COMBINATORICA 2 (1) (1982) 1—7

LARGEST RANDOM COMPONENT OF A k-CUBE

M. AJTAI, J. KOMLÓS and E. SZEMERÉDI

Received 12 May 1981

Let C^k denote the graph with vertices $(\varepsilon_1, \dots, \varepsilon_k)$, $\varepsilon_i = 0,1$ and vertices adjacent if they differ in exactly one coordinate. We call C^k the k-cube.

Let $G = G_{k,p}$ denote the random subgraph of C^k defined by letting

$$\text{Prob}(\{i, j\}) \in G) = p$$

for all $i, j \in C^k$ and letting these probabilities be mutually independent.

We show that for $p = \lambda/k$, $\lambda > 1$, $G_{k,p}$ almost surely contains a connected component of size $c2^k$, $c = c(\lambda)$. It is also true that the second largest component is of size $o(2^k)$.

Introduction

In [2] Erdős and Spencer proved that $G_{k,p}$ is connected with probability ~ 0 for $p < 1/2$, $\sim e^{-1}$ for $p = 1/2$, ~ 1 for $p > 1/2$. They mentioned that in $G_{k,p}$, $p = \lambda/k$, $\lambda < 1$, the size of the largest component is $o(2^k)$ (almost surely), and suggested that there might be a jump at $\lambda = 1$, i.e. for $\lambda > 1$, $G_{k,\lambda/k}$ contains a component of size $c2^k$. We are going to prove this latter conjecture.

Notation

C^k denotes the k-cube, sometimes the graph, mostly its vertex-set; we write edges for those of $G_{k,p}$ and *Edges* for those of the skeleton C^k (i.e. Edges are possible edges);
$G_k(\lambda)$ denotes the random graph $G_{k,\lambda/k}$;
a vertex $v \in C^k$ is *even* or *odd* according to the number of *ones* in v;
if two vertices x, y differ in exactly i coordinates, we say that their *Hamming distance* $d(x, y) = i$;
x and y are *neighbours* if $d(x, y) = 1$, furthermore x is the *up-neighbour* (y is the *down-neighbour*) if x contains more *ones;*
for two sets $A, B \subset C^k$ the Hamming distance is

$$d(A, B) = \inf_{\substack{a \in A \\ b \in B}} d(a, b),$$

A and B are *neighbours* if $d(A, B) \leqq 1$;

AMS subject classification (1980): 05 C 40; 60 C 05

the set of vertices of a connected subgraph is called a *cell* (i.e. components are maximal cells);
for a set $A \subset C^k$ we define the neighbourhood

$$\Gamma(A) = \{x \in C^k;\ \exists y \in A:\ d(x, y) \leqq 1\},$$

and the exact neighbourhood $\gamma(A) = \Gamma(A) - A$;
e_0 is the vertex with all zero coordinates, $e_1, \dots, e_k$ are the unit vectors;
$n = 2^k$ is the cardinality of C^k;
$P(\cdot)$ and E stand for probability and expectation;
$o(1)$ is understood as $k \to \infty$;
all constants appearing in the paper depend on λ of the Theorem, furthermore $c_0, c_1, \dots$ are assumed to be positive and small enough in terms of all previously *defined* c, K_1, K_2 are large constants.

Theorem. *In $G_k(\lambda)$, $\lambda > 1$, there is a connected component of size $> c_0 n$ with probability $1 - o(1)$.*

About the value c_0 see Remark 2 at the end of the paper.

The theorem will be established using a blowing-up (or shrinking) argument, this is expressed in the following series of lemmas.

Lemma 1. *The probability that in $G_k(\lambda)$, $\lambda > 1$, the vertex e_0 belongs to a component of size $> c_1 k$, is at least c_2.*

Lemma 2. *In $G_k(\lambda)$, $\lambda > 1$, with probability $1 - o(1)$, all vertices v except for at most n^{1-c_3}, have the following property:*

Property 2. There are $c_4 k$ disjoint cells neighbouring v, each of size $c_5 k$.

Lemma $\bar{2}$. *In $G_k(\lambda)$, $\lambda > 1$, with probability $1 - o(1)$, all vertices v except for at most n^{1-c_3}, have the following property:*

Property $\bar{2}$. The vertices of any neighbouring cell of size $c_5 k$ have Property 2.

Lemma 3. *In $G_k(\lambda)$, $\lambda > 1$, with probability $1 - o(1)$, all vertices v except for at most n^{1-c_6}, have the following property:*

Property 3. There are $c_7 k$ neighbours of v belonging to components of size $> c_8 k^2$.

Proof of Lemma 1. When we refer to the *Laws of Large Numbers,* we always apply the following very weak statement: If $A_1, \dots, A_m$ are independent events, $P(A_i) \geqq p$, $1 \leqq i \leqq m$, $mp = E$, and X denotes the number of events occurred, then

$$P(X < E/10) < e^{-E/2}$$

if only $E > E_0$ (an absolute constant).

Indeed,

$$P(X < E/10) \leqq \sum_{j < E/10} \binom{m}{j} p^j (1-p)^{m-j} < \sum_{j < E/10} (mpe^p)^j e^{-E}/j!$$

$$< 2(mpe)^{E/10} e^{-E}/(E/10)! < e^{-E/2}$$

We will also use Markov's inequality

$$P(X > t) \leqq (EX)/t$$

for any non-negative random variable X and positive number t.

The proof of Lemma 1 is based on the following result of the theory of branching processes (for notations and the quoted result see e.g. [4]):

Given a Galton–Watson process with distribution $p_i = P$(number of offsets$=i$), $i=0, 1, \dots$ set $f(x) = \sum_i p_i x^i$ (the generating function). If the expectation $\sum_i i p_i > 1$, then with probability $q=1-p$ the process does not stop, where p is the root of the equation

$$f(p) = p, \quad 0 \leqq p < 1$$

(since f is convex and $f(1)=1$, p is unique unless $p_1=1$ in which case we take $p=0$).

Now it is easy to see ([1], proof of Claim 3) that the quoted result implies:

If the distribution of the number of offsets is binomial $p_i = \binom{m}{i} \alpha^i (1-\alpha)^{m-i}$, and the expectation $m\alpha \geqq 1+\varepsilon$, then $q \geqq \delta$, where $\delta > 0$ depends only on $\varepsilon > 0$ and *not* on m and α.

Now we take $c_1 = (\lambda-1)/(3\lambda)$ and apply the above result with $m = k(1-2c_1)$, $\alpha = \lambda/k$.

Pick m out of the k neighbours of e_0 and randomize the edges going to them from e_0. The number X_1 of neighbours that get connected to e_0 follows a binomial distribution $B(m, \alpha)$. If $X_1 > c_1 k$, we are home. Otherwise denote these neighbours (if any) by $d_1, \dots, d_{X_1}$.

Pick m out of the $k-1$ up-neighbours of d_1. The number $X_{2,1}$ of those connected to d_1 is again a $B(m, \alpha)$. Then pick m up-neighbours of d_2, different from those $X_{2,1}$ connected to d_1 and randomize to get $X_{2,2}$ of them connected to d_2, etc. Then we go one level further, etc.

We can always choose m up-neighbours of the current vertex, which are different from those obtained earlier as long as the total number of vertices on the tree up to this point is not more than $c_1 k$. If it is more than that, we stop the process.

Now the probability that our process gets stuck (dies out) before we get $c_1 k$ points and stop it, is at least some $c_2 > 0$, since

$$m\lambda/k = (1-2c_1)\lambda > 1. \quad \blacksquare$$

Note that for reason of symmetry, the vertex e_0 can be replaced by any other vertex $v \in C^k$.

Remark 1. In the proof of this "Shrinking lemma" we have not used any special structural property of C^k, only the fact that

Property A. All valencies are at least k.

The further lemmas, however, must use something more, for if the skeleton is not C^k but the Turán-graph ($n/(k+1)$ disjoint cliques of size $k+1$) then Lemma 1 holds but Lemmas 2, $\bar{2}$, 3 and the theorem itself do not. In the proof of Lemma 2 we will use the property that C^k can be split to k disjoint "smaller copies", but it can be easily seen that the following property would suffice:

Property B. For every $\varepsilon>0$ there is a $\delta>0$ such that for $i=\delta k$ the following condition holds:
any i vertices have altogether at least $i(1-\varepsilon)k$ neighbours.

For finishing the proof we will need the additional property that the skeleton graph does not contain large isolated subsets of vertices:

Property C. For every $d_1>0$ there is a $d_2>0$ such that every vertex-set of size between $d_1 n$ and $(1-d_1)n$ has a boundary of size at least $d_2 n/k^{1/2}$.

(Here 1/2 might be changed to larger exponent, but the boundary would still have to be fairly large. This is thus the strong condition, and it fails to hold for the lattice of small dimensions.) Properties A, B, C can substitute the special structure of the k-cube.

Proof of Lemma 2. Let us fix a vertex v. For reason of symmetry, we can assume it is e_0. Consider its neighbours e_i, $1\leqq i\leqq c_9 k$. For given i we fix the first $i-1$ coordinates to 0, the i-th to 1, and vary only the remaining $k-i$. We actually have a cube C^{k-i}. Note that $k-i\geqq(1-c_9)k$, and these cubes are disjoint for different i.

Thus Lemma 1 implies that each e_i belongs to a component of size $>c_5 k$ in its own cube with probability $>c_{10}$ each, if only $(1-c_9)\lambda>1$. By the Laws of Large Numbers:

P (less than $c_4 k$ of these events occur) $<e^{-c_{11}k}=n^{-c_{12}}$. Thus every (fixed) vertex has Property 2 with probability $>1-n^{-c_{12}}$. Hence the expected value of the number N of vertices not having Property 2 is less than $n^{1-c_{12}}$, whence by Markov's inequality $N<n^{1-c_3}$ with probability $>1-n^{c_3-c_{12}}=1-o(1)$.

(For later purposes, we assume that $\binom{k}{c_5 k}<n^{c_3/2}$. If this would not hold, throw out points from the above neighbouring cells to make c_5 smaller.) ∎

Proof of Lemma $\bar{2}$. If N denotes the set of vertices not having Property 2, then the vertices v not having Property $\bar{2}$ are in N or have a neighbouring cell of size $c_5 k$ with a vertex in N.

Thus these vertices v are in a distance less than $c_5 k$ from N. But the number of such vertices is not more than

$$|N|\binom{k}{c_5 k}<n^{1-c_3}n^{c_3/2}=n^{1-c_3/2}=n^{1-\bar{c}_3}. \quad ∎$$

Proof of Lemma 3. Choose a λ', $1<\lambda'<\lambda$. We construct $G_k(\lambda)$ by first randomizing with $p=\lambda'/k$ (i.e. producing a $G_k(\lambda')$), and then make an additional randomization with $p=d/k$.

We assume (Lemma $\bar{2}$) that in the starting graph $G_k(\lambda')$ all vertices, except for at most $n^{1-\bar{c}_3'}$, have Property $\bar{2}$ (with parameters c_4', c_5'). We fix to every one of those vertices v a system S_v of $c_4'k$ disjoint cells of size $c_5'k$, such that all their vertices have Property 2. Denote the set of exceptional vertices by N ($|N|<n^{1-\bar{c}_3'}$).

Now we perform the second randomization. We show that with probability $1-o(1)$ all the above fixed cells, except for at most $n^{1-c_{13}}$, melt into components of size $>c_8 k^2$. This will obviously imply Lemma 3.

Let $S \subset C^k - N$ be an arbitrary cell of $G_k(\lambda')$ of size $c_{14}k$, and let $T = \{t_1, t_2, \ldots\}$ be the set of even resp. odd vertices of S, whichever are more. I.e. $|T| \geqq c_{14}k/2 = l$.

t_1 (as any other vertex of S) has Property 2, and in the second randomization it gets connected with a random number of its $c_4'k$ disjoint neighbouring cells of $G_k(\lambda')$; the probability that with at least one of them, is at least

$$1-(1-d/k)^{c_4'k} > c_{15}.$$

Denote by A_1 the (possibly empty) union if those connected with t_1. Now the same holds for t_2 (and the Edges starting from t_2 are different from those starting from t_1 since t_1 and t_2 are of the same parity). Out of the $c_4'k$ disjoint cells of $G_k(\lambda')$ neighbouring t_2 at least $c_4'k/2$ intersect in less than half with A_1 if only $|A_1| < c_{18}k^2$. Randomize the edges $(p=d/k)$ from t_2 to these $c_4'k/2$ cells. The probability that t_2 gets connected with at least one of them, is at least

$$1-(1-d/k)^{c_4'k/2} > c_{15}/2 = c_{15}'.$$

Denote by A_2 the union of those connected with t_2, and set $B_2 = A_1 \cup A_2$. Now we pass to t_3, etc. Using the notation $B_j = \bigcup_{i=1}^{j} A_i$, we have, as before, that out of the $c_4'k$ disjoint cells neighbouring t_j at least $c_4'k/2$ intersect in less than half with B_{j-1} if only $|B_{j-1}| < c_{18}k^2$. Randomizing the edges $(p=d/k)$ from t_j to these cells, the probability that t_j gets connected with at least one of them, is at least c_{15}'. Thus the expected size of B_l is at least

$$E|B_l| > c_{15'}\, l(c_5'^{k/2}) = c_{16}k^2$$

(if only $c_{16} < c_{18}$, otherwise c_{16} is replaced by min (c_{16}, c_{18})).
Hence, by the Laws of Large Numbers

$$P(|B_l| < c_8k^2) < e^{-c_{17}k} = n^{-c_{18}}$$

(if only c_8 was chosen small enough), since the $c_{19}k^2$ randomizations involved were independent.

We have shown that if $S \subset C^k - N$ is any cell of $G_k(\lambda')$ of size $c_{14}k$, then S will melt into a component of size $> c_8k^2$ with probability $> 1 - n^{-c_{19}}$. Now apply this to the cells of the system S_v fixed at the beginning of this proof. Since their number is less than nk, the expected number of those not melting into components of size $> c_8k^2$, is less than $nkn^{-c_{19}} < n^{-c_{20}}$. ∎

Proof of the Theorem. Choose a λ', $1 < \lambda' < \lambda$, and consider $G_k(\lambda')$ first, and add new edges afterwards with $p = d/k$. We will write $P_2(\,.)$ for the probability of events in this latter randomization.

We assume (Lemma 3) that in $G_k(\lambda')$ all vertices except those in a set N of size $< n^{1-c_6'}$, have Property 3 (with parameters c_7', c_8'). The components of $G_k(\lambda')$ of size $> c_8'k^2$ will be called atoms (even after the second randomization atom will mean that in $G_k(\lambda')$).

$|N| \ll n$ implies, in particular, that at least $c_7'n/2 = c_{21}n$ points belong to atoms.

Set $c_0 = c_{21}/3$.

Now we make the second randomization. We claim that, with probability $1-o(1)$, there will be a component of size $>c_0 n$, such that even the union of atoms within this component will be larger than $c_0 n$.

Indeed, assume that the union of atoms within every component is of size $\leqq c_0 n$. Then there would be a union A of atoms that would be separated (even in $G_k(\lambda)$) from the union B of the remaining atoms, and

$$c_0 n \leqq |A| < 2c_0 n, \quad \text{consequently} \quad |B| \geqq c_0 n.$$

We show that this event has a negligible probability. Since there are at most $n/(c_8' k^2)$ atoms, the number of possibilities for A is at most $2^{n/(c_8' k^2)}$. Thus, it is enough to show that the probability that A and B are separated, for a particular pair A, B, is less than

$$e^{-K_1 n/k^2} \quad (K_1 \text{ large}).$$

Consider the neighbourhoods $\Gamma(A)$, $\Gamma(B)$, and define

$$D = \Gamma(A) \cap \Gamma(B), \quad F = C^k - \Gamma(A) - \Gamma(B).$$

Note that $F \subset N$, and thus $|F| < n^{1-c_6'}$.

There are two possibilities:

(I) $|D| > K_2 n/k$

(II) $|D| \leqq K_2 n/k$

In case (I) set $D' = D - N$. We have $|D'| > (K_2/2)n/k$ points x such that both A and B are neighbouring x, and at least one of them contains $>(c_7'/2)k$ neighbours of x. Choose $(K_2/4)n/k$ points x of the same parity. Thus,

$$P_2(x \text{ connects } A \text{ and } B) > c_{23}/k$$

and these events are independent for different x. Hence

$$P_2\,(\text{no } x \in D' \text{ connects } A \text{ and } B) < (1 - c_{23}/k)^{(K_2/2)n/k} < e^{-K_1 n/k^2}$$

if only K_2 was large enough.

In case (II) note that $|\Gamma(A)| \geqq |A| \geqq c_0 n$ and $|\Gamma(A)| \leqq n - |B| \leqq (1-c_0)n$. It is well-known [3] that such a set $\Gamma(A)$ has a boundary $> c_{24} n/\sqrt{k}$, and this obviously implies that there are $c_{25} n/\sqrt{k}$ disjoint Edges going out from $\Gamma(A)$ (to its complement) If we drop those which have an endpoint in D or N, we still have at least $(c_{25}/2)n/\sqrt{k} = c_{26} n/\sqrt{k}$ disjoint Edges from $\gamma(A)$ to $\gamma(B)$.

For such an Edge $e = (x, y)$, $x \in \gamma(A)$, $y \in \gamma(B)$, there are $c_7' k$ neighbours of x in A and $c_7' k$ neighbours of y in B, thus

$$P_2\big(A \text{ and } B \text{ get connected through } (x, y)\big) > c_{27}/k.$$

Now these events are independent for different e, hence

$$P_2\,(A \text{ and } B \text{ are not connected}) < (1 - c_{27}/k)^{c_{26} n/\sqrt{k}} < e^{-K_1 n/k^2}.$$

Remark 2. A slight modification of the proof given above shows that the obtained giant component is actually everywhere dense in the following strong sense: all vertices v, except for at most $n^{1-c_{28}}$, have the following property:

There are $c_{29} k$ neighbours of v belonging to the giant component.

This however does not automatically imply (though makes it very likely) that the second largest component is of size $o(n)$ (as conjectured by Erdős). This is true though, for a close inspection of the proof shows that c_1, and thus c_0 too, can be chosen arbitrarily close to the constant $q=q(\lambda)$, where q is the probability that a branching process with Poisson distribution of parameter λ does not die out. I.e. $q=1-p$, where p is the solution of the equation $f(p)=p$, $0<p<1$, and $f(x)=$ $=e^{\lambda(x-1)}$ is the generator function of the Poisson distribution. In other words, $p=x/\lambda$, where x (the conjugate of λ) is the solution of

$$xe^{-x} = \lambda e^{-\lambda}, \quad 0 < x < 1.$$

Thus the giant component is as large as possible (for $(p+o(1))n$ vertices belong to bounded components), and there is no room for another component of size cn. (Actually, the proof in [5] can be adapted to C^k and that shows that the second largest component is only of size $O(k)$.)

References

[1] M. Ajtai, J. Komlós and E. Szemerédi, The longest path in a random graph, *Combinatorica* **1** (1981) 1—12.

[2] P. Erdős and J. Spencer, Evolution of the n-cube, *Computers and Math. with Applications* **5** (1979) 33—40.

[3] L. H. Harper, Optimal numberings and isoperimetric problems on graphs, *Journal of Comb. Th.* **1** (1966) 358—394.

[4] T. E. Harris, *The theory of branching processes,* Springer (1963).

[5] J. Komlós, M. Sulyok and E. Szemerédi, Underdogs in a random graph, *submitted to Studia Sci. Math. Hung.*

M. Ajtai, J. Komlós, E. Szemerédi

Mathematical Institute of the
Hungarian Academy of Sciences
Budapest, Hungary, H—1053