ISRAEL JOURNAL OF MATHEMATICS, Vol. 62, No. 3, 1988

WHEN ARE RANDOM GRAPHS CONNECTED

BY
S. KALIKOW AND B. WEISS
Institute of Mathematics, The Hebrew University of Jerusalem, Jerusalem, Israel

ABSTRACT
Several results concerning the connectivity of infinite random graphs are considered. A necessary sufficient condition for a zero–one law to hold is given when the edges are chosen independently. Some specific examples are treated including one where the vertex set is N and the probability that an edge joining i to j is present depends only on $|i-j|$.

The random graphs that will be considered here are formed by taking a fixed vertex set V and then deciding for each possible edge $\{u, v\}$ whether or not the edge is actually in the random graph $\Gamma(\omega)$ by tossing independent coins, one for each possible edge, with success probability p_{uv}. The nature of the random graph so formed will depend of course on the nature of these probabilities $\{p_{uv}\}$. We will be concerned with one of the simplest questions one can ask about $\Gamma(\omega)$ — namely is it connected or not. If the vertex set V is finite the probability of this event is usually some number between 0 and 1 which is not easy to evaluate. We work throughout with V infinite when we might expect some zero–one laws to hold. Indeed our first main result is a rather simple necessary and sufficient condition for the validity of a zero–one law for this question. The condition is formulated directly in terms of the p_{uv}'s and asserts that for any partition of the vertex set into non-empty sets A, B we have that

$$\sum_{\substack{a\in A\\ b\in B}} p_{ab} = +\infty$$

which is equivalent of course to the assertion that for any A, B as above $\mathrm{Prob}\{A \text{ is connected to } B\} = 1$.

Received February 1, 1988

Next we take $X = N$ and suppose that p_{uv} depends only on $|u - v|$, say $p_{uv} = p_{|u-v|}$. Here our main result is that if

$$\sum_{n=1}^{\infty} p_n = +\infty, \quad \text{and} \quad \text{g.c.d.}\{n : p_n > 0\} = 1,$$

then with probability one the graph is connected. This result for $V = Z$ was obtained several years ago by Grimmett, Marstrand and M. Keane. Their proof didn't work for N although it did extend to analogous results for Z^2 where the method we use in §2 breaks down.

Finally we consider an inhomogenous case on N that was suggested by Lester Dubins. Here we introduce a parameter $\lambda > 0$ and define

$$p_{ij} = \left(\frac{\lambda}{\text{Max}\{i, j\}}\right) \wedge 1.$$

A heuristic argument suggests that for $\lambda > 1$ the graph should be connected while for $\lambda < 1$ the graph should be disconnected. (L. Dubins' original question was the case $\lambda = 1$.) We can establish the first assertion, but instead of the second can only show the disconnectedness for $\lambda < \frac{1}{4}$. The appearance of a critical phenomenon of this type is reminiscent of the situation when covering the circles with random arcs (see [S]) but we weren't able to find any deeper connections. The study here was sparked by some talks given by M. Keane at MSRI in 1983–4 and we thank him for introducing us to this fascinating subject and for several discussions on the ideas in this paper.

§1. The zero–one law

Here is a simple example that shows that the zero–one law is not universally valid:

$$p_{12} = \tfrac{1}{2}; \quad p_{1j} = 0 \quad \text{all } j \geqq 3; \quad p_{ij} = 1 \quad \text{all } 2 \leqq i < j.$$

Clearly $p(\Gamma(\omega)$ is connected$) = \frac{1}{2}$, since all hinges upon whether or not $\{1\}$ is connected to the rest of the vertices. More generally such a situation can arise whenever there is a partition of N into two sets A, B with

$$0 < p(A \text{ is connected to } B) < 1.$$

Since the possibility $p(A$ is connected to $B) = 0$ is anyway a triviality for the connectedness question we may take for a necessary condition that $p(A$ is connected to B) = 1 for any partition $A \cup B = N$. By the usual Borel–Cantelli lemma this is equivalent to

$$\sum_{\substack{a \in A \\ b \in B}} p_{ab} = +\infty.$$

We see that this is a necessary condition for $\Gamma(\omega)$ to be connected with probability one. The next theorem says that this is sufficient.

THEOREM 1. *If for any partition of N into sets A, B we have*

$$\sum_{\substack{a \in A \\ b \in B}} p_{ab} = +\infty$$

then

$$p(\Gamma(\omega) \text{ is connected}) \in \{0, 1\}.$$

Moreover, if $p(\Gamma(\omega)$ is connected$) = 0$ then with probability one $\Gamma(\omega)$ has infinitely many components.

We shall denote by $L(a, b)$ the event that a and b are linked, so that our basic model is that the events $L(a, b)$ are independent events and $p(L(a, b)) = p_{ab}$. By $C(a, b)$ we shall denote the event that a and b are connected by some finite chain of links. Note that the event *"the graph $\Gamma(\omega)$ has infinitely many components"* is a tail event with respect to the collection of independent events $L(a, b)$ and so has probability either zero or one. To prove the theorem it therefore suffices to show that the event

(*) *"the graph $\Gamma(\omega)$ has preciesly n components"*

has probability zero for all $n \geqq 2$.

LEMMA 1. *If for some $n \geqq 2$, (*) has positive probability then*

(**) *"the graph $\Gamma(\omega)$ has precisely two components"*

has positive probability.

PROOF. Observe that for any a and b $p(C(a, b)) > 0$, since for fixed a, if A denotes the set of such b's then clearly A is not empty and if A is not all of N, the partition A, $N \setminus A$ would contradict the hypotheses of the theorem. Next consider all possible n tuples $(e_1, e_2, \ldots, e_n)$ and for each consider the event $E(e_1, \ldots, e_n)$ that there are precisely n components and each e_i is in a distinct one. Since there are only countably many such n-tuples, for some choice we must have

$$p(E(e_1, \dots, e_n)) > 0.$$

Now by the observation above, there is some chain from e_1 to e_2, say $e_1 = f_0$, $f_1, \dots, f_k = e_2$ such that $p_{f_i, f_{i+1}} > 0$ for $0 \leqq i < k$. For some $i_0 < k$ if we modify the event $E(e_1, \dots, e_n)$ by connecting all pairs f_i, f_{i+1} for $0 \leqq i \leqq i_0$ we will reduce the number of components by exactly one with positive probability. Thus the fact that (∗) has positive probability for some n implies that also (∗) with $n-1$ has positive probability which by induction proves the lemma. □

We assume, therefore, that for some fixed $u, v \in N$ the event

$$E(u, v) = \{\Gamma(\omega) \textit{ has precisely two components, } u \textit{ and } v \textit{ are in distinct components}\}$$

has positive probability and proceed to derive a contradiction. By the lemma and the preceding remarks this will prove the theorem. Since the events $L(a, b)$ generate the σ-algebra, it follows that for n sufficiently large there is some choice of either $L(a, b)$ or $0(a, b)$ (the *complement* of $L(a, b)$, namely the event that the link (a, b) is open) for all $a < b \leqq n$ such that if S denotes the intersection of all these events,

$$(*) \qquad p(E(u, v) \mid S) > .999.$$

Fix $i \in N$, and notice that the events $C(i, u)$, $C(i, v)$ conditioned on S are monotonic events so that by T. Harris' correlation inequality ([K] p. 72)

$$p(C(i, u) \text{ and } C(i, v) \mid S) \geqq p(C(i, u) \mid S) p(C(i, v) \mid S).$$

Thus if both $p(C(i, u) \mid S) > .1$ and $p(C(i, v) \mid S) > .1$ we would get $p(C(u, v) \mid S) > .01$ contradicting (∗). Thus at least one of them is at most .1. Again by (∗) we have that the probability that i is connected to either u or v given S is at least .9 so that we have for each i one of the following two possibilities:

$$\text{(I)} \qquad \begin{cases} p(C(i, u) \mid S) \leqq .1, \\ p(C(i, v) \mid S) \geqq .8, \end{cases}$$

or

$$\text{(II)} \qquad \begin{cases} p(C(i,u) \mid S) \geqq .8, \\ p(C(i,v) \mid S) \leqq .1. \end{cases}$$

Define a partition of N by putting $i \in A$ if (I) holds and $i \in B$ if (II) holds. Consider now a list of all possible pairs (a_i, b_i), $1 \leqq i < +\infty$ where $a_i \in A$, $b_i \in B$ and S doesn't specify whether $L(a_i, b_i)$ or $0(a_i, b_i)$ holds. By the basic hypothesis of the theorem almost every ω belongs to at least one event $L(a_i, b_i)$ and we denote by

$$E_i = \{\omega : \omega \in L(a_i, b_i), \omega \in 0(a_j, b_j) \text{ for all } j < i\}.$$

Since the events E_i are disjoint we conclude that

$$\sum_{i=1}^{\infty} p(E_i \mid S) = 1.$$

Let k be the first index such that

$$\sum_{i=1}^{k} p(E_i \mid S) \geqq \tfrac{1}{2} \quad \text{and} \quad \sum_{i=k}^{\infty} p(E_i \mid S) \geqq \tfrac{1}{2}.$$

Since $0(a_1, b_1) \cap 0(a_2, b_2) \cap \cdots \cap 0(a_i, b_i) = \bigcup_{i+1}^{\infty} E_j$ we have

$$p(0(a_1, b_1) \cap \cdots \cap 0(a_j, b_j) \mid S) \geqq \tfrac{1}{2} \qquad \text{for } j < k$$

and so (*) implies that

$$p(E(u,v) \mid 0(a_1, b_1) \cap \cdots \cap 0(a_j, b_j) \cap S) > .998.$$

Clearly

$$p(C(a_{j+1}, u) \mid 0(a_1, b_1) \cap \cdots \cap 0(a_j, b_j) \cap S) \leqq p(C(a_{j+1}, u) \mid S) \leqq .1$$

so that by the argument that led to (I) and (II) we have

$$p(C(a_{j+1}, v) \mid 0(a_1, b_1) \cap \cdots \cap 0(a_j, b_j) \cap S) \geqq .8$$

and thus

$$p(C(a_{j+1}, v) \mid E_{j+1} \cap S) \geqq .8.$$

Similarly

$$p(C(b_{j+1}, u) \mid E_{j+1} \cap S) \geqq .8$$

and so

$$p(C(a_{j+1}, v) \cap C(b_{j+1}, u) \mid E_{j+1} \cap S) \geqq .6.$$

However, given E_{j+1}, $C(a_{j+1}, v)$ and $C(b_{j+1}, u)$ imply $C(u, v)$, and thus we conclude that

$$p(C(u, v) \mid E_{j+1} \cap S) \geqq .6$$

for all $j < k$. Since trivially

$$p(C(u, v) \mid E_1 \cap S) \geqq .6$$

we obtain

$$p(C(u, v) \mid \left(\bigcup_1^k E_i \cap S \right) \geqq .6$$

and since $p(\bigcup_1^k E_i \mid S) \geqq \frac{1}{2}$ we finally obtain

$$p(C(u, v) \mid S) \geqq .3$$

contradicting (*) and proving the theorem.

§2. Homogeneous random graphs on N

In this section we analyze the connectivity properties of random graphs on N when the probability of joining i to j depends only on $|i - j|$. There is a dichotomy, depending upon whether these probabilities form a convergent or a divergent series. In the first case the assumptions of the theorem in §1 do not apply so that while it is completely trivial to see that there is a positive probability that the graph is not connected, a slight new argument is required to show that this probability, in fact, equals one. Let us denote by p_n the probability that i and j are connected for $|i - j| = n$. Then for any k the event E_k = "the vertex k is isolated in $\Gamma(\omega)$" has probability at least

$$p = \prod_1^\infty (1 - p_n)^2.$$

If p is positive, then it is not too hard to see that whenever $n_1 < n_2 < \cdots < n_N$ are sufficiently far apart the events $E_{n_1}, E_{n_2}, \ldots, E_{n_N}$ are nearly independent so that

$$p(E_{n_1} \cup E_{n_2} \cup \cdots \cup E_{n_N})$$

is approximately $1-(1-p)^N$ and thus the probability of $\bigcup_1^\infty E_n$ is equal to one which means that the graph $\Gamma(\omega)$ is not connected with probability one.

The remainder of this section is devoted to the proof of the following theorem which deals with the remaining case.

THEOREM 2. *If* $\Sigma_1^\infty p_n = +\infty$, $1 = \text{g.c.d.}\{n : p_n > 0\}$ *and* $p_{ij} = p\,|\,i-j\,|$ *then with probability one* $\Gamma(\omega)$ *(the random graph formed by joining* $i, j \in N$ *independently with probabilities* p_{ij}*) is connected.*

We shall need an elementary lemma on nonnegative sequences which we leave as an exercise:

LEMMA 2. *If* $\{a_n \geqq 0\}_1^\infty$, *then for any* $\varepsilon > 0$ *there exist infinitely many indices* n *for which* $a_{n-1} - a_n < \varepsilon$.

Denote by $\Gamma_n(\omega)$ the random graph on $\{0, 1, \ldots, n\}$ obtained by performing the experiments which determine whether or not i is linked to j for all $0 \leqq i < j \leqq n$. Let $X_n(\omega)$ denote the number of components of $\Gamma_n(\omega)$ and denote

$$a_n = E(X_n).$$

Fix some small $\varepsilon > 0$ and then by the lemma find an n so that $a_{n-1} - a_n < \varepsilon$. Now

$$\varepsilon > a_{n-1} - a_n = E(X_{n-1} - X_n) = E((X_{n-1} - X_n)^+) - E((X_{n-1} - X_n)^-)$$

and $(X_{n-1} - X_n)^-$ is positive only when $X_n > X_{n-1}$. For this to happen "n" must be an isolated vertex in $\Gamma_n(\omega)$ and this event has probability $\Pi_{i=1}^n (1-p_i)$. Thus if n is large enough we can certainly assert that

$$(*) \qquad E((X_{n-1} - X_n)^+) \leqq 2\varepsilon.$$

Next observe that if $(X_{n-1} - X_n)^+$ is positive it must equal one or more and so $(*)$ implies that with probability at least $1 - 2\varepsilon$ we have $X_{n-1} = X_n$. By the homogeneity we can interchange the role of 0 and n, so if $\overline{\Gamma}_{n-1}$ denotes the random graph on $\{1, 2, \ldots, n\}$ and $\overline{X}_{n-1}$ the number of its components we have the same statement concerning $\overline{X}_{n-1}$ and X_n. What this means is that, if we look at the component of $\Gamma_n(\omega)$ that contains 0 and cut all the links to 0, then the component less $\{0\}$ remains connected. Finally fix i_0 with $p_{i_0} > 0$ and suppose that n is large compared to i_0. Set

$$E_n = \{\text{in } \Gamma_n(\omega), 0 \text{ is connected to some vertex different from } i_0, \text{ and } i_0 \text{ is not connected to 0 by any path that avoids the link } 0 - i_0\}.$$

Since $\Sigma_1^\infty p_k = +\infty$, the first condition in E_n has probability that approaches one as $n \to \infty$. By the independence, we have

$$p(E_n \text{ and the link } 0 - i_0 \text{ is closed}) = p_{i_0} \cdot p(E_n).$$

We have just seen that this last event has probability at most 2ε and therefore $p(E_n) \to 0$. This means that with probability one 0 is connected to any i_0 with $p_{i_0} > 0$. Our arithmetic hypothesis on the p_n's now concludes the proof. □

In case the arithmetic hypothesis fails to hold, we get a finite number of components consisting of the various residue classes modulo the g.c.d.$\{n : p_n > 0\}$.

§3. An inhomogeneous random graph

Once again the vertex set is N and now with $\lambda > 0$ a positive parameter, we define

$$p_{ij}(\lambda) = \max\{1, \lambda/\max\{i, j\}\}.$$

The original question raised by Lester Dubins was: is the graph connected with probability one when $\lambda = 1$? Our first result will be that if $\lambda > 1$ then almost surely the graph is connected. For this we will need one of the results of Erdös and Renyi in their fundamental study of finite random graphs ([ER]). The result we need appears there as part of a very detailed analysis of the behavior of random graphs over a wide range of the parameters. For the reader's convenience we will include a short direct proof of a less precise result (which suffices for our purposes). We follow a suggestion made by Eli Shamir.

THEOREM 1 (Erdös–Renyi). *If $p > 1$, then there is some positive constant γ (for example $\gamma = (p-1)/3p$) such that*

$$\lim_{n \to \infty} \text{Prob}\{\textit{in the random graph on } n \textit{ vertices where each edge is present independently with probability } p/n \textit{ there is a connected component of size at least } \gamma n\} = 1.$$

PROOF. Choose $\gamma > 0$ so that $p(1 - 2\gamma) > 1$. We will estimate the probability in question by a dynamic process in which we sequentially choose to reveal more and more of the random connections in the graph. To begin with we fix some set of vertices of size M and reveal the connections between these vertices. Fix some $\varepsilon > 0$ and some a_0 which will be specified shortly. If M is large enough then, with probability $(1 - \varepsilon)$, there will be a connected component here of size at least a_0.

Now we look at all the possible connections between this component (say the largest one in the random graph on M vertices) and the remaining $n - M$ vertices. Denote by A_1 the set of vertices that are connected to this component and let a_1 denote the size of A_1. If $a_0 + a_1 \geqq \gamma n$ we stop. Otherwise we look now at the connections between A_1 and the remaining $n - M - a_1$ vertices. Denote the set of vertices so connected by A_2, and $a_2 = |A_2|$. If $a_0 + a_1 + a_2 \geqq \gamma n$ we stop, otherwise we continue as before and define A_k, $a_k = A_k$. Naturally if A_1 is empty we can't go beyond the first step and, in general, if at any stage $A_k = \emptyset$ the procedure stops. What we shall see is that, with probability close to one, the procedure continues until we see a component of size $\geqq \gamma n$.

Let $m = n - M - a_1 - \cdots - a_k$. Naturally m is a random variable and depends upon k but we suppress this dependence. For each of the new m-vertices let X_j $(j = 1, 2, \ldots, m)$ be a random variable that equals 1 if the j-th vertex is connected to some element of A_k and 0 if the j-th vertex is not so connected. Then clearly the X_j are independent identically distributed random variables and

$$a_{k+1} = \sum_{j=1}^{m} X_j$$

so that given $A_1, \ldots, A_k$; a_{k+1} is a binomially distributed random variable with parameter

$$\mathrm{Prob}(X_j = 1) = 1 - \left(1 - \frac{p}{n}\right)^{a_k} \cong \frac{pa_k}{n}$$

(to simplify the writing we drop the higher-order terms which are negligible). Thus

$$E(a_{k+1} \mid A_1, \ldots, A_n) \cong a_k \cdot p \, \frac{m}{n}.$$

Now if n is large enough so that $M/n < \gamma$, there is some fixed constant $c > 1$ so that

$$p\,\frac{m}{n} \geqq c.$$

Furthermore, a direct estimation of the relevant binomial coefficient using Stirling's formula gives the existence of some $\delta > 0$ so that

$$\text{Prob}\left(a_{k+1} \leqq \left(\frac{1+c}{2}\right) a_k \mid A_1, \ldots, A_k\right) \leqq e^{-\delta a_k}.$$

It follows that with probability at least $(1 - e^{-\delta a_k})$

$$a_{k+1} \geqq \left(\frac{1+c}{2}\right) a_k.$$

It is now clear that if a_0 is chosen large enough (the above calculation is valid of course for a_1 as well) with probability at least $1 - \varepsilon$, the A_k's grow by a fixed multiple so that the procedure continues until a component of size at least γn is seen. □

Fix now $\lambda > 1$, and set $B = 2\lambda/(\lambda - 1)$. For $n = 1, 2, \ldots$ consider the block of integers between B^n and B^{n+1}; there are $B^n(B-1)$ integers there and the probability that any two are connected is at least λ/B^{n+1}. Thus, comparing with the Erdös–Renyi situation we see that with $p = (1+\lambda)/2$ we have a lower bound on the probability that there is a connected component within the block (B^n, B^{n+1}) whose size is some constant γB^{n+1}. Fix now the vertices 1 and k_0, and consider only those blocks with $B^n > k_0$. Now we check to see whether or not both 1 and k_0 are connected to the large component in (B^n, B^{n+1}). The preceding analysis shows that there is some positive constant $\delta > 0$ such that

$$p_r\{\text{both 1 and } k_0 \text{ are connected to the large component in } (B^n, B^{n+1})\} \geqq \delta.$$

But clearly these events are independent and therefore by the Borel–Cantelli lemma we have that 1 is connected to k_0 with probability 1. Thus we have proved:

Theorem 2. *For $\lambda > 1$, the random graph formed on the vertex set $\{1, 2, \ldots\}$, with the edge (i, j) present, (independently) with probability* $\max(1, \lambda/\max(i, j))$, *is connected with probability* 1.

For $\lambda < 1$ the situation is that the probability that n is connected to a vertex $< n$ is less than 1, and this suggests that there are no infinite connected chains. However we haven't been able to make this argument precise and the best we can do is show that with probability 1 the graph is not connected when $\lambda < \frac{1}{4}$. For this we need an elementary estimate on the norm of the matrix $A = (a_{ij})_{1,1}^{\infty}$ defined by

$$a_{ij} = \begin{cases} 0 & \text{if } j = 1 \\ \dfrac{1}{\max(i,j)} & \text{if } j \neq 1 \end{cases}$$

acting on l_2. It's a little easier to estimate the norm of $A + D$ where D is the diagonal matrix with $d_{ii} = 1/i$, which is clearly greater than the norm of A since both A and D are positive symmetric matrices. Let B be the lower triangular matrix defined by

$$b_{ij} = \begin{cases} \dfrac{1}{i} & j \leqq i \\ 0 & j > i \end{cases}$$

and observe that $BB^* = A + D = B + B^* - D$. It follows that $(I - B)(I - B)^* = I - D$ whence $\| I - B \| \leqq 1$, and then $\| B \| \leqq 2$, $\| BB^* \| \leqq 4$ and we have proved

Lemma 3. *With A defined as above $\| A \| \leqq 4$.*

From here we see that with $\lambda < \frac{1}{4}$, the series $\Sigma_1^{\infty} \lambda^n A^n$ converges in norm, and thus for any $\varepsilon > 0$ there are indices i, j such that

$$(*) \qquad \sum_1^{\infty} (\lambda^n A^n)_{ij} < \varepsilon.$$

This last expression contains in particular, for any simple path from i to j, a term giving the probability that all edges in that path are connected. Thus the left-hand side of $(*)$ is an upper bound for the probability that i is connected to j. It follows that the probability that the graph is connected is less than ε. Since ε is arbitrary we have proved.

Proposition 4. *For $\lambda < \frac{1}{4}$ the random graph described above is disconnected with probability one.*

While we are fairly certain that $\frac{1}{4}$ can be replaced by 1, what happens at the critical value $\lambda = 1$ remains for us a mystery.

REFERENCES

[ER] P. Erdös and A. Renyi, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutato Int. Kozl. **5** (1960), 17–61.

[K] H. Kesten, *Percolation Theory for Mathematicians*, Birkhauser, 1982.

[S] L. Shepp, *Covering the circle with random arcs*, Isr. J. Math. **11** (1972), 328–345.