ON CONNECTEDNESS OF A RANDOM GRAPH WITH A SMALL NUMBER OF EDGES

by

Z. FÜREDI

Abstract

Consider the $n\times n$ lattice graph $G(n)$. Let $G(n,p)$ denote the random subgraph of $G(n)$ defined by choosing the edges of $G(n)$ with probability p mutually independendently.

We prove: $p<1/3$ is fixed then $G(n,p)$ is highly nonconnected, i. e. the largest component of $G(n,p)<C(p)\log n$, if $p>2/3$ then $G(n,p)$ is nearly connected, i.e. there is a giant component. But $G(n,p)$ will be connected if p is very close to 1, more precisely if $p=1-c/\sqrt[4]{n}$ then

$$\lim_{\infty\to n} \text{Prob}\,(G(n,p) \text{ is connected}) = e^{-c^4}.$$

It was P. ERDŐS and A. RÉNYI who first posed the problem of investigating the properties of random graphs (see [2], [3], [4], [5]). One of their most known result is as follows [2].

Let us consider the $\binom{n}{2}$ edges of the complete graph on n points as (completely) independent random variables. Let us denote by e the random variable which corresponds to the edge e. It takes the value 1 or 0 according to whether this edge belongs to our random graph or not. Further we suppose that $p=\text{Prob}\,(e=1)=$ $=\frac{\log n}{n}+\frac{c}{n}$ and $\text{Prob}\,(e=0)=1-p$. This random graph (or random vector variable) will be denoted by $G_{n,p}$. Now as n tends to infinity and c is fixed,

$$\lim \text{Prob}\,(G_{n,p} \text{ is connected}) = e^{-e^{-c}} =$$

$$= \lim \text{Prob}\,(G_{n,p} \text{ has no isolated point}).$$

This theorem was generalized in many ways, e.g. [8] if

$$p=\frac{\log n}{n}+\frac{w(n)}{n}$$

then

$$\lim_{n\to\infty} \text{Prob}\,(G_{n,p} \text{ contains a Hamiltonian circuit}) = 1 \text{ or } 0$$

according to $\lim_{n\to\infty} w(n)=\infty$ resp. $-\infty$.

(The threshold function is not yet known.)

1980 *Mathematics Subject Classification*. Primary 05C40; Secondary 60C05.
Key words and phrases. Random graph, threshold function.

Another direction of generalization which is considered in this paper, is that one regards not a complete graph but some other one as the underlying graph. In this case only the edges of this underlying graph are chosen randomly. There are several theorems similar to the above cited one in this topic, e.g. for the complete bipartite graph or the n-cube [1], [6], [7]. However; the investigated underlying graphs usually have $O\left(\binom{|V(G)|}{2}\right)$ edges [9]. It is surprising that the case when G has only few edges has not been investigated. Questions of this type are of similar character and have important applications in physics.

In this paper we consider the following special case. Let $G(n)$ be the graph formed by an $n\times n$ square lattice. $|V(G)|=(n+1)^2$ $|E(G)|=2n(n+1)$ .Choose the edges of $G(n)$ independently with probability p, and denote this random graph by $G(n,p)$.

THEOREM 1. *If* $p=1-\frac{c}{\sqrt{n}}$ *where c is fixed, then*

$$\lim_{n\to\infty} \text{Prob}\,(G(n,p) \text{ is connected}) = e^{-c^4} =$$

$$= \lim_{n\to\infty} \text{Prob}\,(G(n,p) \text{ has no isolated point}).$$

THEOREM 2. *Let p be fixed not depending on n. Then*
a) *If* $p>2/3$ *then* $G(n,p)$ *contains a giant component and the*

$$\text{2-nd largest component of } G(n,p) < \left(\frac{\log n}{\log(1/3(1-p))}+w(n)\right)^2.$$

b) *If* $p<1/3$ *then*

$$\text{the largest component of } G(n,p) < C(p)\log n.$$

(The inequalities are meant in the sense that they hold true with probability tending to 1 as $n\to\infty$.)

The Theorems mean that $G(n,p)$ is highly nonconnected if p is small, but for large p $G(n,p)$ is nearly connected and it will be really connected if p is very close to 1. It seems to be difficult to determine at what value of p the threshold value lies, but it is very likely 1/2 (cf. [6]). The method of the proof of Theorem 2 yields that the constant 2/3 can be improved to 0.658.

PROOF of Theorem 1.

A cutset of $G(n)$ is said to be *connected* if its edges can be linearly ordered in such a way that consecutive edges are neighbouring. The edges are *neighbouring* if they belong to the same small square in $G(n)$ (see Fig. 3). It is easy to see that if $G(n,p)$ is not connected then there exists a cutset of $\overline{G(n,p)}$ so it has a connected cutset by its planarity.

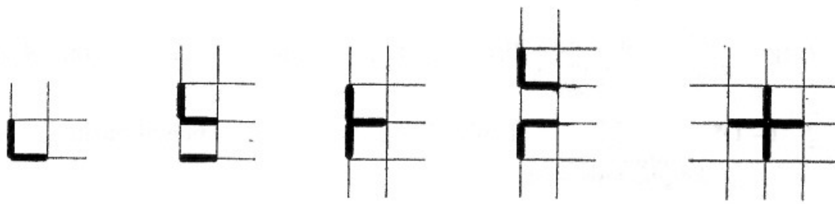
We shall see from the proof of Theorem 2 that

Prob (there exists a connected cut with k or more edges) $< 2n(n+1)3^k(1-p)^k/$ $/(1-3(1-p))$.

Put $k=5$, then

$$\text{Prob (there exists a connected cutset with } \geqq 5 \text{ edges)} <$$
$$< 2n(n+1)\big(3c/\sqrt{n}\big)^5/\big(1-3c/\sqrt{n}\big) = O\big(c^5/\sqrt{n}\big).$$

On the other hand any connected cutset with 2, 3 or 4 edges must be one of the following five types (Figure 1).



(i) (ii) (iii) (iv) (v)

Fig. 1

(i) *A connected cutset with 2 edges is in the corner.* There are 4 such cuts.

$$\text{Prob}\big((\text{i})\big) < 4(1-p)^2 = 4c^2/n.$$

(ii) *A connected cutset with 3 edges is in the corner*. There are 8 such cutsets.

$$\text{Prob}\big((\text{ii})\big) < 8(1-p)^3 = 8c^3/n\sqrt{n}.$$

(iii) *Isolated point on the boundary of* $G(n)$. There are $4(n-1)$ such cutsets.

$$\text{Prob}\big((\text{iii})\big) < 4(n-1)(1-p)^3 < 8c^3\sqrt{n}.$$

(iv) *A connected cutset with* 4 *edges on the boundary of* $G(n)$. There are $4(n-2)$ such cutsets.

$$\text{Prob}\big((\text{iv})\big) < 4(n-2)(1-p)^4 < 4c^4/n.$$

(v) *Isolated point inside* $G(n)$. There are $(n-2)^2$ such cutsets.

From these we get.

Prob $\big(G(n, \text{p})$ has no isolated point$\big)=$
$=O(1)+$Prob (There are no isolated points and no connected cutsets with $\geqq 5$ edges)$=$
$=O(1)+$Prob (There are no cutsets with $\geqq 2$ edges)$+O(1)=$
$=O(1)+$Prob $\big(G(n, \text{p})$ is connected$\big)$.

We are finished with the proof of the equation

$$\lim_{n\to\infty} \text{Prob}\big(G(n,p) \text{ is connected}\big) = \lim_{n\to\infty} \text{Prob}\big(G(n,p) \text{ has no isolated point}\big).$$

Further

$$\text{Prob}\big(G(n,p) \text{ has no isolated point}\big) =$$
$$= \text{Prob (There is no isolated point inside the lattice } G(n,p)) - o(1) =$$
$$= \text{Prob}\big(\neg(\text{v})\big) - o(1).$$

By an application of the sieve method we determine Prob $(\neg(\mathrm{v}))$. Write A_i for the event that the i-th inner point is isolated in $G(n,p)$ $(1 \leqq i \leqq (n-1)^2)$. A_u and A_v are *neighbouring* if the u-th and v-th points are joined by an edge in $G(n)$.

$$\operatorname{Prob}(\neg(\mathrm{v})) = \operatorname{Prob}(\neg(\mathrm{v}) \text{ and there are no neighbouring } A_u A_v \text{ at all}).$$

(Clearly, the non-existence of isolated points implies the non-existence of pairs of neighbouring isolated points.)

$$\operatorname{Prob}(\neg(\mathrm{v})) = \sum_{k=0}^{(n-1)^2} (-1)^k \sum_{i_1, i_2, \dots, i_k} \operatorname{Prob}(A_{i_1} A_{i_2} \dots A_{i_k} \text{ and } \nexists^1 \text{ neighbouring } A_u A_v) =$$

$$= \sum_{k=0}^{(n-1)^2} (-1)^k \sum_{\substack{i_1, i_2, \dots, i_k \text{ and} \\ \nexists A_{i_\alpha} A_{i_\beta} \text{ neighbouring}}} \operatorname{Prob}(A_{i_1} A_{i_2} \dots A_{i_k} \text{ and } \nexists \text{ neighbouring } A_u A_v).$$

In the last sum it suffices to consider only those k with $0 \leqq k < \log_2 n \cdot \max(6c^4, 3)$, because if k is more than this upper bound then

$$0 < \sum_{\substack{i_1, i_2, \dots, i_k \text{ and} \\ \nexists \text{ neighbouring } A_{i_\alpha} A_{i_\beta}}} \operatorname{Prob}(A_{i_1} \dots A_{i_k} \text{ and } \nexists \text{ neighbouring } A_u A_v) \leqq$$

$$\leqq \sum_{\substack{i_1, i_2, \dots, i_k \text{ and} \\ \nexists \text{ neighbouring } A_{i_\alpha} A_{i_\beta}}} \operatorname{Prob}(A_{i_1} \dots A_{i_k}) = \sum_{\substack{i_1, i_2, \dots, i_k \text{ and} \\ \nexists \text{ neighbouring } A_{i_\alpha} A_{i_\beta}}} (1-p)^{4k} <$$

$$< \binom{(n-1)^2}{k} (1-p)^{4k} < \frac{n^{2k}}{k!} \frac{c^{4k}}{n^{2k}} < \left(\frac{c^4 e}{k}\right)^k < \frac{1}{n^3}.$$

So it suffices to sum over $k < (6c^4+3)\log_2 n$.

$$\operatorname{Prob}(\neg(\mathrm{v})) =$$

$$= O\left(\frac{1}{n}\right) + \sum_{k=0}^{c \log_2 n} (-1)^k \sum_{\substack{i_1, i_2, \dots, i_k \\ \nexists A_{i_\alpha} A_{i_\beta} \text{ neighbouring}}} \operatorname{Prob}(A_{i_1} \dots A_{i_k} \text{ and } \nexists \text{ neighbouring } A_u A_v) =$$

$$= o(1) + \sum_{k=1}^{c \log_2 n} (-1)^k \sum_{\substack{i_1, i_2, \dots, i_k \text{ and} \\ \nexists A_{i_\alpha} A_{i_\beta} \text{ neighbouring}}} \operatorname{Prob}(A_{i_1} \dots A_{i_k}) -$$

$$- \sum_{k=0}^{c \log_2 n} (-1)^k \sum_{\substack{i_1, i_2, \dots, i_k \text{ and} \\ \nexists A_{i_\alpha} A_{i_\beta} \text{ neighbouring}}} \operatorname{Prob}(A_{i_1} \dots A_{i_k} \text{ and } \exists \text{ neighbouring } A_u A_v) =$$

$$= o(1) + S_1 - S_2,$$

where S_1 and S_2 are the sums.

Now we show that S_1 is equal to $e^{-c^4} + o(1)$. To do so we have to count how many times one can choose k out of the A_i' s so that among the chosen k ones there

[1] The symbol $\nexists$ stands for $\neg\exists$.

are no neighbours. The number of such choices is at most $\binom{(n-1)^2}{k}$ and at least $\binom{(n-1)^2}{k}-4(k-1)\binom{(n-1)^2}{k}$. From this we get that

$$S_1 = 1-\frac{c^4}{1!}+\frac{c^8}{2!}-\ldots+O\left(\frac{\log n}{n^2}\right) = e^{-c^4}+o(1).$$

We divide every term of S_2 into two parts.

$$\sum_{\substack{i_1, i_2, \ldots, i_k \text{ and} \\ \nexists A_{i_\alpha} A_{i_\beta} \text{ neighbouring}}} \text{Prob}\,(A_{i_1} \ldots A_{i_k} \text{ and } \exists \text{ neighbouring } A_u A_v) =$$

$$= \sum_{\substack{A_{i_1}, \ldots, A_{i_k} \\ \text{contain no neighbouring pair,} \\ \text{but they imply a neighbouring pair} \\ A_u A_{i_\alpha} \text{ (see Fig. 2)}}} \text{Prob}\,(A_{i_1} \ldots A_{i_k} \text{ and } \exists \text{ neighbouring } A_u A_v) +$$

$$+ \sum_{\substack{A_{i_1}, \ldots, A_{i_k} \\ \text{contain no eighbouring pair, and} \\ \text{they do not imply a neighbouring pair.}}} \text{Prob}\,(A_{i_1} \ldots A_{i_k} \text{ and } \exists \text{ neighbouring } A_u A_v).$$

If the position of $A_{i_1}, \ldots, A_{i_k}$ implies the existence of a neighbouring pair (see Fig. 2) (but A_{i_α} and A_{i_β} are not neighbouring) then

$$\text{Prob}\,(A_{i_1} \ldots A_{i_k} \text{ and } \exists \text{ neighbouring } A_u A_v) =$$

$$= \text{Prob}\,(A_{i_1} \ldots A_{i_k}) = (1-p)^{4k} = c^{4k}/n^{2k}.$$

But the number of $A_{i_1}, \ldots, A_{i_k}$ in such a position is at most $(n-1)^2\binom{(n-1)^2}{k-4}$.

If this is not the case then

$$\text{Prob}\,(A_{i_1} \ldots A_{i_k} \text{ and } \exists \text{ neighbouring } A_u A_v) =$$

$$= \text{Prob}\,(A_{i_1} \ldots A_{i_k} \text{ and } \exists \text{ neighbouring } A_u A_v \text{ and}$$

$$[\text{one of } A_{i_\alpha} \text{ and } A_u \text{ or } A_v \text{ are neighbouring}]) +$$

$$+ \text{Prob}\,(A_{i_1} \ldots A_{i_k} \text{ and } \exists \text{ neighbouring } A_u A_v \text{ and } \neg$$

$$[\text{one of } A_{i_\alpha} \text{ and } A_u \text{ or } A_v \text{ are neighbouring}]) <$$

$$< (1-p)^{4k}\left(4k(1-p)+2n^2(1-p)^k\right).$$

Using these facts we get that S_2 is $o(1)$, consequently

$$\text{Prob}\,(\neg(\text{v})) = e^{-c^4}+o(1). \quad \text{Q. E. D.}$$

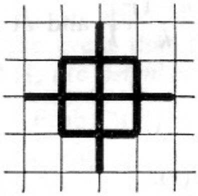


Fig. 2

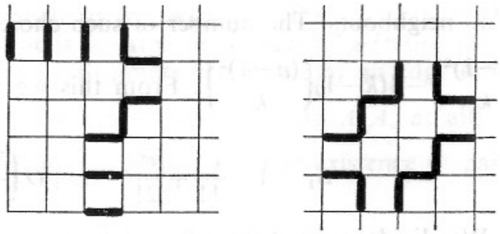


Fig. 3

PROOF of Theorem 2.

a) To begin with we show that the number of connected cuts of $G(n)$ with k edges is at most $n(n+1)3^{k-1}\cdot 4$ (see Fig. 3). This follows from the fact that on every connected cutset we can choose a starting edge and find (at most) two orderings of the edges of this cutset. Thus we can encode the form of the cut by 4 signs. We get a sequence of k signs with no consecutive identical signs. The number of such sequences is at most $4\cdot 3^{k-1}$. Hence

$$\operatorname{Prob}\left(G(n,p) \text{ has a connected cut with } \geqq k \text{ edges}\right) \leqq$$

$$\leqq \sum_{i=k}^{\infty} \operatorname{Prob}\left(G(n,p) \text{ has a connected cut with } i \text{ edges}\right) <$$

$$< \sum_{i=k}^{\infty} n(n+1)4\cdot 3^{k-1}(1-p)^i = n(n+1)4\cdot 3^{k-1}(1-p)^k \frac{1}{1-3(1-p)}.$$

Thus if $k > 2\log n/\log(1/3(1-p)) + w(n)$ then

$$\operatorname{Prob}\left(G(n,p) \text{ has a connected cut with } \geqq k \text{ edges}\right) = o(1).$$

In view of the fact that a cutset with k edges surrounds a set of at most $\frac{1}{4}k^2$ points:

$$\operatorname{Prob}\left(G(n,p) \text{ has a component with } \geqq \frac{k^2}{4} \text{ points}\right) = o(1).$$

b) Starting from an arbitrary point A let us go on the edges of $G(n)$, in every point P we can choose from at most three edges. Those new points connected with P are called the successors of P. As we can list the edges of $G(n)$ arbitrarily, so there is an appropriate branching process for the building of the component of $G(n,p)$ containing A (see [0]). Since

$$\mathrm{E}\,(\text{number of successors of } P) \leqq 3p < 1,$$

this process is subcritical (Galton—Watson), i.e. it extincts exponentially, more precisely, there are positive constants c, ε such that

Prob (the cardinality of the component of $G(n,p)$ containing A is more than $k) < ce^{-\varepsilon}K$.

Thus

$$\text{Prob}\,(G(n,p) \text{ contains a component greater then } C(p)\log n) \leqq$$

$$\leqq \sum_A \text{Prob}\,(G(n,p) \text{ contains a component containing } A \text{ and greater than } C(p)\log n) <$$

$$< n^2 c e^{-c(p)\log n} = o(1).$$

REMARK. The methods presented here can be generalized to connected graphs for which the maximal degree is very small compared with the number of vertices, e.g. the lattice points of d dimensional cubes of size n ($n\to\infty$, d is fixed).

REFERENCES

[0] ATHREYA, K. B. and NEY, P. E., *Branching processes,* Die Grundlehren der mathematischen Wissenschaften, Band 196, Springer-Verlag, New York—Heidelberg, 1972. *MR* **51** # 9242.

[1] BURTIN, YU. D., The probability of connectedness of a random subgraph of an n-dimensional cube, *Problemy Peredači Informacii* **13** (1977), 90—95 (in Russian). *MR* **58** # 21833.

[2] ERDŐS, P. and RÉNYI, A., On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17—61. *MR* **23** # A2338.

[3] ERDŐS, P. and RÉNYI, A., Asymmetric graphs, *Acta Math. Acad. Sci. Hungar.* **14** (1963), 295—315. *MR* **27** # 6258.

[4] ERDŐS, P. and RÉNYI, A., On the existence of a factor of degree one of a connected random graph, *Acta Math. Acad. Sci. Hungar.* **17** (1966), 359—368. *MR* **34** # 85.

[5] ERDŐS, P. and SPENCER, J., *Probabilistic methods in combinatorics,* Probability and Mathematical Statistics, Vol. 17, Akadémiai Kiadó, Budapest, 1974; Academic Press, New York—London, 1974. *MR* **52** # 2895.

[6] ERDŐS, P. and SPENCER, J. H., Evolution of the n-cube, *Comput. Math. Appl.* **5** (1979), 33—39. *MR* **80g**: 05054

[7] HART, S., A note on the edges of the n-cube, *Discrete Math.* **14** (1976), 157—163. *MR* **53** # 161.

[8] KOMLÓS, J. and SZEMERÉDI, E., Limit distribution for the existence of Hamilton cycles in a random graph, *Discrete Math.* (to appear).

[9] MARGULIS, G. A., Probabilistic characteristics of graphs with large connectivity, *Problemy Peredači Informacii* **10** (1974), 101—108. *MR* **57** # 12300.

(Received April 22, 1980)

MTA MATEMATIKAI KUTATÓ INTÉZETE
REÁLTANODA U. 13—15.
H—1053 BUDAPEST
HUNGARY

81-4202 — Szegedi Nyomda — Felelős vezető: Dobó József igazgató