density d' we have $|d'-d|<\varepsilon$.

As a consequence, if we subdivide into ℓ equal parts each cluster in an ε-regular partition, the obtained new partition is still regular, but the parameter ε may have changed to $\varepsilon'=\ell\varepsilon$, the new reduced graph has now ℓ times more vertices, and the edges in the old reduced graph were replaced by copies of the complete bipartite graph $K_{\ell,\ell}$.

We also need a tool which is a special case of the Key Lemma in [19,17].

Lemma 11. *Given $d,\gamma>0$, and two graphs R and H, there is a positive ε such that the following holds for all positive integers m. Let us construct two graphs as follows: The graph $R(m)$ on $n=mv(R)$ vertices is obtained from R by replacing every vertex of R by m vertices, and replacing the edges of R with copies of the complete bipartite graph $K_{m,m}$, while the graph G is obtained similarly except that now we replace the edges of R with ε-regular pairs of density at least d. If $R(m)$ contains an H-matching with at least $(1-\gamma)n$ vertices then so does G.*

4.2. H-factors in complete r-partite graphs

Lemma 12. *If the r-chromatic graph H has a color-vector $\underline{\alpha}$ and G_n is a complete r-partite graph with color-vector $\underline{\beta}\succ\underline{\alpha}$, then G_n has an H-factor covering more than $n-K(r)v(H)$ vertices (where the constant $K(r)$ depends only on r).*

We will use the following simple lemma.

Lemma 13. *Let $\underline{a}^1,\underline{a}^2,\dots,\underline{a}^{r!}$ be the vectors obtained from $\underline{\alpha}$ by permuting the coordinates in all possible ways. Then $\underline{\beta}$ is a convex combination of these vectors: $\underline{\beta}=\sum x_i\underline{a}^i$, $x_i\ge 0$, $\sum x_i=1$.*

Proof. While a direct proof would also be easy, the lemma follows from standard theorems about doubly stochastic matrices (see in [20]). By a theorem of Hardy, Littlewood and Pólya, $\underline{\beta}\succ\underline{\alpha}$ if and only if there is a doubly stochastic matrix M such that $\underline{\beta}=M\underline{\alpha}$. Also, by Birkhoff's theorem, M is a convex combination of permutation matrices, $M=\sum x_iP_i$, and hence $\underline{\beta}=\sum x_iP_i\underline{\alpha}=\sum x_i\underline{a}^i$. ∎

Proof of Lemma 12. Let us write $\underline{\beta}$ as a convex combination $\underline{\beta}=\sum x_i\underline{a}^i$. Rounding all coefficients x_i down to $\lfloor x_i\rfloor$, we get the lemma with $K(r)=r!$. ∎

Remark by Endre Boros. *Lemma 12 is true even with $K(r)=r$.*
Proof. The dimension of P^r is $r-1$. Thus, by Carathéodory's theorem (see [9]), $\underline{\beta}$ is a convex combination of at most r of the vectors $\underline{a}^i$. ∎

Lemma 11 implies the following extension of Lemma 12.

Lemma 14. *Let H and G_n be as in Lemma 12 and let $d,\gamma>0$. There is an $\varepsilon=\varepsilon(H,d,\gamma)>0$ such that, if G is obtained from G_n by replacing the complete bipartite graphs between the color-classes of G_n by ε-regular pairs with density at least d, the resulting G contains an H-factor covering at least $(1-\gamma)n$ vertices of G.*

4.3. Bottle-graphs

Definition 2. A *bottle-graph* B of chromatic number r is a complete r-partite graph with color-class sizes $\sigma, w, w, \ldots, w$, where $\sigma \le w$. The number σ is the *neck* of the bottle and w is the *width* of B.

Note that $\chi_{cr}(B) = h/w = r-1+\sigma/w$.

Given an arbitrary r-chromatic graph H, we say that an r-chromatic bottle-graph B is a bottle-graph of H if the color-vector of B is $\underline{\beta} = (s,t,\ldots,t)$, where $s = \sigma(H)/v(H)$, and $t = (1-s)/(r-1)$. (Keep the smallest proportion and even out the rest.) Clearly, the vector $\underline{\beta}$ dominates any color-vector of H, and, since the critical chromatic number of a graph G only depends on $\chi(G)$ and $\sigma(G)/v(G)$, we have $\chi_{cr}(B) = \chi_{cr}(H)$.

Thus, *it is sufficient to prove Theorem 8 for bottle-graphs.* Indeed, if H is an arbitrary graph, then we first find a bottle-graph B of H with sufficiently many (but still a constant number of) vertices, then apply Theorem 8 for B to find an almost-perfect B-matching in any large G_n satisfying the degree condition (11), and then apply Lemma 12 to find an almost-perfect H-matching inside each copy of B, providing an almost-perfect H-matching of G_n.

Remark. Theorem 8 makes it clear that, in general, it is not the chromatic number $\chi(H)$ that determines the asymptotic behavior of the tiling problem for H but rather the related but more subtle quantity $\chi_{cr}(H)$. The Alon-Yuster proofs (similarly to that of Theorem 2) all start by equalizing the color-class sizes of H, that is, by embedding H into a complete $\chi(H)$-partite graph K with equal color-class sizes, and then tiling the large graph G with copies of K using the Regularity Lemma and the Hajnal-Szemerédi theorem. While this loss of information about H is too crude, our theorem says that only the smallest color-class has to be treated with more care, the leftover $\chi(H)-1$ classes can be equalized – hence the use of bottle-graphs.

Convention. In the rest of the paper, H is a fixed r-chromatic **bottle-graph**. We will assume $\sigma < w$ (that is, $\chi_{cr} < r$), since otherwise the theorem was proved already by Alon and Yuster (Theorem 5). We will also use two auxiliary graphs, K_r and H'. The latter one is obtained from H by removing one vertex from each color-class of size w. Note that they are both bottle-graphs, and they both dominate H (their color-vectors do).

The following lemma is a crucial step in the proof of Theorem 8.

Lemma 15. *Let H be a bottle-graph of order h, chromatic number r, critical chromatic number χ_{cr} and width w. For fixed $x \in (0,1)$ and $\varepsilon > 0$,* *let $n \geq n_0(H, x, \varepsilon)$, and let G_n be an n-graph with minimum degree $\delta \geq g(x)n$, and maximum number of vertices in G covered by an H-matching $M \leq (1-\varepsilon)xn$. Then G_n has a tiling with vertex-disjoint copies of H, H' and K_r that covers at least $M+\varepsilon' n$ vertices, where*

$$\varepsilon' = \frac{1}{h}\left(\frac{1}{r-1} - \frac{1}{\chi_{cr}}\right)\varepsilon x > 0.$$

Proof. Let us write $Z = (1-1/(r-1)+\varepsilon')|L|^2/2$. If $\mathcal{L}$ is the set of left-over vertices ($|\mathcal{L}| = L$), then, assuming L is large enough, the graph on these vertices has at most Z edges according to Theorem 2. Hence, at least $\varepsilon' L$ of the vertices $x \in \mathcal{L}$ have degrees at most $2Z/((1-\varepsilon')L)$ into $\mathcal{L}$. Let's pick such an x. Since $\deg(x) \geq \delta \geq g(x)n$, the number of copies of H in the matching into which x is connected by more than $h-w$ edges is at least

$$\left[\delta - \frac{2Z}{(1-\varepsilon')L} - \frac{h-w}{h}M\right]\frac{1}{w}$$
$$\geq \left[g(x)n - \left(1 - \frac{1}{r-1} + \varepsilon'\right)\frac{L}{1-\varepsilon'} - \frac{h-w}{h}M\right]\frac{1}{w} =: C$$

Now

$$\begin{aligned} wC = &\left[g(x) - \left(1 - \frac{1}{r-1} + \varepsilon'\right)\frac{1}{1-\varepsilon'}\right]n \\ &- \left[-\left(1 - \frac{1}{r-1} + \varepsilon'\right)\frac{1}{1-\varepsilon'} + \left(1 - \frac{1}{\chi_{cr}}\right)\right]M \\ \geq &\left[g(x) - \left(1 - \frac{1}{r-1} + \varepsilon'\right)\frac{1}{1-\varepsilon'}\right]n \\ &- \left[-\left(1 - \frac{1}{r-1} + \varepsilon'\right)\frac{1}{1-\varepsilon'} + \left(1 - \frac{1}{\chi_{cr}}\right)\right](1-\varepsilon)xn \\ \geq &\left(\frac{1}{r-1} - \frac{1}{\chi_{cr}}\right)\varepsilon xn - 2\varepsilon' n \geq w\varepsilon' n. \end{aligned}$$

Let us select one such copy $H(x)$. Since fewer than w edges are missing from x to $H(x)$, there is an edge from x to each color-class of $H(x)$ of size w. By connecting x to one vertex in each such class, we are actually splitting $H(x) \cup \{x\}$ into a copy of K_r and a copy of H'. Using a greedy algorithm to pair up $\varepsilon' n$ such vertices in $\mathcal{L}$ with appropriate copies of H in the matching, we obtained the required tiling of G_n with copies of H, H' and K_r. ∎

4.4. Proof of Theorem 8

Let $x \in (0,1)$ be given.

Remark. It is sufficient to prove the theorem under the stronger assumption

$$\delta(G_n) \geq (g(x) + \varepsilon)n. \tag{11'}$$

Indeed, if G_n only satisfies $\delta(G_n) \geq g(x)n$, then by introducing εn new vertices and connecting them to all vertices of G_n, we get a new graph G' that satisfies (11′) Thus, G' has an H-matching covering most vertices. Disregarding those copies of H that contain new vertices, we get the desired conclusion. We will thus assume that H is an r-chromatic bottle-graph, and G_n satisfies (11'). We need to cover at least xn vertices of G_n by copies of H.

We first apply the Regularity Lemma to G_n, and get k equal-size clusters of vertices with reduced graph R, and a left-over set V_0. Note that all degrees in R are at least $(g(x)+\varepsilon/2)k$.

Now, the rest of the proof will go as follows. Starting from $R = R_0$, we construct a constant-length sequence R_i corresponding to reduced graphs for G_n, where the vertices correspond to smaller and smaller clusters. One refinement step consists of two stages.

In Stage I, we start with a maximum-size H-matching in R_i, and apply Lemma 15 to get a larger tiling of R_i with copies of H, H' and K_r. Each of these copies corresponds to h clusters in G_n connected by some ε-regular pairs.

In Stage II, we subdivide the clusters corresponding to vertices of R_i to ℓ equal-size parts, where ℓ is large (we throw some vertices into V_0 if ℓ does not divide the cluster size.) In this new reduced graph R_{i+1}, the copies of H, H' and K_r correspond to large complete r-partite graphs in R_i each one dominating H. Hence, by Lemma 14, we can almost perfectly tile them with copies of H.

When by using these alternating steps we reach the desired proportion of covered vertices, we finish the proof by one final application of Lemma 14. ∎

Acknowledgement. I thank Endre Boros for the remark after Lemma 12.

Addendum. It came to our attention that Conjecture 4 has also been conjectured by Eldar Fischer, Robert Johansson and Sarmad Abbasi. Also, Noga Alon and Eldar Fischer proved the sufficiency of (11) for bipartite graphs H.

References

[1] Sarmad Abbasi: The solution of the El-Zahar problem, Ph. D. Thesis, Rutgers University, 1998.
[2] N. Alon and E. Fischer: 2-factors in dense graphs, *Discrete Math.,* to appear.
[3] N. Alon and R. Yuster: Almost H-factors in dense graphs, *Graphs and Combinatorics,* **8** (1992), 95–102.
[4] N. Alon and R. Yuster: H-factors in dense graphs, *J. Combinatorial Theory Ser. B,* **66** (1996), 269–282.
[5] B. Bollobás: *Extremal Graph Theory*, Academic Press, New York 1978.
[6] B. Bollobás and S. E. Eldridge: Problem in: *Proc. Colloque Intern. CNRS* (J.-C. Bermond et. al. eds.), 1978.
[7] B. Bollobás and S. E. Eldridge: Packings of graphs and applications to computational complexity, *J. Combinatorial Theory Ser. B,* **25** (1978), 105–124.
[8] K. Corrádi and A. Hajnal: On the maximal number of independent circuits in a graph, *Acta Math. Acad. Sci. Hungar.,* **14** (1963), 423–439.
[9] L. Danzer, B. Grünbaum and V. Klee: Helly's theorem and its relatives, in: *Convexity* (V.L. Klee, ed.), AMS, Providence, R.I. 1963, 101–180.
[10] G. A. Dirac: Some theorems on abstract graphs, *Proc. London Math. Soc.,* **2** (1952), 69–81.
[11] M. H. El-Zahar: On circuits in graphs, *Discrete Math.,* **50** (1984), 227–230.
[12] H. Enomoto, A. Kaneko and Zs. Tuza: P_3-factors and covering cycles in graphs of minimum degree $n/3$, *Combinatorics,* (A. Hajnal, L. Lovász and V. Sós eds.), Colloq. Math. Soc. J. Bolyai, vol. 52, Eger (Hungary) (1987), 213–220.
[13] P. Erdős: Some recent combinatorial problems, Preprint, November 1990.
[14] P. Erdős and M. Simonovits: A limit theorem in graph theory, *Studia Sci. Math. Hungar.,* **1** (1966), 51–57.
[15] P. Erdős and A. H. Stone: On the structure of linear graphs, *Bull. Amer. Math. Soc.,* **52** (1946), 1089–1091.
[16] A. Hajnal and E. Szemerédi: Proof of a conjecture of Erdős, *Combinatorial Theory and its Applications* vol. II (P. Erdős, A. Rényi and V. T. Sós eds.), Colloq. Math. Soc. J. Bolyai 4, North-Holland, Amsterdam (1970), 601–623.
[17] J. Komlós: The Blow-up Lemma (survey), *Combinatorics, Probability and Computing,* **8** (1999), 161–176.
[18] J. Komlós, G. N. Sárközy and E. Szemerédi: Blow-up Lemma, *Combinatorica,* **17** (1997), 109–123.
[19] J. Komlós and M. Simonovits: Szemerédi's Regularity Lemma and its applications in graph theory (survey), Bolyai Society Mathematical Studies 2, Combinatorics, Paul Erdős is Eighty (Volume 2) (D. Miklós, V. T. Sós and T. Szőnyi eds.), Keszthely (Hungary) (1993), Budapest (1996), 295–352.
[20] Henryk Minc: *Nonnegative Matrices*, John Wiley & Sons, New York 1988.
[21] N. Sauer and J. Spencer: Edge disjoint placements of graphs, *J. Combinatorial Theory Ser. B,* **25** (1978), 295–302.
[22] E. Szemerédi: Regular partitions of graphs, *Proc. Colloques Internationaux CNRS* № 260 – *Problèmes Combinatoires et Théorie des Graphes* (J.-C. Bermond, J.-C. Fournier, M. Las Vergnas and D. Sotteau eds.), Orsay (1976), 399–401.
[23] P. Turán: On an extremal problem in graph theory (in Hungarian), *Matematikai és Fizikai Lapok,* **48** (1941), 436–452.

János Komlós

Rutgers University
Pisvataway, NJ 08854, USA
komlos@math.rugers.edu