

THE LONGEST PATH IN A RANDOM GRAPH

by

Miklós AJTAI, János KOMLÓS

and

Endre SZEMERÉDI

Mathematical Institute of the Hungarian Academy of Sciences

Budapest, Hungary H-1053

Received 12 September 1979

A random graph with $(1+\varepsilon)\frac{n}{2}$ edges contains a path of length cn . A random directed graph with $(1+\varepsilon)n$ edges contains a directed path of length cn . This settles a conjecture of Erdős.

0. Introduction

A. We are going to investigate the length of the longest path (directed path) in a random graph. As a model for randomness, we will work with the random graphs $G_{n,p}$ and $D_{n,p}$, where the edges of the graph $G_{n,p}$ on n vertices are chosen at random, independently of each other, with a probability p , while in $D_{n,p}$ all *directed* edges are chosen independently with a probability p ; thus it may happen that both (x, y) and (y, x) occur in $D_{n,p}$.

Note that the expected number of edges in $G_{n,p}$ is $\binom{n}{2}p \sim \frac{n^2p}{2}$, while in $D_{n,p}$ it is $n(n-1)p \sim n^2p$.

A random graph $G'_{n,N}$ (random directed graph $D'_{n,N}$) with n vertices and N edges (directed edges, possibly going both ways) is defined as a graph that is chosen at random from among all such graphs in such a way that each individual graph has the same probability to be chosen. There is of course an intimate relationship

between $G'_{n,N}$ and $G_{n,p}$, $p = \frac{N}{\binom{n}{2}} \sim 2N/n^2$, and also between $D'_{n,N}$ and $D_{n,p}$, $p = \frac{N}{n(n-1)} \sim N/n^2$.

B. Since

$$\binom{n}{k} k! \left(\frac{\alpha}{n}\right)^{k-1} \sim n\alpha^{k-1} e^{-k^2/2n},$$

the longest path in a random graph with $(1-\varepsilon)\frac{n}{2}$ edges (resp. the longest directed path in a random directed graph with $(1-\varepsilon)n$ edges) is of length $O(\log n)$ with a probability near 1, while in a random graph with $\frac{n}{2}$ edges (directed graph with n edges) it is $O(\sqrt{n \log n})$. (Here we mixed the two different notions $\mathbf{G}_{n,p}$ and $\mathbf{G}'_{n,N}$, see Remark 2.)

C. We are going to show that $(1+\varepsilon)\frac{n}{2}$ edges (or $(1+\varepsilon)n$ directed edges) guarantee the existence of a path (directed path) of length cn with a probability approaching 1 as $n \rightarrow \infty$ ("almost surely", for short).

D. Throughout the paper c and ε will mean positive constants of possibly different values each time we use them and ε is always assumed to be as small as needed in the actual situation.

E. The following results have been conjectured by P. Erdős [4].

Theorem 1. *The random directed graph $\mathbf{D}'_{n,\alpha n}$ with n vertices and αn directed edges, $\alpha > 1$, almost surely contains a directed path of length cn , $c = c(\alpha)$.*

Theorem 2. *The random (undirected) graph $\mathbf{G}'_{n,\beta n}$ with n vertices and βn edges, $\beta > 1/2$, almost surely contains a path of length cn , $c = c(\beta)$.*

Exponential rate. *More precisely we have that for any $\alpha > 1$ there are positive numbers c , K and $\vartheta < 1$ such that the probability that $\mathbf{D}_{n,p}$, $p = \frac{\alpha}{n}$, contains a directed path of length cn is at least $1 - K\vartheta^n$.*

Corollary. *For arbitrarily prescribed $c < 1$ and $\vartheta > 0$, there are α and K such that the exponential rate holds with these parameters.*

Same remark applies for Theorem 1, i.e. for $\mathbf{G}_{n,p}$, $p = \frac{\alpha}{n}$, $\alpha > 1$.

F. Remark 1. We have learned that W. Fernandez de la Vega [3] proved the following theorem:

For $p = 1 - e^{-d/n}$, $\mathbf{G}_{n,p}$ almost surely contains a path of length $(1 - 2.21/d)n$, i.e. our Theorem 2 for $\beta > 1.105$. He also described an algorithm for finding such a long path.

Note that, although his theorem does not cover the case $1/2 < \beta < 1.105$, but for larger β it gives a longer path than ours.

Remark 2. We have stated our theorems in terms of \mathbf{G}' and \mathbf{D}' , but it is much easier to work with \mathbf{G} and \mathbf{D} (the well-known difference between sampling without or with replacement). So we are going to prove the above Exponential rate rather than Theorems 1 and 2. This is, indeed, sufficient since \mathbf{G}' (or \mathbf{D}') can be obtained via \mathbf{G} (or \mathbf{D}) as follows:

Given n and N , define $p = N \binom{n}{2}$. For given $\delta > 0$ set $p_1 = (1 - \delta)p$ and $p_2 = (1 + \delta)p$. Consider a random graph \mathbf{G}_{n,p_1} and modify it as follows: If the number

N_1 of edges in G_{n,p_1} is at most N , then choose $N - N_1$ additional edges at random (i.e. each one of the $\binom{\binom{n}{2} - N_1}{N - N_1}$ possible $(N - N_1)$ -tuples are equally likely to be chosen). If $N_1 > N$, then throw G_{n,p_1} away, and select a random graph with n vertices and N edges.

Obviously, this composite selection method always leads, by symmetry, to the required random graph $G'_{n,N}$, but Step 2 is needed only with a small probability (laws of large numbers), and thus $G'_{n,N}$ almost surely has all those properties of G_{n,p_1} which are not destroyed by adding additional edges (like the existence of a path). Now starting from G_{n,p_2} and deleting $N_2 - N$ edges out of the N_2 edges if $N_2 \geq N$, and starting from scratch if $N_2 < N$, we get that $G'_{n,N}$ almost surely shares all properties with G_{n,p_2} which are not destroyed by deleting edges (like the non-existence of a path). Thus, $G'_{n,N}$, being sandwiched between G_{n,p_1} and G_{n,p_2} , behaves like $G_{n,p}$, if only the problem is "continuous" in p , i.e. G_{n,p_1} and G_{n,p_2} are not too different from the given point of view. (In short, the random graph $G_{n,p}$, $p = N / \binom{n}{2}$, almost surely has $N \pm O(\sqrt{N})$ edges, and $O(\sqrt{N})$ random edges cannot change much.)

Theorems on large deviations show that even the exponential rates are not influenced (only the value of \mathcal{G}) by changing G to G' . Same remark applies for D and D' .

Remark 3. By throwing additional εn edges we can see that a graph with βn edges, $\beta > 1/2$, (directed graph with αn directed edges, $\alpha > 1$) contains a cycle (directed cycle) of length cn , $c = c(\beta)$, (resp. $c = c(\alpha)$) with a probability exponentially near to 1.

Remark 4. We give the proof for the directed case first. This, in turn, immediately implies the undirected case with $\beta > 1$ ($(1 + \varepsilon)n$ edges rather than $(1 + \varepsilon)\frac{n}{2}$ edges), for we can define $G_{n,p}$ through $D_{n,\frac{p}{2}}$ (more precisely, through $D_{n,p'}$, $2p' - (p')^2 = p$) by simply undirecting the directed edges (and counting the double edges only once). But between $\frac{n}{2}$ and n edges we need another method. That is why we chose another way and proved the undirected theorem by introducing a "shrinking method" (Lemma S), which will show that it is sufficient to prove Theorem 2 for any particular (possibly very large) value of β . Actually it will show that Theorem 2 could directly be obtained from the weaker statement that $\varepsilon n \log n$ random edges almost surely ensure the existence of a path of length cn .

This shrinking could also be applied to the directed case, but we preferred to give a direct proof, for we found the applied method — the use of branching processes — worth reproducing here.

We remark that an idea very similar to shrinking is the core of the proof of Theorem 9a in Erdős—Rényi [1], the basic work on random graphs.

G. Finally we show how Corollary follows from Exponential rate. The

proofs for Theorems 1 and 2 (or rather for Exponential rate in the directed and the undirected cases) will be presented in Sections 1 and 2, respectively.

First we show that for *some* $c_0 > 0$ the value ϑ can be made arbitrarily small by choosing α and K large enough. Indeed, let $c_0 > 0$, $\vartheta_0 < 1$, α_0 , $K_0 > 0$ be some values for which Exponential rate holds. Make k independent randomizations with the given probability $p = \frac{\alpha_0}{n}$, which amounts to one randomization with probability $p' = 1 - (1-p)^k \sim kp = \frac{k\alpha_0}{n}$. The probability that none of them produces a path of length $c_0 n$ is less than

$$(K_0 \vartheta_0^n)^k = K(\vartheta_0^k)^n < K\vartheta^n$$

if only k is large enough to ensure $\vartheta_0^k < \vartheta$.

To show that for *any* $c < 1$ the value ϑ can be arbitrarily small, start from the above quadruple $(c_0, \vartheta, \alpha, K)$, with ϑ small. With probability $> 1 - K\vartheta^n$ we have a path P_0 of length $c_0 n$. By making an additional randomization with $p = \frac{\alpha}{(1-c_0)n}$ we have a path P_1 of length $c_1 n = c_0(1-c_0)n$ disjoint from the first path P_0 with a probability $> 1 - K\vartheta^{(1-c_0)n}$. Then a randomization with probability $\frac{\alpha}{(1-c_0-c_1)n}$ produces a path P_2 of length $c_2 n = c_0(1-c_0-c_1)n$ disjoint from the first two paths, with a probability $> 1 - K\vartheta^{(1-c_0-c_1)n}$, etc. In a finite number k of steps (depending only on c_0 and c and *not* depending on ϑ) we get k disjoint paths of lengths $c_0 n, \dots, c_{k-1} n$, $c_0 + \dots + c_{k-1} > c$, with probability $> 1 - K'(\vartheta')^n$, where even ϑ' is small if ϑ was chosen small enough (in terms of c_0 and c). Set $\varepsilon = (c_0 + \dots + c_{k-1} - c)/2k$. Now make a last randomization with probability A/n , where A is chosen in such a way that $e^{-\varepsilon^2 A} < \vartheta'$. The probability that for some i , $0 \leq i \leq k-2$, the last εn points of P_i are not connected to the first εn points of P_{i+1} , is less than

$$ke^{-\varepsilon^2 n^2 A/n} < k(\vartheta')^n,$$

proving the Corollary.

1. The directed case

A. The proof goes along the following lines. We have a fixed numbering v_1, \dots, v_n of our vertices, and have to make a randomization in which the directed edges are randomized independently with probability $p = \alpha/n$. We make this in the following way: we select first the children of v_1 , i.e. the points into which edges go from v_1 . The probability that the children form a given set of k vertices, $0 \leq k \leq n-1$, is $p^k(1-p)^{n-1-k}$. Then we choose the oldest child, i.e. the one with the smallest index, and select his (her) children, i.e. find those vertices, different from v_1 and the children of v_1 , into which edges go from the oldest child. If the oldest child has no children, then we pass to the next child. If it does, then we select the oldest grandchild, and proceed as before. If none of these grandchildren have further children, only then do we go back to the second child. Thus we always go down on the branch of the first-born child as long as we can, and go back on the constructed tree only if

necessary, i.e. if a branch dies out. The main rule is that in defining this tree we *always exclude the points that appeared already* (otherwise we would not get a tree). If the whole tree dies out, we take the smallest point from the rest, and start again. The reason why we proceed in this "first-born children first" manner is that if we construct the whole family tree, then in $\log n$ levels it exhausts almost all points, thus giving us only a path of length $\log n$. Following only the first children however we hope to be able to go down cn generations in this tree. Of course, the constructed tree strongly depends on our *a priori* numbering, the reason for this lack of symmetry is that we randomize only a small portion of the directed edges. To make the procedure complete, we should make a final randomization at the end, where all directed edges whose fate has not yet been decided (like edges going to v_1 or between children, etc.) are randomized. This however does not influence the already existing paths, actually it may make them even longer.

The described process is a little complicated, for the distributions involved are changing from generation to generation and even from point to point, e.g. the probability that v_1 has k children is

$$\binom{n-1}{k} p^k (1-p)^{n-1-k}$$

while the probability that the first child has l children is

$$\binom{n-1-k}{l} p^l (1-p)^{n-1-k-l}$$

where k is the number of children of v_1 ; and it is again different for the second child, etc. This is the price we have to pay for the exclusion of the already appeared points from the next generation. However, this is the only way we can get a tree, i.e. we can maintain independence of the past and the new randomizations. We can overcome this difficulty of changing distributions by allowing only a certain number m (and always the same m) for prospective children.

B. More precisely we proceed as follows. Fix a number $\delta > 0$ to be chosen later (but only in terms of α and not of n). Define $m = (1-\delta)n$, $\lambda = (1-\delta)\alpha$. δ will be chosen so small that $\lambda > 1$. Take v_1 , and let the varying set M of m vertices first consist of the points v_2, \dots, v_{m+1} . Determine the random number **Ch** of children of v_1 in M according to the binomial distribution

$$(1.1) \quad \lambda_k = P(\text{Ch} = k) = \binom{m}{k} p^k (1-p)^{m-k} \sim \frac{(mp)^k}{k!} e^{-mp} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

After having k , let us select k points from M at random (each k -tuple has the probability $1/\binom{M}{k}$ to be chosen). Let us call the one with the smallest index the oldest child or first child and denote it by D (for Daniel). Before selecting the children of D we change M by deleting the k children of v_1 and joining k new points, $v_{m+2}, \dots, v_{m+k+1}$ to M . Then we repeat the procedure with this M , and D in the role of the parent. If D has the children D_1, \dots, D_l then we delete these points from M and add l brand new points, and repeat the procedure with D_1 and M , etc. If at a

stage the actual parent-to-be would not have children, we say that this branch died out, go back one step and try to build the tree there.

The basic rule is that at each step we change M so that it always contains new points, i.e. ones that have never been children at any stage, and M always consists of exactly m vertices.

If the whole tree dies out we take a new point (say the one with the smallest index), and start again. The whole procedure gets stuck only when we do not have enough points left to enlarge M with, i.e. we used up all the δn surplus points. Thus the question is whether this happens earlier than we would have cn generations, i.e. a directed path of length cn .

C. This problem, however, can be formulated without any reference to finitely many points, thus enabling us to formulate the question in the language of branching processes.

We define a branching process (Galton—Watson process) with branching distribution given by (1.1). (This distribution is the only place where the finite number m appears, and since the distribution is asymptotically Poisson, this dependence on m or n is very weak.) From now on we consider λ to be fixed and $p = \frac{\lambda}{m}$.

Now we will calculate (or estimate) three probabilities P_1 , P_2 and P_3 , all depending on n , α and δ , whose sum will bound the probability that we do not have a directed path of length cn in the graph.

Assuming the process dies out after a finite time, let T denote the total population until extinction. Let T_1, T_2, \dots be a sequence of independent random variables all having the same distribution as our T . Set $S_k = T_1 + \dots + T_k$. We define

$$P_1 = P\left(S_{\varepsilon n} > \frac{\delta}{2}n\right),$$

and will show that P_1 is exponentially small if ε was chosen small enough (in terms of α and δ). As is well known, this will follow from

Claim 1. T has a finite moment-generating function, more precisely there exist $t_0 > 0$, $K > 0$, depending only on λ and not on m , such that

$$\mathbb{E}e^{tT} \leq K.$$

This shows us that even if we make εn tries and the trees always die out, which only has an exponentially small probability — this will be our P_3 —, they do not use up more than $\frac{\delta}{2}n$ points altogether, and since $n - m = \delta n$ we still have $\frac{\delta}{2}n$ points to manoeuvre with (to replace the used points of M with).

Now we pass to defining the probability P_2 , which is a little more complicated.

Assume the branching process never dies out, i.e. it has an infinite path (this has a positive probability determined later). We actually need an “ordered” branching process, i.e. for each node we not only make a random decision on how many children it will have (using the distribution (1.1)), but we also order the children-nodes some way. This, of course, can be made after having the whole family-

tree, by simply assigning a random order at each level to the branches. This ordering enables us to speak of left and right.

Now consider the leftmost infinite path **LIP**, in the obvious sense. What we want to show is that until we use up $\frac{\delta}{2}n$ points on the left of **LIP**, we go down cn levels, i.e. we have a cn -path before our original procedure gets completely stuck.

Thus let L_k be the number of points on the left of **LIP** (including those of **LIP**) until the level k .

Claim 2. *There is a $c=c(\lambda)>0$ such that the probability $P_2=P(L_{cn}>\delta n/2)$ is exponentially small.*

(Since all appearing distributions depend on n , we have to clarify that when we say exponentially small, we mean $<K\vartheta^n$, where K and ϑ depend only on α and δ , and of course not on n .)

Since at each step we use up less than **LIP**+**Ch** points, and **Ch** obviously has uniformly bounded moment generating functions, thus for proving Theorem 1 (or rather Exponential bound in the directed case) it is sufficient to show that P_1 , P_2 and P_3 are exponentially small, i.e. to prove Claim 1 (P_1), Claim 2 (P_2) and the fact that the probability Q that our branching process dies out, is less than 1 (remind that $P_3=Q^{en}$). Since $Q=Q(\lambda, m)$, this latter is understood as follows

Claim 3. *For the probability Q that our branching process dies out we have for all m*

$$Q < Q_0 = Q_0(\lambda) < 1.$$

D. Proof of Claim 3. Since $ECh=\lambda>1$, it is well known that $Q<1$, all we have to show is that this inequality is uniform in m . Let $f(x)=\sum_{k=0} \lambda_k x^k$ be the generating function of the distribution (1.1). It is easy to see (and well-known) that Q is the unique root of the equation

$$(1.2) \quad Q = f(Q) \quad \text{on the interval} \quad 0 \leq Q < 1.$$

Now $f(x)=(1-p+px)^m (\sim e^{-\lambda(1-x)})$, and since $f(x)$ is monotone increasing in m , so is Q , and thus $Q < Q_0 < 1$ for all m , where Q_0 is the unique solution on $(0, 1)$ of the equation

$$Q_0 = e^{-\lambda(1-Q_0)}$$

i.e. if $x<1$ is the unique value for which

$$xe^{-x} = \lambda e^{-\lambda}$$

(this x will play an important role in Section 2, just as in the investigation of the largest connected component), then $Q_0 = \frac{x}{\lambda}$.

E. Proof of Claim 1. If the distribution of the total population \bar{T} is $\bar{t}_l = P(\bar{T}=l)$, then we have $\bar{t}_0=0$, $\bar{t}_\infty=1-Q$, and

$$(1.3) \quad \bar{t}_l = \sum_{k=0} \lambda_k \sum_{\substack{l_1+\dots+l_k=l-1 \\ l_j \geq 1}} \bar{t}_{l_1} \dots \bar{t}_{l_k} \quad (l \geq 1),$$

where for $k=0$ the sum after λ_k is 0 unless $l=1$, when it is 1. Multiplying by x^l and summing for $l=1, 2, \dots$ we get for the function $\bar{g}(x) = \sum_{l=1} \bar{t}_l x^l$

$$(1.4) \quad \bar{g}(x) = x f(\bar{g}(x)).$$

This defines $\bar{g}(x)$ uniquely for $x < 1$ (then $\bar{g}(x) < Q$), and by continuity it also defines \bar{g} up to

$$y_0 = \frac{x_0}{f(x_0)} \quad \left(\sim \frac{e^{\lambda-1}}{\lambda} \right),$$

where

$$x_0 = \frac{1-p}{(m-1)p} = \frac{1-p}{\lambda-p} \quad \left(\sim \frac{1}{\lambda} \right)$$

is the maximum of the function $x/f(x)$, if we add the assumption $\bar{g} < x_0$. (Taking $x=1$ we get back to $Q=f(Q)$.) If \mathbf{T} is the total population under the condition that it is finite, i.e.

$$t_l = P(\mathbf{T} = l) = P(\bar{\mathbf{T}} = l | \bar{\mathbf{T}} < \infty) = \bar{t}_l / Q, \quad l = 1, 2, \dots;$$

then for the generating function $g(x) = \sum_{l=0} p_l x^l$ of \mathbf{T} we have $g(x) = \bar{g}(x)/Q$. Differentiating (1.4) we get

$$\bar{g}'(1) = \frac{f(Q)}{1-f'(Q)},$$

whence

$$E\mathbf{T} = g'(1) = \frac{1}{1-f'(Q)} \sim \frac{1}{1-\lambda Q}$$

this is the expected total number of points, assuming the population dies out. This is also true for $\lambda < 1$, when $Q=1$.

Now $Ee^{t_0}\mathbf{T} = g(e^{t_0})$, and thus all we have to show is that for some $a > 1$ (depending on λ but not on m) $g(a)$ is bounded uniformly in m . But this is so, since

$$(1.5) \quad g(y_0) = \frac{x_0}{Q} < \frac{1}{Q}$$

(which is uniformly bounded since Q is increasing in m)

$$y_0 = \frac{x_0}{f(x_0)} \cong \lim_{n \rightarrow \infty} y_0 = e^{\lambda-1}/\lambda > 1.$$

F. Proof of Claim 2. Let \mathbf{T}_L be defined as the number of points in the union of all trees on the left of \mathbf{LIP} which start from the root and do not contain any edge from \mathbf{LIP} . In other words these are the trees hanging on the older brothers of the node which is the second point on \mathbf{LIP} (first one is the root).

Obviously, \mathbf{L}_k is the sum of k independent copies of \mathbf{T}_L . Thus all we have to show is that Claim 1 holds for \mathbf{T} replaced by \mathbf{T}_L , or that for the generating func-

tion $h(x)$ of \mathbf{T}_L we have $a > 1$ and $K < \infty$ depending only on λ such that $h(a) < K$. Let the distribution of \mathbf{T}_L be (h_1, h_2, \dots) , $h(x) = \sum_{i=0}^{\infty} h_i x^i$. Then similarly to (1.3)

$$(1.6) \quad h_l = \sum_{k=0}^{\infty} \lambda'_k \sum_{\substack{l_1 + \dots + l_k = l-1 \\ l_j \geq 1}} t_{l_1} \dots t_{l_k}, \quad l \geq 1 \quad (h_0 = 0),$$

where λ'_k is the probability that the node which originates an infinite path, has k children on the left (strictly) of the leftmost infinite path. Then clearly

$$\lambda'_k = Q^k \sum_{j=k+1}^{\infty} \lambda_j$$

whence one easily gets that

$$(1.7) \quad \tilde{f}(x) = \sum_{k=0}^{\infty} \lambda'_k x^k = \frac{1 - f(xQ)}{1 - xQ}.$$

Now multiplying (1.6) by x^l and summing we get

$$(1.8) \quad h(x) = x \tilde{f}(g(x))$$

and in order to prove that for some $a > 1$

$$h(a) = a \tilde{f}(g(a)) = a \frac{1 - f(g(a)Q)}{1 - g(a)Q}$$

is bounded uniformly in m , we only have to show that $1 - g(a)Q$ is bounded away from 0, uniformly in m , for some $a > 1$. But by (1.5)

$$g(y_0)Q = x_0 = \frac{1-p}{\lambda-p} < \frac{1}{\lambda} < 1, \quad y_0 \equiv \frac{e^{\lambda-1}}{\lambda} > 1$$

and this is sufficient.

G. To get a lower bound on our c in Theorem 1, we calculate the number $E = E\mathbf{T}_L + E\mathbf{Ch}_L$, since for every point on \mathbf{LIP} we have obtained, in the average, $E\mathbf{T}_L - 1$ points on dead - end trees and on the right $E\mathbf{Ch}_R = E\mathbf{Ch}_L$ other points, thus

$$cn \geq \frac{\delta}{2} n/E, \quad \text{i.e.} \quad c \geq \delta/2E.$$

Of course, the number $\delta/2$ in our arguments could have been replaced by any $\delta' < \delta$, and thus taking $c_0 = \delta/E$, it is asymptotically a lower bound for c . Now

$$E\mathbf{T}_L = h'(1) = 1 + g'(1)\tilde{f}'(1) = \frac{1}{1-Q},$$

whence

$$E = E\mathbf{T}_L + \tilde{f}'(1) = \frac{1+Q-Qf'(Q)}{1-Q} \sim \frac{1+Q-\lambda Q^2}{1-Q}.$$

Thus $c_0 \sim \frac{\delta(1-Q)}{1+Q-\lambda Q^2}$ and to get a good lower bound we finally have to specify the value of δ in such a way that it make $\delta(1-Q)/(1+Q-\lambda Q^2)$ small.

Using the equations

$$f(Q) = Q, \quad \lambda = \alpha(1-\delta)$$

some calculations show that (in the limit as $n \rightarrow \infty$) $\delta(1-Q)$ is maximal for the value δ for which $Q=1/\alpha$. This gives $\delta=1-\frac{\log \alpha}{\alpha-1}$, and thus

$$(1.9) \quad c_0 \cong \frac{(\alpha-1)-\log \alpha}{\alpha+\delta} > \frac{(\alpha-1)-\log \alpha}{\alpha+1}.$$

H. In other words we have proved that for $c < c_0$, the graph $\mathbf{D}_{n,p}$, $p=\alpha/n$, contains a path of length cn with a probability $>1-K\vartheta^n$, $K=K(\alpha, c)$, $\vartheta=\vartheta(\alpha, c) < 1$.

Note that for $\alpha=1+\varepsilon$, $\varepsilon \sim 0$, $c_0 \gtrsim \frac{\varepsilon^2}{2}$; while for $\alpha \sim \infty$, $c_0 \gtrsim 1 - \frac{\log \alpha}{\alpha}$. This is much worse than $1 - \frac{1}{\alpha}$, that one expects for any $\alpha > 1$.

2. The undirected case

Lemma S. (*Shrinking lemma*)

Given $\alpha > 1$, there are positive numbers $\delta=\delta(\alpha)$, $\gamma=\gamma(\alpha)$, $0 < \vartheta=\vartheta(\alpha) < 1$ and $K=K(\alpha)$ such that the random (undirected) graph $\mathbf{G}_{n,p}$, $p=\alpha/n$, contains at least δn vertices that are covered by disjoint cycles of lengths $> \gamma \log n$ with a probability $> 1 - K\vartheta^n$.

Actually, any value $\delta < x_0$ is good for $\delta(\alpha)$, where $0 < x_0 < 1$ is the root of the equation

$$f(x) = 1/2\alpha + x \log(1/x) + (1-x) \log \alpha - (1-x)^2 \frac{\alpha}{2} = 0$$

and the values γ , ϑ , K are chosen appropriately (in terms of α , δ).

We say that a graph is a forest if it contains no cycles, i.e. if it is a disjoint union of trees. For proving Lemma S we are going to use the following estimation.

Lemma 2.

For the random graph $\mathbf{G}_{n,p}$, $p=\alpha/n$ we have

$$P(\mathbf{G}_{n,p} \text{ is a forest}) < 2\sqrt{n}\vartheta_F^n,$$

where $\vartheta_F=\vartheta_F(\alpha)=\alpha e^{1/2\alpha}e^{-\alpha/2}$ and $n > 4\alpha^2$.

Proof of Lemma 2. We use the following exponential generating function for $G_k(n)$, the number of graphs with n labelled vertices consisting of k disjoint trees (cf. Rényi [2]):

$$(2.2) \quad \sum_{n=0}^{\infty} G_k(n) \frac{x^n}{n!} = \frac{\left(y - \frac{y^2}{2}\right)^k}{k!} \quad (k = 0, 1, 2, \dots)$$

where

$$x = ye^{-y}, \quad 0 < y < 1; \quad G_0(0) = 1, \\ G_0(n) = 0 \quad \text{for } n > 0; \quad G_k(0) = 0 \quad \text{for } k > 0.$$

Since a forest with k components has $n-k$ edges, we have

$$\begin{aligned} P(\mathbf{G}_{n,p} \text{ is a forest}) &= \sum_{k=1}^n G_k(n) p^{n-k} (1-p)^{\binom{n}{2}-n+k} = \\ &= \left(\frac{p}{1-p}\right)^n (1-p)^{\binom{n}{2}} \sum_{k=1}^n G_k(n) \left(\frac{1-p}{p}\right)^k. \end{aligned}$$

Now multiplying (2.2) by t^k and summing for k we obtain

$$\sum_{n=0}^{\infty} \left(\sum_{k=0}^n G_k(n) t^k \right) \frac{x^n}{n!} = e^{t\left(y - \frac{y^2}{2}\right)}.$$

Hence for any particular n and $0 < y < 1$

$$\sum_{k=0}^n G_k(n) t^k \leq n! (ye^{-y})^{-n} \cdot e^{t\left(y - \frac{y^2}{2}\right)}$$

and applying the latter inequality with $y=1$ and $t = \frac{1-p}{p} = \frac{n}{\alpha} - 1$ (one can take $y=1$ because of continuity); we get the lemma.

Note that ϑ_F is strictly monotone decreasing in α for $\alpha > 0$, and thus $\vartheta_F < 1$ iff $\alpha > 1$.

Proof of Lemma S. Consider a collection of disjoint cycles maximal w.r. to the total number of vertices covered. Let us write the number m of these covered vertices in the form $m = xn$. We have thus in $\mathbf{G}_{n,p}$ a subgraph of $(1-x)n$ vertices, which contains no cycles. The probability of this event is, according to Lemma 2.1, at most

$$\binom{n}{(1-x)n} 2\sqrt{n} [\vartheta_F(\alpha(1-x))]^{(1-x)n} \approx e^{nf(x)}.$$

Note that $f(x)$ is strictly monotone increasing on $[0, x_0]$. Hence Lemma S and the remark after are proved, except for the part that these covering cycles are large. So it remains to remark that for any $\delta_1 > 0$ there is a $\gamma > 0$ such that the probability that $\delta_1 n$ vertices in $\mathbf{G}_{n,p}$ are covered by disjoint cycles of length not exceeding $\gamma \log n$ is less than 2^{-n} . Indeed, the probability that the whole graph is covered with such cycles is less than

$$\frac{2^n n! p^n}{\left(\frac{n}{\gamma \log n}\right)!} < (npe^{-\frac{1}{\gamma}})^n,$$

for a permutation and a subset of indices always define a partition into cycles, but this way each partition is counted at least $\left(\frac{n}{\gamma \log n}\right)!$ times.

Thus, the above probability is less than

$$\binom{n}{\delta_1 n} (\delta_1 n p e^{-\frac{1}{\gamma}})^{\delta_1 n} < 2^{-n}$$

if γ is small enough.

Proof of Theorem 2. We will drop additional εn edges, and see that the $m > \delta n$ vertices that are covered by long cycles (Lemma S) will contain an almost Hamiltonian path. More precisely: For arbitrarily small positive numbers ε , δ_1 and ϑ there is a constant $K = K(\varepsilon, \delta_1, \vartheta)$ such that if we drop additional εn edges (make an additional randomization with $p = 2\varepsilon/n$) to the graph then the subgraph consisting of a disjoint union of long cycles that is ensured by Lemma S, will contain a path with at least $(1 - \delta_1)m$ of these m (say) vertices with probability exceeding $1 - K\vartheta^n$.

Let us consider the above union of cycles, and cut them into arcs of length L (throw away the leftover), where L is so large that even the number $\alpha' = 2\varepsilon\delta_1^2 L$ is large (in terms of δ_1). Divide each arc into three arcs of length l , $L - 2l$, l , ($l = \delta_1 L$), and call them the head, the middle and the tail of the arc. Now drop the additional εn edges. We define a directed graph **D** whose vertices are the above arcs of length L , two arcs A_1 and A_2 will be connected with a directed edge $A_1 A_2$ if at least one of the additional edges connects the tail of A_1 with the head of A_2 . The obtained shrunk graph **D** contains $n' = \frac{m}{L} \cong (1 - o(1)) \frac{\delta n}{L}$ vertices, and is very dense; the probability that A_1 is connected to A_2 is

$$p' = 1 - \left(1 - \frac{2\varepsilon}{n}\right)^{l^2} \cong (1 - o(1)) \frac{2\varepsilon\delta_1^2 L}{Ln'} \approx \frac{2\varepsilon\delta_1^2 L}{n'} = \frac{\alpha'}{n'}.$$

Thus, by Theorem 1, **D** contains a directed path of at least $(1 - \delta_1)n'$ of its "points". This gives a path of length $\cong (1 - \delta_1)n'(L - 2l) > (1 - 3\delta_1)m$ in the original graph with probability $> 1 - K\vartheta^n$ if L , and thus α' too, are large enough.

Remark 5. The proof has shown that outside the longest path (or cycle) we can typically expect a graph in which the longest path is of length $o(n)$. The only natural reason for having only such short paths is that in the remaining graph of m points the probability of edges is only $\frac{1}{m}$. I.e. we would have $\frac{\alpha}{n} = \frac{1}{m}$ or $m = \frac{n}{\alpha}$. In other

words, a natural guess for the maximum length of paths is $\left(1 - \frac{1}{\alpha}\right)n + o(n)$. (An

upper bound, anyway, is $G(\alpha)n$, where $G(\alpha) = 1 - \frac{x}{\alpha}$, $xe^{-x} = \alpha e^{-\alpha}$, ($0 < x < 1$); since this is the size of the largest connected component with probability $1 - o(1)$.) (See Erdős—Rényi [1], Theorem 9.B.)

References

- [1] P. ERDŐS and A. RÉNYI, On the evolution of random graphs, *Publications of the Math. Inst. the Hung. Acad. of Sci.*, 5 (1960), 17—61.
- [2] A. RÉNYI, Some remarks on the theory of trees, *Publications of the Math. Inst. of the Hung. Acad. of Sci.*, 4 (1959), 73—85.
- [3] W. F. DE LA VEGA, Sur la plus grande longueur des chemins élémentaires de graphes aléatoires, *Preprint of Laboratoire d'informatique pour les Sciences de l'Homme, C.N.R.S.*, 1979.
- [4] P. ERDŐS, Problems and results on finite and infinite graphs, in *Proc. Symp. Prague 1974*, Akademia Praha 1975.