On the connectivity and diameter of small-world networks

Ayalvadi Ganesh[*]   Feng Xue[†]

October 15, 2007

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.

Keywords: Random graphs, small-world models, connectivity.
Subject classification: 05C80

1 Introduction

A classical random graph model introduced by Erdős and Rényi [5] 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

[*]Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK (e-mail: a.ganesh@bristol.ac.uk). This work was carried out while the author was with Microsoft Research, Cambridge, UK.

[†]Communications Technology Laboratory, Intel Corporation, Santa Clara, CA 95052 USA (e-mail: feng.xuel@intel.com)

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, [9, 11]. 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 Penrose [10] that the probability of connectivity goes to zero if $\pi n r_n^2 - \log n \to -\infty$, and goes to 1 if $\pi n r_n^2 - \log n \to +\infty$. Similar results were also obtained by Gupta and Kumar [6] in the context of a model of wireless networks.

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 [13]. 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 the probability of connectivity goes to 1 if $m_n > 5.1774 \log n$ and to 0 if $m_n < 0.074 \log n$. These results were improved by Balister et al. [1] who showed that $m_n \geq 0.3043 \log n$ is necessary and $m_n > 0.5139 \log n$ is sufficient for connectivity (with probability going to 1 as $n \to \infty$). It is not known whether there is a threshold for connectivity at $c \log n$ for some constant 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.

One of the motivations for interest in spatial random graphs is their applicability to wireless communication networks [6, 13]. 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 shortcut 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, [2, 4, 12]), and also on whether efficient decentralised routing is possible [7].

In this paper, we consider two variants of the above small world model. We model node locations by a stochastic point process, e.g., iid uniform 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 graph 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 (whp) 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 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: There are n nodes, located uniformly at random on the torus obtained by identifying the opposite sides of the square $[-\sqrt{n}/2, \sqrt{n}/2]^2$ of area n centred at the origin. 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.

The model 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 nth random graph is connected, We denote by D_n the diameter of the graph, namely the maximum over all node pairs of the 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

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

Conversely, if

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

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 $m_n = (1 + \delta)\sqrt{2 \log n}$ and $np_n = (1 + \delta)\sqrt{2 \log n}$.

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

$$\pi r_n^2 + np_n = \log n + c_n, \quad and \lim_{n\to\infty} c_n = c. \tag{4}$$

Then, the number of isolated nodes in the random graph generated by Model B with parameters r_n and p_n converges in distribution to a Poisson random variable with mean e^{-c}. Moreover, if $\lim_{n\to\infty} c_n = -\infty$, then the random graph generated by Model B is disconnected with high probability.

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. The second claim of the theorem, that the random graph is disconnected whp if $c_n \to -\infty$, can be proved directly using the second moment method. Specifically, if W denotes the number of isolated nodes in this random graph, we can use the inequality $P(W = 0) \leq \mathrm{Var}(W)/(EW)^2$, which follows from Chebyshev's inequality, to show that $P(W = 0) \to 0$. (The variance of W can be bounded by bounding the probability that pairs of nodes are isolated.) It then follows that there is at least one isolated node, and hence the graph is disconnected whp. Instead, we shall obtain the result as a corollary of the first claim, which provides more detailed information regarding the number of isolated nodes. In particular, it shows that if $c_n \to c$, then the probability of being connected is asymptotically bounded by $e^{-e^{-c}}$.

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.

5

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 the red discs in any sequence. Say the red disc centred at x is chosen. 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 u and v in the red disc centred at x either come from the black disc that was centred at x or they come from black discs that were centred at y and z and overlapped the black disc centred at x. (Possibly, $y = z$ or $y = x$ or $z = x$). In the former case, the graph distance (number of edges along a shortest path) between nodes u and v in G is at most 2; in the latter case, it is at most 6 (since, if the discs centred at x and y intersect, there is a node a which is a neighbour of both x and y; thus, there is a path $u \rightarrow y \rightarrow a \rightarrow x \rightarrow b \rightarrow z \rightarrow v$ for some nodes a and b). 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 graph distance 6 apart (there is a path $u \rightarrow y \rightarrow a \rightarrow x \rightarrow b \rightarrow z \rightarrow v$ 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

$$\tilde{p}_n \;=\; 1 - (1 - p_n)^{(m_n+1)^2}$$

$$\begin{aligned}
&\geq\quad 1 - e^{-(m_n+1)^2 p_n} \\
&\geq\quad 1 - \exp\!\Big(-2(1+\delta)\frac{m_n+1}{n}\log\frac{n}{m_n+1}\Big), \quad\quad\quad (5)
\end{aligned}$$

where the last inequality follows if (1) is assumed to hold. Moreover, 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

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

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

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

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 [5], $P(C_n) \to 1$ as $n \to \infty$. This establishes the first claim in (2).

Moreover, it follows from [3, 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, $\text{diameter}(G) \leq 7\,\text{diameter}(\tilde{G}) + 6$, and the second claim in (2) follows.

Conversely, consider a sequence $n_k, k \in \mathbb{N}$ such that (3) is satisfied along the sequence n_k, m_{n_k}, p_{n_k}, for some $\delta > 0$. First, we argue that n_k nodes can be partitioned into sets of size either $\tilde{m}_{n_k} + 1$ or $\tilde{m}_{n_k} + 2$, where the $\tilde{m}_{n_k}$ are such that $(\tilde{m}_{n_k} + 1)/(m_{n_k} + 1) \to 1$ as $k \to \infty$. To see this, write

$$n_k = q(m_{n_k} + 1) + r, \text{ where } 0 \leq r \leq m_{n_k}, \quad\quad\quad (8)$$

and q and r are integers. Now define $s = \lfloor r/q \rfloor$ to be the integer part of r/q, and let $b = r - qs$ and $a = q - b$. Then a and b are integers, $0 \leq b < q$ and $a > 0$. Moreover, $r = qs + b = as + b(s+1)$. Hence, we can rewrite (8) as

$$n_k = a(m_{n_k} + s + 1) + b(m_{n_k} + s + 2), \text{ where } a + b = q.$$

Defining $\tilde{m}_{n_k} = m_{n_k} + s$, we see that n_k nodes can be partitioned into q sets, each of size $\tilde{m}_{n_k} + 1$ or $\tilde{m}_{n_k} + 2$.

7

Moreover, observe from (8) that

$$q \geq \frac{n_k}{m_{n_k} + 1} - 1, \text{ and so } s \leq \frac{m_{n_k}(m_{n_k} + 1)}{n_k - m_{n_k} - 1}.$$

Hence, by (3), $\frac{s}{m_{n_k}+1}$ tends to zero as k tends to infinity; consequently, $\frac{\tilde{m}_{n_k}+1}{m_{n_k}+1}$ tends to 1, as claimed above.

We now consider the following deterministic sequence of node configurations. The configuration on n_k nodes consists of clusters of size $\tilde{m}_{n_k} + 1$ or $\tilde{m}_{n_k} + 2$, where $\tilde{m}_{n_k}$ is defined as above. The nodes within each cluster are within Euclidean distance ϵ_{n_k} of each other, and any two nodes in distinct clusters are more than ϵ_{n_k} apart, for some positive constant ϵ_{n_k}. It is clear that such an ϵ_{n_k} and node configuration can be found.

Denote the clusters by $A_1, A_2, \ldots, A_{q_k}$. There are clearly no nearest neighbour edges between clusters, only shortcuts. Let G_1 denote the graph on q_k nodes obtained by replacing each cluster A_i by a single node i, and putting an edge between nodes i and j only if there is a shortcut in G between clusters A_i and A_j. Clearly, G is connected only if G_1 is (though the converse may not be true as the clusters A_i may not be connected). Now, conditional on the cluster sizes, there is a shortcut (at least one) between clusters A_i and A_j with probability

$$1 - (1 - p_{n_k})^{|A_i| \cdot |A_j|} \leq \tilde{p}_{n_k} := 1 - (1 - p_{n_k})^{(\tilde{m}_{n_k}+2)^2}. \tag{9}$$

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 $\tilde{m}_{n_k} + 1$ in G by adding a pseudo-node which is within distance ϵ_{n_k} of all nodes in this cluster. Shortcuts are present between pseudo-nodes and other nodes with the same probability p_{n_k} 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_i by a single node i, and put an edge between i and j in $\tilde{G}_1$ only if there is a shortcut between the (augmented) clusters A_i and A_j in $\tilde{G}$. It is clear from this construction that $\tilde{G}_1$ is a Bernoulli random graph on q_k nodes with edge probability $\tilde{p}_{n_k}$ given by (9). 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 [5] to show that, with high probability, $\tilde{G}$ fails to be connected.

Observe that $p_{n_k} \to 0$ as $k \to \infty$ by the assumption that m_{n_k} and p_{n_k} satisfy (3), i.e., that $n_k(m_{n_k} + 1)p_{n_k} < \log n_k$. Hence, for any $\epsilon > 0$, we have for all k sufficiently large that the probability of a shortcut between two clusters is bounded by

$$\tilde{p}_{n_k} \leq 1 - e^{-(1+\epsilon)(\tilde{m}_{n_k}+2)^2 p_{n_k}} \leq (1+\epsilon)(\tilde{m}_{n_k} + 2)^2 p_{n_k}.$$

Since the number of clusters q_k lies in $[\frac{n_k}{\tilde{m}_{n_k}+2}, \frac{n_k}{\tilde{m}_{n_k}+1}]$, we obtain using (3) that

$$
\begin{aligned}
q_k \tilde{p}_{n_k} &\leq (1+\epsilon) \frac{(\tilde{m}_{n_k}+2)^2}{(\tilde{m}_{n_k}+1)(m_{n_k}+1)}(m_{n_k}+1)n_k p_{n_k} \\
&\leq (1+\epsilon)(1-\delta) \frac{(\tilde{m}_{n_k}+2)^2(m_{n_k}+1)}{(\tilde{m}_{n_k}+1)(m_{n_k}+2)^2} \log \frac{n_k}{m_{n_k}+1}.
\end{aligned}
\tag{10}
$$

Now,

$$
\log \frac{n_k}{m_{n_k}+1} = \log \frac{n_k}{\tilde{m}_{n_k}+2} + \log \frac{\tilde{m}_{n_k}+2}{m_{n_k}+1} \leq (1+\epsilon) \log q_k
$$

for all k sufficiently large, because q_k tends to infinity whereas $\frac{\tilde{m}_{n_k}+1}{m_{n_k}+1}$ tends to 1, so that $\frac{\tilde{m}_{n_k}+2}{m_{n_k}+1}$ remains bounded as k tends to infinity. Since $\epsilon > 0$ can be chosen arbitrarily small in (10) and $\frac{\tilde{m}_{n_k}+1}{m_{n_k}+1}$ tends to 1, it follows that

$$
q_k \tilde{p}_{n_k} \leq (1-\delta') \log q_k,
$$

for any $\delta' \in (0, \delta)$ and all k sufficiently large. Moreover, $q_k \to \infty$ as $k \to \infty$ by the assumption that $m_n/n \to 0$. Hence, using the results in [5] on the connectivity of Bernoulli random graphs, we obtain that $P(\tilde{G}_1$ is connected$) \to 0$ as $k \to \infty$. But G is connected only if $\tilde{G}_1$ is. Therefore, $P(C_{n_k}) \to 0$ as $k \to \infty$. This completes the proof of the theorem.

Proof of Theorem 2: Let W denote the number of isolated nodes in the graph generated by Model B with parameters n, r_n and p_n. We do not make the dependence of W on the parameters explicit in the notation. Let Z denote a Poisson random variable with mean EW. We use the Stein-Chen method to show that W is close to Z in total variation distance.

Let U_1 have the distribution of the number of isolated nodes, and let $1 + V_1$ have the distribution of the number of isolated nodes conditional on node 1 being isolated. We shall construct U_1 and V_1 on the same probability space and show that $E|U_1 - V_1|$ is small. Then, we shall use [8, Chapter II, Theorem 24.3] to deduce that the total variation distance between W and Z is small.

The random variables U_1 and V_1 are constructed on the same probability space as follows. First, nodes 1 through n are placed uniformly at random on the square $[-\sqrt{n}/2, \sqrt{n}/2]^2$, and nearest neighbour and shortcut links are generated according to Model B. Next, all nodes located within distance r_n of node 1 (including node 1) are coloured red, while all other nodes are coloured green. Now, for each red node other than node 1, we place an associated blue node uniformly at random in the portion of the square $[-\sqrt{n}/2, \sqrt{n}/2]^2$ which excludes the circle of radius r_n centred at node 1. (With some abuse of notation, we will use the same node label to refer to such associated nodes, distinguishing them by their colour.) The blue nodes carry the same shortcuts as the red nodes with which they are associated. In other

9

words, if there is a shortcut between red nodes i and j (not equal to 1), then there is one between blue nodes i and j, and likewise if there is a shortcut between red node i and green node j. Finally, we put down nearest neighbour links between blue nodes and other blue or green nodes if they are within distance r_n of each other. Now, we define U_1 as the number of isolated nodes in the subgraph induced by red and green nodes, and V_1 as the number of isolated nodes in the subgraph induced by blue and green nodes. It is obvious that U_1 has the same distribution as W, the number of isolated nodes. To see that $1 + V_1$ has the distribution of the number of isolated nodes conditional on node 1 being isolated, observe that, conditional on this event, the remaining nodes are uniformly distributed outside the circle of radius r_n around node 1, and that the shortcut distribution between these nodes is unchanged, while there are no shortcuts to node 1. This is precisely the law of the subgraph on the blue and green nodes, and hence the number of isolated nodes in this subgraph has the same distribution as V_1.

Let $\mathcal{U}_1$ and $\mathcal{V}_1$ denote the set of isolated nodes in the red-green and blue-green subgraphs respectively. Let $\mathcal{R}$, $\mathcal{B}$ and $\mathcal{G}$ denote the set of red, blue and green nodes, with respective cardinalities R, B and G. In particular, $\mathcal{R} = \mathcal{B} \cup \{1\}$ and $\mathcal{R}$ and $\mathcal{G}$ partition the node set. Now,

$$EW = EU_1 = \sum_{i=1}^{n} P(i \in \mathcal{U}_1) = n\Big(1 - \frac{\pi r_n^2}{n}\Big)^{n-1}(1 - p_n)^{n-1} \sim e^{-c_n}, \qquad (11)$$

where the last equivalence follows from (4). Moreover,

$$E[|U_1 - V_1|] \leq \sum_{i=2}^{n} P(i \in \mathcal{U}_1, i \notin \mathcal{V}_1) + \sum_{i=2}^{n} P(i \notin \mathcal{U}_1, i \in \mathcal{V}_1). \qquad (12)$$

If i is one of the green nodes, then conditional on the event $\{i \in \mathcal{U}_1\}$, the event $\{i \notin \mathcal{V}_1\}$ can only occur if one of the blue nodes happens to fall within distance r_n of i; this happens independently for each blue node, with probability at most $\pi r_n^2 / (n - \pi r_n^2)$. Moreover, conditional on $\{i \in \mathcal{U}_1\}$, the number of blue nodes is binomially distributed with parameters $n - 2$ and $\pi r_n^2 / (n - \pi r_n^2)$ if nodes 1 and i are more than distance $2r_n$ apart; if they are less than $2r_n$ apart, the number of blue nodes is stochastically dominated by such a binomial random variable. Hence,

$$P(i \notin \mathcal{V}_1 | i \in \mathcal{U}_1 \cap \mathcal{G}) \leq 1 - \Big(1 - \frac{\pi^2 r_n^4}{(n - \pi r_n^2)^2}\Big)^{n-2} \leq \frac{(\log n + c_n)^2}{n}(1 + o(1)). \qquad (13)$$

We have used (4) to obtain the last inequality. On the other hand, if $i \neq 1$ is one of the red nodes, then the event $\{i \in \mathcal{U}_1\}$ can't occur (as all red nodes are neighbours of node 1, by definition). Thus, for $i \neq 1$, we have

$$
\begin{aligned}
P(i \in \mathcal{U}_1, i \notin \mathcal{V}_1) &= P(i \in \mathcal{U}_1 \cap \mathcal{G})P(i \notin \mathcal{V}_1 | i \in \mathcal{U}_1 \cap \mathcal{G}) \\
&\leq P(i \in \mathcal{U}_1)P(i \notin \mathcal{V}_1 | i \in \mathcal{U}_1 \cap \mathcal{G}) \leq \frac{(\log n + c_n)^2}{n^2}e^{-c_n}(1 + o(1)), (14)
\end{aligned}
$$

where the last inequality is obtained using (11) and (13).

Next, if i is a green node and $i \in \mathcal{V}_1$, then a necessary condition for the event $i \notin \mathcal{U}_1$ to occur is that node i either has a shortcut to node 1, or node i is at distance between r_n and $2r_n$ from node 1 (so that it has a neighbour in the red-green subgraph, which is absent in the blue-green subgraph). Thus, we have for each green node i that

$$P(i \notin \mathcal{U}_1, i \in \mathcal{V}_1 \cap \mathcal{G}) \leq P(\{i \in \mathcal{V}_1 \cap \mathcal{G}\} \cap A) + P(\{i \in \mathcal{V}_1 \cap \mathcal{G}\} \cap B), \qquad (15)$$

where A is the event that there is a shortcut between node i and node 1 and B the event that node i is at distance between r_n and $2r_n$ from node 1.

Any node $i \neq 1$ is isolated in the blue-green subgraph if it has no shortcuts to other blue or green nodes, and if there is no other node within distance r_n of it. Now, node i is located at least distance r_n away from node 1. Hence, at least half of the circle of radius r_n centred around it lies outside the circle of radius r_n centred around node 1, from which all nodes are moved out. Hence, the probability that a given node j lies within distance r_n of node i is at least $\pi r_n^2 / 2(n - \pi r_n^2)$. Thus, we see that

$$P(i \in \mathcal{V}_1 | A) = P(i \in \mathcal{V}_1) \leq (1 - p_n)^{n-2} \Big(1 - \frac{\pi r_n^2}{2(n - \pi r_n^2)} \Big)^{n-2}. \qquad (16)$$

Hence,

$$
\begin{aligned}
P(\{i \in \mathcal{V}_1 \cap \mathcal{G}\} \cap A) &\leq P(A)P(i \in \mathcal{V}_1 | A) \\
&\leq p_n (1 - p_n)^{n-2} \Big(1 - \frac{\pi r_n^2}{2(n - \pi r_n^2)} \Big)^{n-2} \\
&\leq \frac{p_n e^{-c_n/2}}{\sqrt{n}} [1 + o(1)] \;\leq\; e^{-c_n/2} n^{-3/2} \log n [1 + o(1)]. \qquad (17)
\end{aligned}
$$

The last two inequalities follow from (4). Likewise, we have

$$
\begin{aligned}
P(\{i \in \mathcal{V}_1 \cap \mathcal{G}\} \cap B) &\leq P(B)P(i \in \mathcal{V}_1 | B) \\
&\leq \frac{3\pi r_n^2}{n - \pi r_n^2} (1 - p_n)^{n-2} \Big(1 - \frac{\pi r_n^2}{2(n - \pi r_n^2)} \Big)^{n-2} \\
&\leq \frac{3\pi r_n^2}{n} \frac{e^{-c_n/2}}{\sqrt{n}} [1 + o(1)] \;\leq\; 3 e^{-c_n/2} n^{-3/2} \log n [1 + o(1)]. \quad (18)
\end{aligned}
$$

Combining (15), (17) and (18), we obtain that

$$P(i \notin \mathcal{U}_1, i \in \mathcal{V}_1 \cap \mathcal{G}) \leq 4 e^{-c_n/2} n^{-3/2} \log n [1 + o(1)]. \qquad (19)$$

On the other hand, if $i \neq 1$ is a red node, then automatically i is not isolated (in the red-green subgraph), so that

$$P(i \notin \mathcal{U}_1, i \in \mathcal{V}_1 \cap \mathcal{R}) \;=\; P(i \in \mathcal{V}_1 \cap \mathcal{R}) \;=\; P(i \in \mathcal{R})P(i \in \mathcal{V}_1).$$

11

Using (16), we obtain that

$$P(i \notin \mathcal{U}_1, i \in \mathcal{V}_1 \cap \mathcal{R}) \leq \frac{\pi r_n^2}{n}\left(1 - \frac{\pi r_n^2}{2(n - \pi r_n^2)}\right)^{n-2}(1-p_n)^{n-2} \leq \frac{\log n + c_n}{n^2}e^{-c_n}[1+o(1)]. \quad (20)$$

Combining (19) and (20), we get

$$P(i \notin \mathcal{U}_1, i \in \mathcal{V}_1) \leq 4e^{-c_n/2}n^{-3/2}\log n[1 + o(1)]. \quad (21)$$

Finally, substituting (14) and (21) in (12) yields

$$E[|U_1 - V_1|] \leq \frac{4\log n}{\sqrt{n}}e^{-c_n/2}\,[1 + o(1)]. \quad (22)$$

Recall that the number of isolated nodes, W, is the sum of Bernoulli random variables corresponding to the indicators that individual nodes are isolated, and that Z is a Poisson random variable with the same mean as W. (The dependence of W and Z on n has been suppressed in the notation.) Hence, by (11), (22) and [8, Chapter II, Theorem 24.3],

$$d_{TV}(W, Z) \leq 2(1 \wedge [EW]^{-1}) \cdot EW \cdot E[|U_1 - V_1|] \leq \frac{8\log(n)e^{-c_n/2}}{\sqrt{n}}[1 + o(1)], \quad (23)$$

where $d_{TV}(W, Z)$ denotes the total variation distance between the random variables W and Z. Suppose $c_n \to c$ and let Y denote a Poisson random variable with mean e^{-c}. Since $EZ = EW \sim e^{-c_n}$, it is easy to see that $d_{TV}(Z, Y)$ tends to zero. By the triange inequality, so does $d_{TV}(W, Y)$. In other words, W converges in distribution to a Poisson random variable with mean e^{-c} as n tends to ∞. This establishes the first claim of the theorem.

To establish the second claim, first note that

$$P(W = 0) \leq P(Z = 0) + d_{TV}(W, Z) = e^{-EW} + d_{TV}(W, Z),$$

so that, by (11) and (23),

$$P(W = 0) \leq \left(\exp(-e^{-c_n}) + \frac{8\log(n)e^{-c_n/2}}{\sqrt{n}}\right)[1 + o(1)].$$

In particular, if c_n is fixed at c, then $P(W = 0) \to e^{-e^{-c}}$. Now, if $c_n \to -\infty$, then for arbitrarily large negative c, we have $c_n \leq c$ for all n large enough. Moreover, if we augment the nth graph with additional random shortcuts so that $\pi r_n + np_n = \log n + c$, then this does not reduce the number of isolated nodes. Hence, $P(W = 0) \leq 2e^{-e^{-c}}$ say, for all n sufficiently large. Since $-c$ can be chosen arbitrarily large, $P(W = 0) \to 0$ as $n \to \infty$. In other words, there is at least one isolated node, and so the graph is disconnected, whp. 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 (up to constants) 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. Finally, the results here are obtained for a uniform shortcut distribution. The extension to random connection models where the probability of a shortcut between two nodes is some function of the Euclidean distance between them is an open problem.

References

[1] Paul Balister, Bela Bollobas, Amites Sarkar and Mark Walters, "Connectivity of random k-nearest-neighbour graphs", *Adv. Appl. Prob.*, 37: 1–24, 2005.

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

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

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

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

[6] 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.

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

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

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

[10] Mathew Penrose, "The longest edge of the random minimal spanning tree", in *Ann. Appl. Probab.*, 7: 340–361, 1997.

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

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

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