On the connectivity and diameter of small-world networks

Ayalvadi Ganesh
Microsoft Research
7 J J Thomson Avenue
Cambridge CB3 0FB, UK

Feng Xue
Coordinated Science Laboratory
University of Illinois
1308 W. Main Street
Urbana, IL 61801, USA

January 2005

Technical Report
MSR-TR-2005-14

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

Abstract

We consider two different models of small-world graphs on nodes whose locations are modelled by a stochastic point process. In the first model, each node is connected to a fixed number of its nearest neighbours, while in the second, it is connected to all nodes located within some fixed distance. In both models, nodes are additionally connected via shortcuts to other nodes chosen uniformly at random. We obtain sufficient conditions for connectivity in the first model, and necessary conditions in the second. Thereby, we show that connectivity is achieved at a smaller value of total degree (nearest neighbours + shortcuts) in the first model. We also obtain bounds on the diameter of the graph in this model.

1 Introduction

A classical random graph model introduced by Erdős and Rényi [4] consists of n nodes, with the edge between each pair of nodes being present with probability p_n, independent of all other edges. This model, which is known as the Bernoulli random graph model, has been extensively studied, and many of its properties are well understood. For instance, Erdős and Rényi showed that this random graph model exhibits a sharp threshold for connectivity at $p_n = \log n/n$. Precisely, if $np_n - \log n \to c$ as $n \to \infty$, for some constant c, then the probability that the graph is connected goes to $\exp(-\exp(-c))$. Note that $(n-1)p_n$ is the expected node degree, so the result says that there is a sharp threshold for connectivity at an expected node degree of $\log n$. A number of variants of the above model, such as random regular graphs, have also been studied extensively. A feature common to these models is that the nodes are exchangeable.

Recently, there has been considerable interest in a different class of models, namely spatial or geometric random graphs; see, for example, [8, 9]. Here, the nodes are associated with co-ordinates in a Euclidean space, and the probability of an edge between a pair of nodes is typically some function of the distance between them (or, more generally, of their spatial co-ordinates). The node positions are also the realisation of some random process. One example is obtained by placing n nodes uniformly on the unit square and putting an edge between any pair of nodes if the distance between them is smaller than a threshold r_n. It was shown by Gupta and Kumar [5] that the probability of connectivity in this model goes to 1 if $\pi n r_n^2 - \log n \to +\infty$, and is smaller than 1 otherwise, and by Penrose [9] that the probability of connectivity goes to zero if $\pi n r_n^2 - \log n \to -\infty$. Since $\pi n r_n^2$ is the expected degree of each node (except near the edges of the square), we see that there is a threshold for connectivity at a mean degree of $\log n$, which is the same as in the Bernoulli random graph. A somewhat different model was studied in Xue and Kumar [11]. Here, each node is connected to its m_n nearest neighbours; more precisely, the edge between u and v is present if either u is one of the m_n nodes closest to v or v is one of the m_n nodes closest to u. The authors show for this model that there are finite positive constants c_1 and c_2 such that the probability of connectivity goes to 1 if $m_n > c_2 \log n$ and to 0 if $m_n < c_1 \log n$. They also conjecture that there is a threshold for connectivity at $m_n = c \log n$ for some c. We remark that in the above models, the same results hold if we consider a Poisson point process of intensity n instead of n points uniformly distributed on the unit square. Indeed, it is easier to prove the results in the Poisson case and then show that the models are equivalent as far as connectivity is concerned.

One of the motivations for interest in spatial random graphs is their applicability to wireless communication networks [5, 11]. Spatial random graphs on high-dimensional spaces might offer good models for social networks, which are poorly described by Bernoulli random graphs. Another class of models that has attracted attention in the latter context are so-called "small world networks". One commonly used way to model such networks is to consider nodes as located at the points of a (finite or infinite) d-dimensional lattice, and to augment the lattice with shortcuts, which are additional edges between pairs of nodes. The shorcut between a pair of nodes is present with a probability that is typically some function of the distance between them. Since the lattice is already connected, interest in these models has focused on how the diameter is reduced by the presence of shortcuts (see, for example, [1, 3, 10]), and also on whether efficient decentralised routing is possible [6].

In this paper, we consider two variants of the above small world model. We model node locations by a stochastic point process, e.g., the Poisson process on a square. Nodes are connected by nearest neighbour links, either to a fixed number of nodes closest to them, or to all nodes within a fixed distance. In addition, nodes are joined by shortcuts to other nodes chosen uniformly at random. We are interested in how connectivity depends on the number of nearest neighbours and the number of shortcuts. In the next section, we address this question after providing a precise definition of the models. We also obtain bounds on the diameter of the small-world network in the connected regime.

2 Main Results

We consider two different models of a small-world network, denoted Model A and Model B. In each case, we consider a sequence of random networks indexed by a parameter $n \in \mathbb{N}$, which we call the size of the network. We say that a property Q holds with high probability if the probability that a random network of size n possesses the property Q goes to 1 as n tends to infinity. In all cases, we consider undirected graphs.

Model A: There are n nodes, and each node is chooses m_n other nodes to connect to, called its nearest neighbours. In addition, a shortcut is present between each pair of nodes with probability p_n, independently of all other edges. (If two nodes are connected by both a nearest neighbour edge and a shortcut, we replace the multiple edge by a simple one.)

Note that the "nearest neighbour" relation need not be symmetric. An edge is present between nodes x and y if either y is one of the m_n nearest neighbours of x or x is one of the m_n nearest neighbours of y or there is a shortcut between them. The model is parametrised by the sequences m_n and p_n, and shortcuts are the only source of randomness in this model. The terminology of "nearest neighbour" may be misleading: as far as our results below are concerned, it only matters that each node connects to m_n other nodes, chosen arbitrarily. However, we have chosen to use this term for concreteness, and because it was motivated by applications.

Example: Suppose that the nodes are located uniformly at random on the unit square. In this case, Model A incorporates elements of both Bernoulli random graphs and the Xue-Kumar model, and our results, stated in Theorem 1 below, apply to *every* realisation of the node locations.

Model B: We consider the square $[-\sqrt{n}/2, \sqrt{n}/2]^2$ of area n centred at the origin. Nodes are located at the points of a Poisson process of unit intensity conditioned to have a point at the origin (i.e., we work with the Palm version of the Poisson process). Each node is connected to all nodes within a radius r_n and, in addition, shortcuts are present between each pair of nodes with probability p_n, independent of all other edges.

Model B is parametrised by the sequences r_n and p_n. It combines elements of Bernoulli and spatial random graphs. Observe that the shortcut distribution is the same in Models A and B. The main difference between the models is thus that the number of nearest neighbours is random in Model B but deterministically bounded below in Model A. We shall see that this greatly improves connectivity in Model A.

We consider a sequence of random graphs indexed by n. We denote by C_n the event that the n^{th} random graph is connected, We denote by D_n the diameter of the graph, namely the maximum over all node pairs of length of the shortest path between them, in terms of number of edges. We take $D_n = \infty$ if the graph is not connected.

Theorem 1 *Suppose that the sequences m_n and p_n are such that*

$$\frac{m_n}{n} \to 0 \text{ as } n \to \infty, \text{ and } (m_n+1)np_n > 2(1+\delta)\log\frac{n}{m_n+1} \tag{1}$$

for some $\delta > 0$ and all n sufficiently large. Then, for the random graph described in Model A above with parameters m_n and p_n, we have

$$\lim_{n\to\infty} P(C_n) = 1, \quad \lim_{n\to\infty} P\Big(D_n \le 7\log\frac{n}{m_n+1}\Big) = 1, \tag{2}$$

Conversely, if

$$\frac{m_n}{n} \to 0 \text{ as } n \to \infty, \text{ and } (m_n+1)np_n < (1-\delta)\Big(\frac{m_n+1}{m_n+2}\Big)^2 \log\frac{n}{m_n+1} \tag{3}$$

for some $\delta > 0$ *and infinitely many* n*, then there is a sequence of node locations such that* $\liminf_{n\to\infty} P(C_n) = 0$.

Remarks:

1. If $m_n = 0$, then Model A reduces to the classical Bernoulli model of Erdős and Rényi. In this case, our upper bound on the required node degree is conservative by a factor of 2, and the lower bound by a factor of 4.

2. If $m_n \to \infty$ as $n \to \infty$, then the upper and lower bounds differ by essentially a factor of 2.

3. Note that m_n is the number of neighbours a node has in terms of the spatial graph, and $(n-1)p_n$ the number of neighbours it has via shortcuts. Thus, the conditions of the theorem say that the product of these quantities must be roughly $\log n$ in order to ensure connectivity. For example, it suffices if $m_n = (1+\delta)\sqrt{\log n}$ and $np_n = (1+\delta)\sqrt{\log n}$.

Theorem 2 *Suppose that the sequences* r_n *and* p_n *are such that*

$$\pi r_n^2 + np_n = \log n - \omega_n, \qquad \textit{and} \lim_{n\to\infty} \omega_n = \infty. \tag{4}$$

Then, with probability going to 1 as n *tends to* ∞*, the random graph generated by Model B with parameters* r_n *and* p_n *fails to be connected.*

Remarks:

1. Observe that $\pi r_n^2 + np_n$ is the mean node degree, so the theorem says that the graph is disconnected if the mean node degree is much smaller than $\log n$, as is the case for both Bernoulli and spatial random graphs. Thus, in this model, there is no synergy between nearest neighbour and shortcut links, at least as far as connectivity as concerned.

2. We recover the necessary condition for connectivity in spatial random graphs by setting $p_n = 0$, and in Bernoulli random graphs by setting $r_n = 0$.

3 Proofs

Proof of Theorem 1: Observe that under Model A, each node belongs to a connected component with at least $m_n + 1$ nodes, since it is connected to m_n nearest neighbours. The intuition behind the proof is that, for the graph to fail to be connected, there must be an isolated component of at least this size. The proof will proceed by reducing the small world graph to a Bernoulli graph between *clusters* of size approximately m_n.

Given a graph G, we divide the nodes into disjoint groups as follows:

1. For each node x, let C_x consist of x and its m_n nearest neighbours. We call C_x the disc centred at x. Initially, the discs C_x are coloured black.

2. Considering the n black discs in (any) sequence, colour each disc red if it does not intersect a disc already coloured red. If a disc C_x gets coloured red, we shall refer to it as the red disc centred at x.

3. Pick a red disc, say the one centred at x. Consider the nodes in all black discs overlapping it, if any. Group these nodes into disjoint sets of size m_n+1 and a residual set with m_n or fewer nodes. Call each group of size m_n+1 a green disc, and absorb the m_n or fewer residual nodes into the red disc. With some abuse of terminology, we shall refer to the (possibly) enlarged red disc and related green discs as all being centred at x.

The procedure terminates with nodes being grouped into disjoint discs A_k, $k=1,2,\cdots,K_n$, each of which is coloured either red or green. All green discs are of size m_n+1 and all red discs are of size between m_n+1 and $2m_n+1$.

Observe that any two nodes in the red disc centred at x either come from the black disc that was centred at x or they come from overlapping black discs that were centred at x and y. In the former case, the distance between these nodes in G is at most 2; in the latter case, it is at most 4 (since, if the discs centred at x and y intersect, there is a node z which is a neighbour of both x and y). Either way, all nodes in the same red disc belong to the same connected component in G, considering only nearest-neighbour edges.

Likewise, any two nodes u and v belonging to the same green disc centred at x come from black discs centred at y and z (possibly, $y=z$) which overlapped the black disc centred at x. Hence, these nodes belong to the same connected component in G and are at most distance 6 apart (there is a path $u \to y \to a \to x \to b \to z \to v$ for some nodes a and b).

In order to show that G is connected, it now suffices to show that the red and green discs form a connected graph when considering only shortcut edges. We shall do this by defining a Bernoulli graph $\tilde{G}$, as follows. Recall that each red disc has between m_n+1 and $2m_n+1$ nodes. We first construct a subgraph G_1 of G by deleting all but m_n+1 nodes in each red disc. (It doesn't matter which m_n+1 nodes are retained, so long as the choice is independent of the presence of edges in G. We can think of the edges of G_1 as being realised after the nodes are chosen.) In G_1, each disc, red or green, has exactly m_n+1 nodes, and if G_1 is connected, then clearly so is G. Next, construct $\tilde{G}$ by replacing each disc A_k in G_1 by a single node k, and putting an edge between nodes j and k if there is at least one shortcut edge between a node in A_j and a node in A_k, in G_1. Clearly, G_1 is connected if $\tilde{G}$ is. But $\tilde{G}$ is a Bernoulli random graph on K_n nodes, with edge probability

$$\begin{aligned} \tilde{p}_n &= 1-(1-p_n)^{(m_n+1)^2} \\ &\geq 1-e^{-(m_n+1)^2 p_n} \\ &\geq 1-\exp\Big(-2(1+\delta)\frac{m_n+1}{n}\log\frac{n}{m_n+1}\Big), \end{aligned} \tag{5}$$

where the last inequality follows if (1) is assumed to hold. Morever, K_n, the number of nodes in $\tilde{G}$, lies between $n/(2m_n+1)$ and $n/(m_n+1)$. Hence, it follows from (5) that

$$\tilde{p}_n \geq 1-e^{-(1+\delta)\log K_n/K_n}. \tag{6}$$

Now, $K_n \to \infty$ as $n \to \infty$ by the assumption that $m_n/n \to 0$. Hence, we obtain from (6) that

$$\tilde{p}_n \geq (1+\delta')\frac{\log K_n}{K_n}, \tag{7}$$

for any $\delta' \in (0,\delta)$ and all n sufficiently large. But $\tilde{p}_n$ is the edge probability in $\tilde{G}$, which is a Bernoulli random graph on K_n nodes. Hence, by the results of Erdős and Rényi [4], $P(C_n) \to 1$ as $n \to \infty$. This establishes the first claim in (2).

Moreover, it follows from [2, Theorem 10.17] that the Bernoulli random graph $\tilde{G}$ on K_n vertices satisfies

$$P\Big(\text{ diameter}(\tilde{G}) \ \leq \frac{\log K_n + 6}{\log\log K_n} + 4\Big) \to 1,$$

as $K_n \to \infty$. Since $K_n \leq \frac{n}{m_n+1}$, and $K_n \to \infty$ as $n \to \infty$, it follows that

$$P\Big(\text{ diameter}(\tilde{G}) \ \leq \log\frac{n}{m_n+1}\Big) \to 1 \text{ as } n \to \infty.$$

Recall that each node in $\tilde{G}$ corresponds to a connected cluster in G, and the diameter of this cluster is at most 6. Thus, the second claim in (2) follows.

Conversely, consider a sequence m_n, p_n satisfying (3) for some $\delta > 0$ and all n sufficiently large. We now consider the following deterministic sequence of node configurations. The configuration on n nodes consists of clusters of size m_n+1 or m_n+2. The nodes within each cluster are within distance ϵ_n of each other, and any two nodes in distinct clusters are more than ϵ_n apart, for some positive constant ϵ_n. It is clear that such an ϵ_n and node configuration can be found.

Denote the clusters by $A_1, A_2, \ldots, A_{K_n}$. There are clearly no nearest neighbour edges between clusters, only shortcuts. Let G_1 denote the graph on K_n nodes obtained by replacing each cluster A_j by a single node j, and putting an edge between nodes j and k only if there is a shortcut in G between clusters A_j and A_k. Now, conditional on the cluster sizes, there is a shortcut (at least one) between clusters A_j and A_k with probability

$$1 - (1 - p_n)^{|A_j| \cdot |A_k|} \leq \tilde{p}_n := 1 - (1 - p_n)^{(m_n+2)^2}. \tag{8}$$

The presence of shortcuts between clusters are not independent events because the existence of one shortcut biases the conditional distribution of the size of the cluster, and thereby the probability of other shortcuts from that cluster. Hence, G_1 is not a Bernoulli random graph. However, this problem is easily circumvented, as follows.

First, we augment each cluster of size m_n+1 in G by adding a pseudonode which is within distance ϵ_n of all nodes in this cluster. Shortcuts are present between pseudonodes and other nodes with the same probability p_n as for ordinary nodes, independent of the presence of other shortcuts. Call the augmented graph $\tilde{G}$. Now construct $\tilde{G}_1$ from $\tilde{G}$ analogous to how G_1 was constructed from G: replace each cluster A_j by a single node j, and put an edge between j and k in $\tilde{G}_1$ only if there is a shortcut between the (augmented) clusters A_j and A_k in $\tilde{G}$. It is clear from this construction that $\tilde{G}_1$ is a Bernoulli random graph on K_n nodes with edge probability $\tilde{p}_n$ given by (8). Moreover, G_1 is a subgraph of $\tilde{G}_1$. Now G is connected only if G_1 is, which in turn requires that $\tilde{G}_1$ be connected. We shall now use the result of Erdős and Rényi [4] to show that, with high probability, $\tilde{G}$ fails to be connected.

Observe that $p_n \to 0$ as $n \to \infty$ by the assumption in (3) that $nm_np_n < \log n$ for all n sufficiently large. Hence, for any $\epsilon > 0$, we have for all n sufficiently large that

$$\tilde{p}_n \leq 1 - e^{-(1+\epsilon)(m_n+2)^2 p_n} \leq (1+\epsilon)(m_n+2)^2 p_n.$$

Since $K_n \in [\frac{n}{m_n+2}, \frac{n}{m_n+1}]$, it follows from (3) that

$$K_n\tilde{p}_n \leq (1+\epsilon)(1-\delta)\log\frac{n}{m_n+1}.$$

Since $\epsilon > 0$ can be chosen arbitrarily small, we can find $\delta' \in (0, \delta)$ such that

$$K_n\tilde{p}_n \leq (1-\delta')\log K_n,$$

for all K_n sufficiently large. Morever, $K_n \to \infty$ as $n \to \infty$ by the assumption that $m_n/n \to 0$. Hence, using the results in [4] on the connectivity of Bernoulli random graphs, we obtain that $P(\tilde{G}_1 \text{ is connected}) \to 0$ as $n \to \infty$. But G is connected only if $\tilde{G}_1$ is. Therefore, $P(C_n) \to 0$ as $n \to \infty$. This completes the proof of the theorem.

Proof of Theorem 2: The graph G is disconnected if it has at least one isolated node, namely, a node that has neither nearest neighbour nor shortcut links with any other node. Let U denote the number of isolated nodes. The key ideas of the proof are (i) to bound the probability that the graph is connected by $P(U = 0)$, and (ii) to obtain a bound on

$P(U=0)$ by calculating EU and showing that U is approximately Poisson with mean EW. Details are provided below.

We neglect edge effects in the following. Denote by $\mathcal{C}_n^1$ and $\mathcal{C}_n^2$ respectively the circles of radius r_n and $2r_n$ centred at the origin. First, we compute the number of isolated nodes conditional on there being N nodes in the square $[-\sqrt{n}/2, \sqrt{n}/2]^2$, excluding the node at the origin. If $N=0$, the graph is trivially connected, and we say that there are no isolated nodes, by definition. If $N \neq 0$, we have

$$P(\text{node at origin is isolated}|N) = \Big(1 - \frac{\pi r_n^2}{n}\Big)^N (1-p_n)^N, \tag{9}$$

whereas, for any other node i,

$$P(i \text{ isolated}|N) = P(i \notin \mathcal{C}_n^1) P(i \text{ isolated}|i \notin \mathcal{C}_n^1, N) = \Big(1 - \frac{\pi r_n^2}{n}\Big)^N (1-p_n)^N. \tag{10}$$

Hence, the expected number of isolated nodes is given by

$$E[U|N] = (N+1)\Big[(1 - \frac{\pi r_n^2}{n})(1-p_n)\Big]^N - \mathbf{1}(N=0). \tag{11}$$

Next, we take expectations with respect to N, which is a Poisson random variable with mean n. Recall that

$$X \sim \text{Poisson}(\nu) \ \Rightarrow \ E[z^X] = e^{-\nu(1-z)}, \quad E[Xz^{X-1}] = \nu e^{-\nu(1-z)}. \tag{12}$$

Hence, (11) implies that

$$EU = \Big(n(1 - \frac{\pi r_n^2}{n})(1-p_n) + 1\Big) \exp\Big[-n\Big(1 - (1 - \frac{\pi r_n^2}{n})(1-p_n)\Big)\Big] - e^{-n}.$$

Defining $\lambda_n = EU$, we get

$$\lambda_n \geq (n - \pi r_n^2 - np_n) e^{-(\pi r_n^2 + np_n)}. \tag{13}$$

Let Z be a Poisson random variable with mean λ_n. We shall now use the Chen-Stein method to show that the distribution of U is close to that of Z. Let $1+V$ have the distribution of the number of isolated points conditional on the point at 0 being isolated. We shall construct both random variables U and V on the same probability space, and show that $E[U-V]$ is small. Then, we can use [7, Theorem 24.3] to deduce that the total variation distance between U and Z is small. We detail the arguments below. For notational convenience, we have not made explicit the dependence of U and V on n.

Recall that $\mathcal{C}_n^1$ and $\mathcal{C}_n^2$ respectively denote the circles of radius r_n and $2r_n$ centred at the origin. On the event that the point at the origin is isolated, there can clearly be no other points in $\mathcal{C}_n^1$. Moreover, there can be no shortcuts between the point at the origin and any other. Since the point process is Poisson, the law of the process outside $\mathcal{C}_n^1$ remains unchanged. Since the shortcuts are iid Bernoulli random variables, the presence of shortcuts between other pairs of nodes is unaffected as well. Hence, we can construct the random variables U and V on the same probability space, namely the space of graphs generated by Model B with parameters r_n and p_n as follows. Given a graph G in the sample space, we define U to be the number of isolated nodes in G and $1+V$ to be the number of isolated nodes when all nodes inside $\mathcal{C}_n^1$, except for the one at the origin, are deleted, and all shortcuts from the origin to other nodes are deleted as well. Note that U cannot have any isolated nodes inside $\mathcal{C}_n^1$, and any node isolated in the definition of U is also isolated in the definition of V. Hence, $U \leq V$, and $V-U$ is the cardinality of the set of nodes all of whose neighbours (nearest and shortcut) are within the circle of radius r_n centred at the origin. Call this set $A(G)$, noting that it depends on the graph G.

The set $A(G)$ can be expressed as $A_1(G) \cup A_2(G)$ where $A_1(G)$ consists of nodes lying in the annulus $\mathcal{C}_n^2 \setminus \mathcal{C}_n^1$ and having no nearest neighbours or shortcut links outside $\mathcal{C}_n^1$. Nodes in $A_1(G)$ should have at least one link to a node inside $\mathcal{C}_n^1$; ignoring this conditions yields an upper bound on $|A_1(G)|$. The set $A_2(G)$ consists of nodes outside $\mathcal{C}_n^2$ which have no neighbours within a distance r_n of themselves, and which have a shortcut link to at least one node inside $\mathcal{C}_n^1$ but to none outside this set. Note that nodes outside $\mathcal{C}_n^2$ can't have a nearest neighbour link to a node inside $\mathcal{C}_n^1$.

Let N_1 denote the number of nodes of G lying within the annulus $\mathcal{C}_n^2 \setminus \mathcal{C}_n^1$. Then N_1 is a Poisson random variable with mean $3\pi r_n^2$. Likewise, the number of nodes outside $\mathcal{C}_n^2$, denoted N_2, is Poisson with mean $n - 4\pi r_n^2$, and indepedent of N_1. Now, if $N_1 = 0$, then $A_1(G) = \emptyset$. Suppose $N_1 \geq 1$, and let x be one of the nodes in $\mathcal{C}_n^2 \setminus \mathcal{C}_n^1$. Then,

$$P(x \in A_1(G)|N_1, N_2) \leq \Bigl(1 - \frac{r_n^2 \cos^{-1}(1/4)}{3\pi r_n^2}\Bigr)^{N_1-1}(1-p_n)^{N_1+N_2-1}.$$

The first term in the product above is a bound on the probability that none of the other $N_1 - 1$ points in $\mathcal{C}_n^2 \setminus \mathcal{C}_n^1$ lies in the intersection of the circle of radius r_n centred at x and $\mathcal{C}_n^2 \setminus \mathcal{C}_n^1$; the intersection is smallest when x is on the boundary of $\mathcal{C}_n^2$, which yields the bound. The second term is the probability that x has no shortcut links to points outside $\mathcal{C}_n^1$. Hence,

$$\begin{aligned} E\Bigl[|A_1(G)| \Bigm| N_1, N_2\Bigr] &\leq N_1(c(1-p_n))^{N_1-1}(1-p_n)^{N_2}, \\ \text{where } c &= 1 - \frac{\cos^{-1}(1/4)}{3\pi} \leq \frac{8}{9}. \end{aligned} \tag{14}$$

Now, using the independence of N_1 and N_2, we obtain from (12) and (14) that

$$\begin{aligned} E|A_1(G)| &\leq E\Bigl[N_1\Bigl(\frac{8}{9}(1-p_n)\Bigr)^{N_1-1}\Bigr]E[(1-p_n)^{N_2}] \\ &= 3\pi r_n^2 \exp\Bigl(-3\pi r_n^2 \frac{1+8p_n}{9}\Bigr)\exp[-(n-4\pi r_n^2)p_n]. \end{aligned}$$

Now, by (4), $p_n r_n^2 \leq \log^2 n/n \to 0$ as $n \to \infty$. Hence, for any $\epsilon > 0$, we have

$$E|A_1(G)| \leq (1+\epsilon)3\pi r_n^2 \exp\Big(\frac{2\pi r_n^2}{3}\Big)\exp(-\pi r_n^2 - np_n),$$

for all n sufficiently large. Using (13), we can rewrite the above as

$$E|A_1(G)| \leq 3(1+\epsilon)\frac{\lambda_n}{n-\pi r_n^2 - np_n}\pi r_n^2 \exp\Big(\frac{2\pi r_n^2}{3}\Big).$$

But, by (4), $\pi r_n^2 \leq \pi r_n^2 + np_n \leq \log n$. Hence,

$$E|A_1(G)| \leq \lambda_n \frac{4\log n}{n^{1/3}}, \tag{15}$$

for all n sufficiently large.

Next, let N_0 denote the number of nodes of G inside $\mathcal{C}_n^1$, excluding the node at the origin. Note that N_0 is Poisson with mean πr_n^2, and that N_0, N_1 and N_2 are mutually independent. For a node $x \notin \mathcal{C}_n^2$, we have

$$P(x \in A_2(G)|N_0, N_1+N_2) = \Big(1 - \frac{\pi r_n^2}{n-\pi r_n^2}\Big)^{N_1+N_2-1}[1-(1-p_n)^{N_0+1}][(1-p_n)^{N_1+N_2-1}].$$

The first term in the product above is the probability that x has no neighbours within distance r_n of itself, the second term is the probability that it has at least one shortcut link to a point inside $\mathcal{C}_n^1$ (including the point at the origin), and the last term is the probability that it has no shortcut link to a point outside $\mathcal{C}_n^1$. Consequently, by the independence of N_0 and N_1+N_2, we have

$$\begin{aligned}
E|A_2(G)| &= E\Big[N_2\Big((1-\frac{\pi r_n^2}{n-\pi r_n^2})(1-p_n)\Big)^{N_1+N_2-1}\Big]E[1-(1-p_n)^{N_0+1}] \\
&\leq E\Big[(N_1+N_2)\Big((1-\frac{\pi r_n^2}{n-\pi r_n^2})(1-p_n)\Big)^{N_1+N_2-1}\Big]E[1-(1-p_n)^{N_0+1}] \\
&= (n-\pi r_n^2)\exp\Big[-(n-\pi r_n^2)\Big(1-(1-p_n)(1-\frac{\pi r_n^2}{n-\pi r_n^2})\Big)\Big]\Big[1-(1-p_n)e^{-\pi r_n^2 p_n}\Big] \\
&\leq n\exp\Big[-(n-\pi r_n^2)+(1-p_n)(n-2\pi r_n^2)\Big]\Big[1-(1-p_n)(1-\pi r_n^2 p_n)\Big] \\
&\leq np_n(1+\pi r_n^2)e^{-\pi r_n^2}e^{-(n-2\pi r_n^2)p_n}.
\end{aligned}$$

Recalling that $p_n r_n^2 \to 0$ as $n \to \infty$, we obtain for any $\epsilon > 0$ and all n sufficiently large, that

$$E|A_2(G)| \leq (1+\epsilon)np_n(1+\pi r_n^2)e^{-\pi r_n^2 - np_n}.$$

But, by (4), $np_n \le \log n$ and $\pi r_n^2 \le \log n$. Hence, by (13),

$$E|A_2(G)| \le \lambda_n \frac{2\log^2 n}{n}, \tag{16}$$

for all n sufficiently large.

Combining (15) and (16), we obtain for all n sufficiently large that,

$$E|V-U| = E|A_1(G)\cup A_2(G)| \le \lambda_n\Big(\frac{4\log n}{n^{1/3}} + \frac{2\log^2 n}{n}\Big). \tag{17}$$

Hence, by the Stein-Chen method, (see, for example, Section 24 and Theorem 24.3 in [7]), it follows from (17) that

$$d_{TV}(U,Z) \le 2(1\wedge \lambda_n^{-1})\lambda_n\Big(\frac{4\log n}{n^{1/3}} + \frac{2\log^2 n}{n}\Big), \tag{18}$$

where Z is a Poisson random variable with mean $\lambda_n = EU$. In particular,

$$\begin{aligned} P(U=0) &\le P(Z=0) + d_{TV}(U,Z) \\ &\le e^{-\lambda_n} + 2(1\wedge \lambda_n^{-1})\lambda_n\Big(\frac{4\log n}{n^{1/3}} + \frac{2\log^2 n}{n}\Big). \end{aligned} \tag{19}$$

But, by (4) and (13),

$$\lambda_n \ge \frac{n-\log n}{n} e^{\omega n} \to \infty \text{ as } n\to\infty\ .$$

Hence, by (19), $P(U=0)$ goes to zero as n tends to infinity. In other words, with probability going to 1, there is at least one isolated vertex, i.e., the graph G is disconnected. This completes the proof of the theorem.

4 Conclusions and open problems

We obtained a sufficient condition for connectivity in Model A and showed that this condition was necessary for a worst-case node configuration. It would be interesting to know whether the condition is tight for some random node configurations, such as the Poisson point process on a square. We have obtained necessary conditions for connectivity in Model B. While it is tempting to conjecture that there is a threshold for connectivity in this model at a mean degree of $\log n$, we do not have a proof of this result.

References

[1] A. D. Barbour and G. Reinert, "Small worlds", *Random Structures and Algorithms*, 19: 54–74, 2001.

[2] Bela Bollobas, *Random graphs*, Cambridge University Press, 2001.

[3] Don Coppersmith, David Gamarnik and Maxim Sviridenko, "The diameter of a long-range percolation graph", *Random Structures and Algorithms*, 21: 1–13, 2002.

[4] P. Erdős and A. Rényi, "On the evolution of random graphs", *Mat Kutato Int. Közl*, 5: 17–60, 1960.

[5] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks", in *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming*, W. H. McEneaney, G. Yin and Q. Zhang (eds.), Birkhäuser, Boston, 1998.

[6] Jon Kleinberg, "The small world phenomenon: an algorithmic perspective", in *Proc. 32nd ACM Symp. Theory Comp. (STOC)*, 2000.

[7] Torgny Lindvall, *Lectures on the Coupling Method*, Dover Publications, New York, 1992.

[8] Ronald Meester and Rahul Roy, *Continuum Percolation*, Cambridge University Press, Cambridge, 1996.

[9] Mathew Penrose, *Random Geometric Graphs*, Oxford University Press, Oxford, 2003.

[10] D. J. Watts, *Small Worlds*, Princeton University Press, 1999.

[11] Feng Xue and P. R. Kumar, "The number of neighbours needed for connectivity of wireless networks", *Wireless Networks*, 10: 169–181, 2004.