Local Connectivity of a Random Graph

P. Erdös
HUNGARIAN ACADEMY OF SCIENCES, BUDAPEST, HUNGARY
E. M. Palmer
MICHIGAN STATE UNIVERSITY, EAST LANSING, MI 48824
R. W. Robinson
SOUTHERN ILLINOIS UNIVERSITY, CARBONDALE, IL 62901

ABSTRACT

A graph is *locally connected* if for each vertex v of degree ≥ 2, the subgraph induced by the vertices adjacent to v is connected. In this paper we establish a sharp threshold function for local connectivity. Specifically, if the probability of an edge of a labeled graph of order n is $p = ((3/2 + \varepsilon_n) \log n/n)^{1/2}$ where $\varepsilon_n = (\log \log n + \log(3/8) + 2x)/(2 \log n)$, then the limiting probability that a random graph is locally connected is $\exp(-\exp(-x))$.

INTRODUCTION

A graph is *locally connected* if for each vertex v of degree ≥ 2, the subgraph induced by the vertices adjacent to v is connected. Our sample space, denoted Ω_n, consists of all $2^{\binom{n}{2}}$ labeled graphs G, if order n. Let $p = p(n)$ be a number between 0 and 1. If G has size q, i.e., q is the number of edges, then the probability of G is defined by

$$P(G) = p^q(1 - p)^{\binom{n}{2} - q}, \tag{1}$$

and we refer to p as the probability of an edge.

Our aim is to study the limiting probability, $\lim_{n\to\infty} P(C)$, where C is the subset of Ω_n which consists of the locally connected graphs. If p is fixed, then the methods of [BIH79] show immediately that the limit above

Journal of Graph Theory, Vol. 7 (1983) 411–417

CCC 0364-9024/83/040411–07$01.70

is 1, i.e., "almost all graphs are locally connected." Therefore we will consider only $p(n)$ such that $\lim_{n\to\infty} p(n) = 0$.

Note that local connectivity is not a monotone property. That is, if G is locally connected and H is obtained by adding an edge to G, then H is not necessarily locally connected whether G is connected or not. Nevertheless, we find two threshold functions for local connectivity. The first and less interesting occurs when the probability of an edge is so small that almost every graph consists of isolated edges and vertices. But the second threshold is non-trivial and appears when the probability of an edge is sufficiently high to cause every edge to belong to a triangle.

For the most part, our notation and terminology follow that of the book [Bo79] where one can also find an introduction to the methods used here. We also assume some familiarity with the fundamental results of the senior co-author in [ErR59] and [ErR60].

THE LOW THRESHOLD

A graph consisting of only isolated edges and vertices is locally connected. If the probability of an edge is sufficiently low, almost all graphs will be of this sort. This remark takes the following precise form.

Theorem 1. If the probability of an edge is $p = 2 \times n^{-3/2}$, then

$$\lim_{n\to\infty} P(C) = e^{-2x^3}, \tag{2}$$

i.e., $2 \times n^{-3/2}$ is a sharp threshold function for local connectivity.

Proof. Let A be the set of graphs in Ω_n which have at least one component which is a tree of order 3. Erdős and Rényi (Theorem 2a of [ErR60]) showed that

$$\lim_{n\to\infty} P(\overline{A}) = e^{-2x^3} \tag{3}$$

Since $A \subseteq \overline{C}$, we have

$$\lim_{n\to\infty} P(C) \leq e^{-2x^3}. \tag{4}$$

On the other hand, we have a lower bound for $P(C)$ if we add the probabilities of all graphs which consist only of isolated vertices and edges:

$$P(C) \geq \sum_{q=0}^{[n/2]} \binom{n}{2q} \frac{(2q)!}{q!2^q} p^q (1-p)^{\binom{n}{2}-q}. \tag{5}$$

It can be shown that this lower bound is also asymptotic to e^{-2x^2}, completing the proof. ∎

GRAPHS WITH EACH EDGE IN A TRIANGLE

Let B be the set of graphs in Ω_n which have each edge in a triangle. The determination of a sharp threshold function for this set plays a crucial role in our study of the limiting probability of the set of locally connected graphs.

We define the random variable $X(G)$ to be the number of edges of G which are *not* in triangles. Thus the expected number of these edges is

$$E(X) = \binom{n}{2} p(1 - p^2)^{n-2}. \tag{6}$$

If we take $p = (c \log n/n)^{1/2}$ with constant $c > 0$, then

$$E(X) \sim \frac{\sqrt{c}}{2} n^{3/2-c}(\log n)^{1/2}. \tag{7}$$

Then, if $c > 3/2$, $E(X) \to 0$ as $n \to \infty$. Therefore, $P(\overline{B}) = P(X \geq 1) \to 0$ or $P(B) \to 1$, i.e., almost every graph has every edge in a triangle.

On the other hand, if $0 < c \leq 3/2$, then $E(X) \to +\infty$. In this case it is easy to show that $E(X^2) \sim E(X)^2$. In fact, $E(X^2) = E(X) + 2P_2$, where P_2 is given below in (14). It follows from Chebyshev's inequality that

$$P(X = 0) \leq (E(X^2)/E(X)^2) - 1. \tag{8}$$

Since $P(B) = P(X = 0)$, we have $P(B) \to 0$ as $n \to \infty$, i.e., almost every graph has an edge not in a triangle.

Thus, we have established a rough threshold for this property and we refine the result in the next theorem.

Theorem 2. If the probability of an edge is

$$p = ((3/2 + \varepsilon_n) \log n/n)^{1/2} \tag{9}$$

where

$$\varepsilon_n = (\log \log n + \log(3/8) + 2x)/(2 \log n), \tag{10}$$

then the limiting probability that a random graph has every edge in a triangle is

$$\lim_{n\to\infty} P(B) = e^{-e^{-x}}. \tag{11}$$

Proof. First we note that with p defined by (9) and (10), we have on substitution in (6):

$$E(X) \sim e^{-x}. \tag{12}$$

It follows from the probabilistic form of the method of inclusion and exclusion that

$$P(X = 0) = \sum_{k=0}^{\binom{n}{2}} (-1)^k P_k, \tag{13}$$

where P_k is the sum, taken over all k-sets of pairs of vertices, of the probabilities of all graphs in which these k pairs of vertices are adjacent but none of these k edges are in triangles. Note that $P_1 = E(X)$, while

$$P_2 = \frac{1}{2!}\binom{n}{2,\,2,\,n-4} p^2(1-p^2)^{2(n-4)}(1 + 0\,(1))$$
$$+ \frac{1}{2} n(n-1)(n-2)p^2(1-p)[(1-p) + p(1-p^2)]^{n-3}. \tag{14}$$

In general the dominant contribution to P_k is made by k-sets of disjoint edges and we have

$$P_k \sim \frac{1}{k!}\binom{n}{2,\,2,\,\ldots,\,2,\,n-2k}[p(1-p^2)^{n-2k}]^k, \tag{15}$$

from which it follows that

$$P_k \sim (e^{-x})^k/k!. \tag{16}$$

Since $P(X = 0)$ is contained between any two consecutive partial sums of the sum on the right side of (13) and (16) holds for fixed k, we have

$$\lim_{n\to\infty} P(X = 0) = \sum_{k=0}^{\infty} (-1)^k (e^{-x})^k/k!, \tag{17}$$

which completes the proof. ∎

THE HIGH THRESHOLD

After passing the threshold of Theorem 1, we observe that local connectivity is not achieved by almost all graphs until at least the stage of connectivity is reached. To see this, one follows the evolution of the random graph as developed in [ErR60]. For example, if $pn^{3/2} \to \infty$ but $pn \to 0$, each component of the random graph is a tree (Theorem 4a of [ErR60]) and hence the graph is not locally connected.

If $pn = c$, a constant, the random graph passes through the "Double Jump" phase of the evolutionary process (see p. 52 of [ErR60]). We define the random variable $X(G)$ to be the number of vertices of degree 1 in G which are adjacent to vertices of degree ≥ 2. A graph with such an end-vertex is not locally connected. The expected value of X is

$$E(X = n(n-1)\, p\, (1-p)^{n-2}[1-(1-p)^{n-2}] \tag{18}$$

and

$$\begin{aligned} E(X^2) = E(X) + n(n-1)\{&(n-2)p^2(1-p)^{2n-7} \\ &+ (n-2)(n-3)p^3(1-p)^{2n-5} \\ &+ (n-2)(n-3)p^2(1-p)^{2n-4}[1-(1-p)^{n-4}]^2\}. \end{aligned} \tag{19}$$

It can be seen that $E(X^2) \sim E(X)^2$ for $pn = c$. Hence $P(X = 0) \to 0$, and so the random graph is sure to be locally disconnected. Furthermore, as we approach full connectivity with $p = c \log n/n$ and $0 < c \leq 1$, the same conclusion holds. Thus, in the late stage when $p = (\log n + x)/n$ and the random graph consists of a single component plus perhaps some isolated points (see Lemma on p. 292 of [ErR59]) we can be certain that the big component has end-vertices, i.e., the giant is fuzzy.

If $p = c \log n/n$ with constant $c > 1$, the random graph is connected but we have seen earlier in Theorem 2 that it is sure to have an edge which does not belong to a triangle. Hence it is locally disconnected and we cannot hope for local connectivity until we reach the threshold of formulas (9) and (10). In this case we have the following precise result.

Theorem 3. If the probability of an edge is given by formulas (9) and (10), the limiting probability that a random graph is locally connected is

$$\lim_{n\to\infty} P(C) = e^{-e^{-x}}. \tag{20}$$

Proof. Let A be the set of all graphs in Ω_n such that the neighborhood of each vertex consists of one component plus perhaps isolated vertices. Let B be those graphs such that each neighborhood has no isolated vertices. Note that B is exactly as before, i.e., every edge is in a triangle. Finally C_0 consists of those graphs which are both connected and locally connected. When the probability of an edge is given by (9) and (10), almost all graphs are connected. Hence $P(C)$ and $P(C_0)$ are equal in the limit. But for each n

$$C_0 = A \cap B. \tag{21}$$

Therefore $P(B) \geqslant P(C_0)$ and since $B \cap \overline{C_0} \subseteq \overline{A}$, the proof can be completed by showing that $P(\overline{A}) \to 0$, i.e., almost all graphs are type A.

Define $X(G)$ to be the number of vertices in G whose neighborhoods do *not* consist of one component and perhaps some isolated vertices. We can estimate $E(X)$ by using the following observation on p. 59 of [ErR60]. If $\varepsilon > 0$ and $pn/\log n \to \infty$, then almost all graphs satisfy

$$(1 - \varepsilon)pn < \deg v < (1 + \varepsilon)pn \tag{22}$$

for all vertices v.

Now we estimate the expected number of bad vertices to be

$$E(X) \sim n \sum \binom{n-1}{m} p^m (1-p)^{n-1-m} P(m), \tag{23}$$

where the sum is over all m in the interval $(1 - \varepsilon)pn < m < (1 + \varepsilon)pn$ and $P(m)$ is the probability that an m-set of vertices does not consist of one component plus perhaps isolated vertices.

For any m in the interval above,

$$p > \frac{\log m}{m} \frac{(1-\varepsilon)p^2 n}{\log((1+\varepsilon)pn)}, \tag{24}$$

and on substitution of p from formula (9) in the right side of (24), we find that for any $\delta > 0$ there is $\varepsilon > 0$ so for n sufficiently large

$$p > (3 - \delta)\log m/m. \tag{25}$$

The proof of the lemma on p. 292 of [ErR59] can be used together with the lowerbound (25) for p to obtain an upper bound for $P(m)$. We find that

$$\begin{aligned} P(m) = \mathcal{O}(&\alpha^{a_1\sqrt{m}} + (a_2/m^{2-\delta})^{\log\log m} \\ &+ m^{2\delta-5}\log m \cdot e^{(\log\log m)^2/2}), \end{aligned} \tag{26}$$

where $0 < \alpha < 1$ and a_1 and a_2 are positive constants. Then it can be shown that $nP(m) \to 0$, completing the proof. ∎

We conclude by observing that when the probability of an edge is beyond the high threshold of Theorem 3, almost every graph is locally connected. To see this let $p = (c \log n/n)^{1/2}$ with $c > 3/2$. Now we can show that the expected number of vertices whose neighborhoods are disconnected tends to zero. As in the proof of Theorem 3 we obtain a lower bound for p using (24), but with $c > 3/2$ we find for n sufficiently large

$$p > 2c(1 - \varepsilon)\frac{\log m}{m}. \tag{27}$$

Therefore by choosing ε small we can be assured that $2c(1 - \varepsilon) = 3 + \delta$. The probability that an m-set is disconnected is $\mathcal{O}(m^{1-(3+\delta)})$ (see [ErR59]). Therefore the expected number of disconnected neighborhoods is $n\mathcal{O}(m^{-2-\delta})$ and this bound tends to zero for m in the range of (22).

References

[BlH79] A. Blass and F. Harary, Properties of almost all graphs and complexes, *J. Graph Theory* 3 (1979) 225–240.

[Bo79] B. Bollobás, *Graph Theory*, Springer, New York (1979).

[ErR59] P. Erdös and A. Rényi, On random graphs I, *Publ. Math. Debrecen* 6 (1959) 290–297.

[ErR60] P. Erdös and A. Rényi, On the evolution of random graphs, *Magyar. Tud Akad. Mat. Kutato Int. Kozl.* 5 (1960). 17–61.