

On the Giant Component in Wireless Multi-hop Networks

Xiaoyuan Ta^{*†}, Guoqiang Mao^{*†}, and Brian D.O. Anderson^{†§}

^{*}School of Electrical and Information Engineering, The University of Sydney

[†]Research School of Information Sciences and Engineering, Australian National University

[‡]National ICT Australia Limited[1], Sydney, Australia

[§]National ICT Australia Limited, Canberra, Australia

Abstract—In this paper, we study the giant component, the largest component containing a non-vanishing fraction of nodes, in a wireless multi-hop network where n nodes are randomly and uniformly distributed in $[0, 1]^d$ ($d = 1, 2$) and any two nodes can communicate directly with each other iff their Euclidean distance is not larger than the transmission range r . We investigate the probability that the size of the giant component is at least a given threshold p with $0.5 < p \leq 1$. For $d = 1$, we derive a closed-form analytical formula for this probability. For $d = 2$, we propose an empirical formula for this probability using simulations. In addition, we compare the transmission range required for having a connected network with the transmission range required for having a certain size giant component for $d = 2$. The comparison shows that a significant energy saving can be achieved if we only require most nodes (e.g. 95%) to be connected to the giant component rather than require all nodes to be connected. The results of this paper are of practical value in the design and analysis of wireless ad hoc networks and sensor networks.

I. INTRODUCTION

A wireless multi-hop network, e.g. wireless ad hoc/sensor network, usually consists of a number of decentralized and self-organized nodes that communicate with each other over wireless channels, and packets are forwarded hop-by-hop from the source to the destination. Connectivity is a prerequisite for many network functions, e.g. routing, topology control, in such wireless multi-hop networks [1], [2], [3]. A network is connected iff (if and only if) for any pair of two nodes, there is at least one path between them. Otherwise, the network is said to be disconnected. Generally, a disconnected network may consist of a number of isolated nodes and several components. We call the largest component containing a non-vanishing fraction of nodes the *giant component* [4].

In the past several years, the connectivity problem in wireless multi-hop networks has been widely investigated and significant outcomes have been achieved [2], [3], [5], [6]. However, from a practical point of view, requiring all nodes to be connected may be a too stringent condition to satisfy. It has been shown by simulations that the transmission range required for a large percentage of nodes to be connected is much less than the transmission range for all nodes to be connected [1], [4], [7]. Fig. 1 shows the average value of the

ratio of the transmission range required for 95% of nodes to be connected to the transmission range required for a connected network. As shown in the figure, when the number of nodes

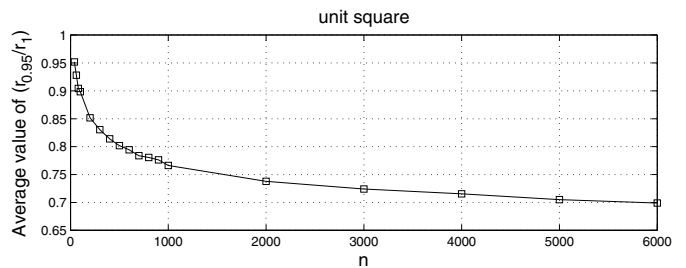


Fig. 1. Simulation results: Average value of ratio $r_{0.95}/r_1$. r_1 : transmission range required for a connected network; $r_{0.95}$: transmission range required for 95% of nodes to be connected. The ratio shown is the average value, and each average value is obtained over 2000 random topologies, in which a total of n nodes are uniformly and randomly distributed on a unit square.

is 1000, the transmission range required for 95% nodes to be connected is 24% less than that required for a connected network. Based on a conservative estimate that the required transmission power increases with the square of the required transmission range, this means an energy saving of at least 42%. In addition, the ratio decreases as the total number of nodes n increases. This means that the energy saving is more significant in a network with a large number of nodes. In practice, many network applications do not require all nodes to be connected and having a few disconnected nodes is not critical [8]. Examples of applications of this type might be habitat monitoring for wild animals, monitoring ocean temperature. Hence, it is practical and significant to investigate the giant component, and to quantify, as far as possible, the extent of this saving (as opposed simply to observing the possibility of saving).

In this paper, we investigate the probability that the *size* of the giant component achieves a given threshold. The *size* of the giant component is defined as the ratio of the number of nodes in the giant component to the total number of nodes in the network [4]. More specifically, we assume that a total of n nodes are randomly and independently distributed in $[0, 1]^d$ ($d = 1, 2$) following a uniform distribution and each node has a uniform transmission range of r . Any two nodes can communicate directly with each other iff their Euclidean

¹National ICT Australia is funded by the Australian Government Department of Communications, Information Technology and the Arts and the Australian Research Council through the *Backing Australia Ability initiative* and the ICT Centre of Excellence Program.

distance is at most r . This network model has been widely used in the area [1], [2], [6], [9]. Based on this result one can derive the minimum transmission range at which a nominated large percentage of nodes are connected with a high probability. The advantage of using this minimum transmission range, rather than a higher transmission range for a connected network, is that both power consumption and interference can be reduced while meaningful services can still be provided.

The rest of this paper is organized as follows. Section II reviews related work. Section III describes the network model. In section IV, we derive a closed-form analytical formula for computing the probability that the giant component size achieves a given threshold for $d = 1$. In section V, we propose an empirical formula for the probability that the giant component size achieves a given threshold for $d = 2$ using simulations. Section VI concludes this paper and discusses future work.

II. RELATED WORK

There has been extensive work on the giant component for *Bernoulli random graphs*, and an analytical formula relating the giant component size and the average node degree² has been found [10]. However, the Bernoulli random graph is not suitable for modeling wireless multi-hop networks, because in Bernoulli random graphs, the probability of existence of a link between any two nodes is independent of their Euclidean distance, whereas in wireless multi-hop networks, such probability is dependent on the Euclidean distance. Therefore, it is inappropriate to apply the results on the giant component size obtained from Bernoulli random graphs directly into wireless multi-hop networks.

In [4], Hekmat *et al.* investigated the giant component size in a log-normal shadowing environment, where a total of n nodes are randomly and uniformly distributed on a square and a link exists between two nodes if the power received at one node from the other node, as determined by the log-normal shadowing model, is greater than a given threshold. Based on the analytical results obtained in Bernoulli random graphs, the authors proposed an empirical formula relating the giant component size and the average node degree in random geometric graphs (refer to [5]). In [11], Németh *et al.* investigated the giant component size by using a fractal propagation model where the probability of having a link between two nodes is determined by their Euclidean distance and two non-negative constants. They found that the giant component size can be characterized by a single parameter, viz., the average node degree. However, the empirical formula of the former paper only works well when the variance of the shadowing is large; the propagation model used in the latter paper may be suitable only for restricted environments.

III. NETWORK MODEL

Generally, a wireless multi-hop network can be represented by an undirected graph $G(V, E)$ with a set of vertices V and

²Average node degree is the average number of neighbors of an arbitrary node.

a set of edges E . Each vertex of the set V uniquely represents a node and each edge of the set E uniquely represents a wireless link, and vice versa. For many purposes, a so-called *random geometric graph* $G_d(n, r)$ can be used for modeling such wireless multi-hop networks, where the subscript d is used to emphasize the dimension. In this paper, we shall model the network by a random geometric graph $G_d(n, r)$ which is typically defined as follows.

Definition 1 ([5], [12]). Given $n \in \mathbb{N}$ and $r \in [0, 1]$, a random geometric graph $G_d(n, r)$ is a graph in which n vertices are randomly and independently distributed in the unit cube $[0, 1]^d$ following a uniform distribution, and any two vertices u and v are directly connected iff $\|u - v\| \leq r$, where the norm $\|\cdot\|$ means the Euclidean norm.

Remark. In this paper, we assume that any two nodes can communicate directly with each other iff their Euclidean distance is less than or equal to the transmission range r , i.e. the unit disk communication model. The unit disk communication model simplifies the analysis, but it is interesting and important to investigate how conclusions obtained under the unit disk communication model extend to a more realistic channel model. In [13], Santi *et al.* claimed that the basic nature of the results would not change if more realistic channel models were considered based on the simulation results in [14]. Therefore, the impact of different channel models on the results in the paper is likely to be quantitative, i.e. the nature and the trend (e.g. the ratio $r_{0.95}/r_1$ decreases quickly with n) revealed in the results will not be affected.

IV. GIANT COMPONENT SIZE IN ONE-DIMENSIONAL SPACE

One-dimensional wireless multi-hop networks are indeed useful in many real world scenarios. Examples of such applications might be a vehicular network built along a highway, a network deployed along an attack route in battlefield, and other networks where nodes are placed on a line [15]. In this section, we shall derive a closed-form analytical formula for the probability of having a giant component with size above a given threshold p ($0.5 < p \leq 1$) in one-dimensional space ($d = 1$), denoted by $P_1(n, r, p)$. We are interested in all values of n , not just asymptotically large n .

In order to derive the probability $P_1(n, r, p)$, the following three lemmas, viz., Lemma 1, Lemma 2 and Lemma 3 are needed. Lemma 2 and Lemma 3 are used to derive $P_1(n, r, p)$; Lemma 1 is used to prove Lemma 2 and Lemma 3.

Lemma 1 (Lemma 1 in [12]). *Let $[x, x+y]$ be a subinterval of length y on a unit interval $[0, 1]$. Let two of k given vertices be placed on the borders of this subinterval. Let $P(k, y, r)$ be the probability that the remaining $k - 2$ vertices placed randomly and uniformly on $[0, 1]$ are inside $[x, x+y]$ and the k vertices form a connected subgraph of length y . Then*

$$P(k, y, r) = \sum_{j=0}^{\min\{k-1, \lfloor y/r \rfloor\}} \binom{k-1}{j} (-1)^j (y - jr)^{k-2}. \quad (1)$$

Lemma 2. Let \mathcal{F}_k^1 denote the event that there exists a connected subgraph with exactly k ($k < n$) vertices in $G_1(n, r)$ and both endpoints of this subgraph are not within distance r from the borders of the unit interval, and none of the remaining $n - k$ vertices is connected to this subgraph (refer to Fig. 2). Then

$$\Pr\{\mathcal{F}_k^1\} = (n - k + 1) \sum_{j=0}^{\min\{k-1, \lfloor \frac{1}{r} \rfloor - 2\}} \binom{k-1}{j} (-1)^j (1 - (j+2)r)^n. \quad (2)$$

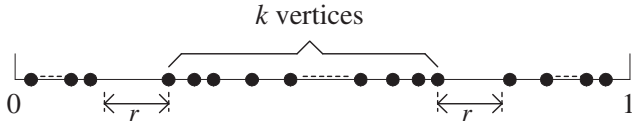


Fig. 2. Illustration of event \mathcal{F}_k^1 .

Proof: There are $\binom{n}{k}$ distinct combinations for selecting k vertices from a total of n vertices. Consider a subinterval $[x, x+y]$, where x and $x+y$ are the positions of the left border and the right border respectively. For any given k vertices, there are $\binom{k}{2}$ different combinations for selecting 2 vertices as endpoints, and two permutations of each selection in placing them on the borders of $[x, x+y]$; the probability that the remaining $k-2$ vertices placed randomly and uniformly on $[0, 1]$ are inside $[x, x+y]$ and the k vertices form a connected subgraph is given by Eq. 1. Then

$$\begin{aligned} \Pr\{\mathcal{F}_k^1\} &= \binom{n}{k} 2 \binom{k}{2} \int_0^{1-2r} \left(\int_r^{1-r-y} dx \right) \\ & P(k, y, r) (1 - y - 2r)^{n-k} dy \\ &= \binom{n}{k} 2 \binom{k}{2} \int_0^{1-2r} P(k, y, r) (1 - y - 2r)^{n-k+1} dy. \end{aligned} \quad (3)$$

Dividing the integration interval $[0, 1 - 2r]$ into subintervals, i.e. $[0, r)$, $[r, 2r)$, ..., and using Lemma 1, Eq. 3 becomes

$$\begin{aligned} \Pr\{\mathcal{F}_k^1\} &= \binom{n}{k} 2 \binom{k}{2} \sum_{i=0}^{\lfloor \frac{1}{r} \rfloor - 3} \int_{ir}^{(i+1)r} (1 - y - 2r)^{n-k+1} \\ & \left[\sum_{j=0}^{L(i)} \binom{k-1}{j} (-1)^j (y - jr)^{k-2} \right] dy \\ & + \binom{n}{k} 2 \binom{k}{2} \int_{\lfloor \frac{1}{r} \rfloor r - 2r}^{1-2r} (1 - y - 2r)^{n-k+1} \\ & \left[\sum_{j=0}^{L(i)} \binom{k-1}{j} (-1)^j (y - jr)^{k-2} \right] dy, \end{aligned} \quad (4)$$

where $L(i) = \min\{k-1, i\}$. Then taking the inner sums outside the integrals, and letting $L' = \min\{k-1, \lfloor 1/r \rfloor - 2\}$,

Eq. 4 becomes

$$\begin{aligned} \Pr\{\mathcal{F}_k^1\} &= \binom{n}{k} 2 \binom{k}{2} \sum_{j=0}^{L'} \binom{k-1}{j} (-1)^j \\ & \cdot \left(\int_{jr}^{1-2r} (y - jr)^{k-2} (1 - y - 2r)^{n-k+1} dy \right) \\ &= \binom{n}{k} 2 \binom{k}{2} \sum_{j=0}^{L'} \binom{k-1}{j} (-1)^j \\ & \cdot (1 - jr - 2r)^n \left(\int_0^1 t^{k-2} (1 - t)^{n-k+1} dt \right). \end{aligned} \quad (5)$$

Note that the integral on the right hand side of Eq. 5 is the *Beta Function*. Therefore, it follows

$$\int_0^1 t^{k-2} (1 - t)^{n-k+1} dt = \frac{(k-2)!(n-k+1)!}{n!}. \quad (6)$$

Inserting Eq. 6 into Eq. 5, Eq. 2 can readily be obtained. ■

Lemma 3. Let \mathcal{F}_k^2 denote the event that there exists a connected subgraph with exactly k ($k < n$) vertices in $G_1(n, r)$ and the leftmost vertex of the subgraph is located within distance r from the left border of the unit interval and the remaining $n - k$ vertices are all located on the right side of this subgraph and none of them is connected to this subgraph (refer to Fig. 3). Then

$$\Pr\{\mathcal{F}_k^2\} = \sum_{j=0}^{\min\{k, \lfloor \frac{1}{r} \rfloor - 1\}} \binom{k}{j} (-1)^j (1 - (j+1)r)^n. \quad (7)$$

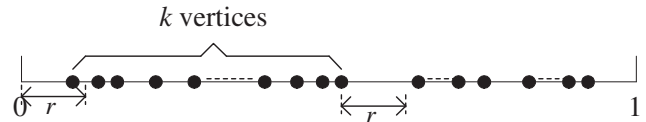


Fig. 3. Illustration of event \mathcal{F}_k^2 .

Proof: Similar to the proof of Lemma 2, we have

$$\Pr\{\mathcal{F}_k^2\} = \binom{n}{k} 2 \binom{k}{2} \int_0^{1-r} \left[\int_0^X P(k, y, r) (1 - y - x - r)^{n-k} dx \right] dy,$$

where $X = \min\{r, 1 - r - y\}$. After simplifications, we have

$$\begin{aligned} \Pr\{\mathcal{F}_k^2\} &= \binom{n}{k} 2 \binom{k}{2} \int_0^{1-r} P(k, y, r) \frac{(1 - r - y)^{n-k+1}}{n - k + 1} dy \\ & - \binom{n}{k} 2 \binom{k}{2} \int_0^{1-2r} P(k, y, r) \frac{(1 - 2r - y)^{n-k+1}}{n - k + 1} dy. \end{aligned}$$

Dividing the integration interval into subintervals, i.e. $[0, r)$, $[r, 2r)$, ..., and using the same method as described in the proof

of Lemma 2, we have

$$\begin{aligned}
 & Pr\{\mathcal{F}_k^2\} \\
 &= \sum_{j=0}^{\min\{k-1, \lfloor \frac{1}{r} \rfloor - 1\}} \binom{k-1}{j} (-1)^j (1 - (j+1)r)^n \\
 &- \sum_{j=0}^{\min\{k-1, \lfloor \frac{1}{r} \rfloor - 2\}} \binom{k-1}{j} (-1)^j (1 - (j+2)r)^n. \quad (8)
 \end{aligned}$$

Combining the two sums in Eq. 8, the result follows. ■

Now we can state the following result in this paper based on Lemma 2 and Lemma 3.

Theorem 1. Consider a random geometric graph in \mathbb{R}^d ($d = 1$), denoted as $G_1(n, r)$. Let p be a fixed real number in $(0.5, 1]$. Let $P_1(n, r, p)$ be the probability that $G_1(n, r)$ has a giant component of size at least p . Then,

$$\begin{aligned}
 & P_1(n, r, p) \\
 &= \sum_{i=\lceil np \rceil}^{n-1} \left[2 \sum_{j=0}^{\min\{i, \lfloor \frac{1}{r} \rfloor - 1\}} \binom{i}{j} (-1)^j (1 - (j+1)r)^n \right. \\
 &+ (n - i + 1) \\
 &\cdot \left. \sum_{j=0}^{\min\{i-1, \lfloor \frac{1}{r} \rfloor - 2\}} \binom{i-1}{j} (-1)^j (1 - (j+2)r)^n \right] \\
 &+ \sum_{j=0}^{\min\{n-1, \lfloor \frac{1}{r} \rfloor\}} \binom{n-1}{j} (-1)^j (1 - jr)^n. \quad (9)
 \end{aligned}$$

Proof: Let $C(G_1) = C(G_1(n, r))$ denote the number of nodes in the giant component in $G_1(n, r)$. Then it is clear that

$$P_1(n, r, p) = \sum_{i=\lceil np \rceil}^n Pr\{C(G_1) = i\}. \quad (10)$$

From Eq. 10, the probability $P(n, r, p)$ can be obtained combining two calculations:

- 1) $Pr\{C(G_1) = n\}$. This probability is actually the probability that the network $G_1(n, r)$ is connected, denoted as P_{con} . It is given by Corollary 1 in [12] as

$$P_{con} = \sum_{j=0}^{\min\{n-1, \lfloor 1/r \rfloor\}} \binom{n-1}{j} (-1)^j (1 - jr)^n. \quad (11)$$

- 2) $Pr\{C(G_1) = i\}$ for $\lceil np \rceil \leq i < n$. This probability equals the probability that there exists a connected subgraph with exactly i ($\lceil np \rceil \leq i < n$) vertices and none of the remaining $n - i$ vertices is connected to this subgraph. There are four different sub-cases in which this event may happen, i.e.

- a) Both endpoints of this subgraph are not within a distance r from the borders of the unit interval. Lemma 2 provides the probability for this case.
- b) The left (right) endpoint of this subgraph is within a distance r from the left (right) border of the unit

interval. Lemma 3 provides the probability for this case.

- c) Both endpoints of this subgraph are within distance r from the borders of the unit interval. This can only happen when $i = n$, but here we require $i < n$. Hence, the probability of this case is zero.

Note that in Theorem 1 it is required that $p > 0.5$. In Lemma 2 and Lemma 3, the connected subgraph with exactly i vertices is not necessarily the largest connected subgraph which we are interested in. To ensure that the connected subgraph with exactly i vertices is the largest connected subgraph, it suffices that we restrict $p > 0.5$, so that for all $i \geq \lceil np \rceil$, the size of a connected subgraph with exactly i vertices is larger than 0.5 and is necessarily the largest connected subgraph.

Finally, the probability $P_1(n, r, p)$ as a key result in this paper can be readily derived as

$$P_1(n, r, p) = \sum_{i=\lceil np \rceil}^{n-1} [2Pr\{\mathcal{F}_i^2\} + Pr\{\mathcal{F}_i^1\}] + P_{con}. \quad (12)$$

Substituting Eq. 2, Eq. 7 and Eq. 11 into Eq. 12, we can readily obtain Eq. 9. ■

V. GIANT COMPONENT SIZE IN TWO-DIMENSIONAL SPACE

In a two-dimensional wireless multi-hop network, the analysis for the giant component becomes much more complicated. It is difficult to obtain an analytical formula comparable to the one-dimensional case. In this section, we shall develop an *empirical* formula for estimating the probability that the giant component size is at least p , denoted by $P_2(n, r, p)$, using simulations. In sub-section V-C, we compare the transmission range required for having a giant component of size above p with the transmission range required for having a connected network. Based on this comparison, we show that a significant energy saving can be achieved if we only require most nodes (95%) to be connected rather than require all nodes to be connected.

A. Simulation

We use the *Monte Carlo* method to compute probability $P_2(n, r, p)$. In the simulation environment, a total of n ($20 \leq n \leq 200$) nodes are randomly and independently distributed in a unit square following a uniform distribution. We vary the transmission range r . For each given r , the links between the nodes are constructed, and at each r we find whether the network has a giant component with size at least p . This process is repeated 10000 times for each r , and finally by averaging on the 10000 random topologies, the empirical probability that the network has a giant component with size at least p at that specific r is obtained.

B. Empirical formula

By fitting extensive simulation results, we find that, for a given $p \in (0.5, 1)$, and for $P_2(n, r, p) \in [0.5, 0.99]$

$$P_2(n, r, p) \approx \frac{\exp\left(\frac{r-r_c}{z}\right)}{1 + \exp\left(\frac{r-r_c}{z}\right)}, \quad (13)$$

where r_c and z are parameters of the model and are functions of p . It is clear that $[0.5, 0.99]$ is the range we are more interested in practice, since the network is not well connected when $P_2(n, r, p)$ is small. Note that this empirical formula has a similar form to the one proposed in [3] which dealt with connectivity of networks.

From Eq. 13, we can obtain an approximately linear relationship between r and $\log\left(\frac{P_2(n, r, p)}{1 - P_2(n, r, p)}\right)$, i.e.

$$r \approx z \cdot \log\left(\frac{P_2(n, r, p)}{1 - P_2(n, r, p)}\right) + r_c. \quad (14)$$

The parameters r_c and z can then be estimated by fitting the linear function given by Eq. 14. Table. I shows the estimated values of r_c and z for different values of n , where the network area is fixed as a unit square. ζ in the table represents the

TABLE I
ESTIMATION OF r_c AND z . UNIT SQUARE NETWORK AREA.

n	p = 0.95			p = 0.90		
	z	r_c	ζ	z	r_c	ζ
20	0.0317	0.3056	0.9976	0.0311	0.2915	0.9962
40	0.0220	0.2213	0.9992	0.0195	0.2114	0.9972
60	0.0159	0.1817	0.9978	0.0133	0.1753	0.9991
80	0.0125	0.1580	0.9993	0.0105	0.1516	0.9985
100	0.0108	0.1407	0.9989	0.0086	0.1357	0.9992
120	0.0091	0.1288	0.9986	0.0074	0.1240	0.9991
140	0.0080	0.1196	0.9957	0.0067	0.1145	0.9973
160	0.0071	0.1116	0.9982	0.0059	0.1071	0.9989
180	0.0063	0.1053	0.9978	0.0051	0.1012	0.9991
200	0.0057	0.0999	0.9982	0.0046	0.0962	0.9990

linear correlation coefficient which indicates the strength of a linear relationship between two random variables. We can see that for both $p = 0.95$ and $p = 0.90$, ζ is very close to 1, which indicates a very good linearity. Hence, the proposed model works very well and is accurate. Note that the proposed model is also valid for lower values of p (e.g. $p = 0.60$, but the corresponding simulation results have been omitted).

In order to further check the accuracy of the proposed model, we compare in Fig. 4 the calculated results from the empirical formula, i.e. Eq. 13, with the corresponding simulation results. For both $p = 0.95$ and $p = 0.90$, we choose four values of n , i.e. 140, 160, 180 and 200, and then choose the corresponding values of z and r_c from Table I, and then substitute them into Eq. 13 to calculate $P_2(n, r, p)$. As shown in Fig. 4, the calculated results and the simulation results match very well.

The model given by Eq. 14 can be easily scaled. For the sake of simplicity, we consider the network area as a square of size $l \times l$. Then set

$$z = \alpha \cdot l \quad \text{and} \quad r_c = \beta \cdot l. \quad (15)$$

Substituting Eq. 15 into Eq. 14, we obtain the modified linear model relating $\frac{r}{l}$ and $\log\left(\frac{P_2(n, r, p)}{1 - P_2(n, r, p)}\right)$, i.e.

$$\frac{r}{l} \approx \alpha \cdot \log\left(\frac{P_2(n, r, p)}{1 - P_2(n, r, p)}\right) + \beta. \quad (16)$$

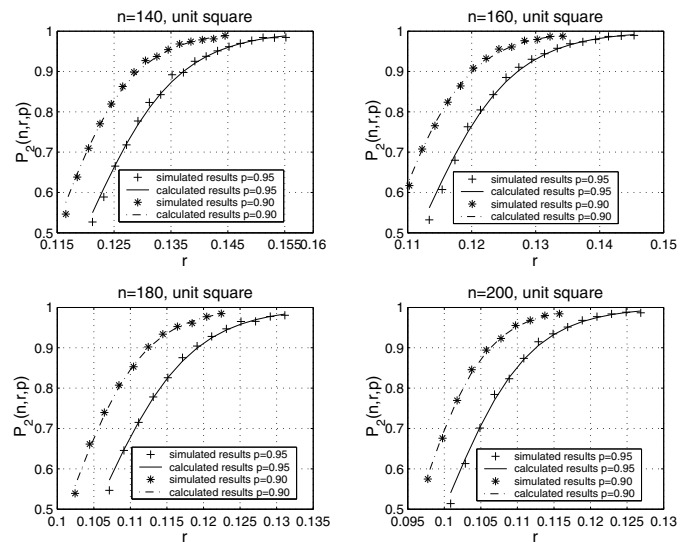


Fig. 4. Comparison between the calculated results from the empirical formula and the simulation results for $n = 140, 160, 180$ and 200 . The calculated results are obtained using Eq. 13 and values of z and r_c in Table I.

The left hand term of Eq. 16 is the normalized transmission range with respect to l . In this paper, the network area corresponds to the special case when $l = 1$. From the simulation results, we find that for fixed p , the parameters α and β are only determined by n , i.e. $\alpha n^2 \log n$ is approximately a linear function of n ; and β is approximately a linear function of $\sqrt{\log n/n}$ which is similar to the functional dependence revealed in [3] for connectivity. Fig. 5 shows the values of

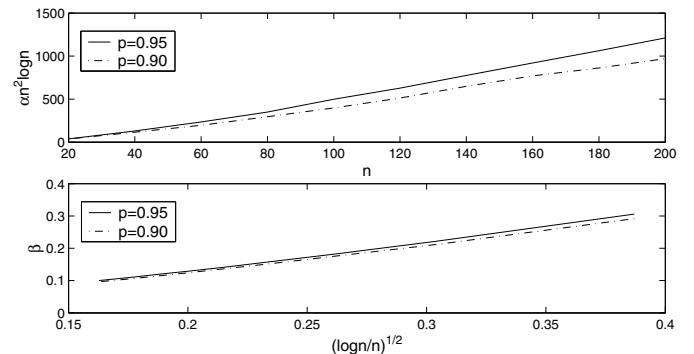


Fig. 5. Plots of $\alpha n^2 \log n$ versus n , and β versus $\sqrt{\log n/n}$.

$\alpha n^2 \log n$ with respect to n , and β with respect to $\sqrt{\log n/n}$ respectively. For $p = 0.95$, the linear correlation coefficients for α and β are 0.9980 and 0.9987 respectively; for $p = 0.90$, the linear correlation coefficients for α and β are 0.9977 and 0.9990 respectively. The linear models for α and β are given by

$$p = 0.95 : \begin{cases} \alpha \approx \frac{6.6418n - 145.9807}{n^2 \log n}, \\ \beta \approx 0.9107 \sqrt{\frac{\log n}{n}} - 0.0521. \end{cases} \quad (17)$$

$$p = 0.90 : \begin{cases} \alpha \approx \frac{5.3535n - 108.5792}{n^2 \log n}, \\ \beta \approx 0.8659 \sqrt{\frac{\log n}{n}} - 0.0482. \end{cases} \quad (18)$$

Now, we can estimate the probability $P_2(n, r, p)$ given n , l , r and p either based on Eq. 13 and Eq. 15 when $l = 1$ or based on Eq. 16 when $l \neq 1$. For instance, given $p = 0.95$ and $l \neq 1$, we can estimate α and β from Eq. 17, and then we can estimate $P_2(n, r, p)$ from Eq. 16. We can also estimate the transmission range required to achieve a desired value of probability $P_2(n, r, p)$ from Eq. 16.

Using the method described above, we record in Table II, for fixed p , some specific values of n and the normalized transmission range $\frac{r}{l}$ which yield a giant component of size above p with probability 95% (i.e. $P_2(n, r, p) = 0.95$). These

TABLE II
VALUES OF THE NORMALIZED TRANSMISSION RANGE $\frac{r}{l}$ GUARANTEEING $P_2(n, r, p) = 95\%$ WITH $p = 0.90$ AND 0.95 .

n	$p = 0.95$		$p = 0.90$	
	r/l	r/l	r/l	r/l
20	0.3988	0.3830	120	0.1556
40	0.2859	0.2688	140	0.1432
60	0.2286	0.2145	160	0.1325
80	0.1947	0.1826	180	0.1239
100	0.1725	0.1611	200	0.1168

results can be applied directly into the design procedure of the network and may provide useful insight for the relative magnitude of the transmission range with respect to network area required for having a certain size giant component with a high probability.

C. Energy savings

In this sub-section, we estimate the energy saving if most nodes rather than all nodes are required to be connected based on a conservative estimate that the required transmission power increases with the square of the required transmission range. Let R_p be the transmission range required for having a giant component of size above p with probability 95% (i.e. $P_2(n, R_p, p) = 0.95$, one may also be interested in other values of probability such as 98%, 99%); and let R_1 be the transmission range required for having a connected network with probability 95%. Then we use $\frac{R_1^2 - R_p^2}{R_1^2}$ as the metric to measure the energy saving. R_p can be obtained in the same way as we obtain results in Table II. R_1 can be obtained using the method given in [3]. Since the empirical method given in [3] is developed for $n \leq 125$, we focus only on $n \leq 125$. The

TABLE III
ESTIMATION OF ENERGY SAVINGS IF MOST NODES RATHER THAN ALL NODES ARE REQUIRED TO BE CONNECTED. R_1 IS THE TRANSMISSION RANGE REQUIRED FOR HAVING A CONNECTED NETWORK WITH PROBABILITY 95%; $R_{0.95}$ ($R_{0.90}$) IS THE TRANSMISSION RANGE REQUIRED FOR HAVING A GIANT COMPONENT OF SIZE ABOVE 0.95 (0.90) WITH PROBABILITY 95%.

n	R_1/l	$R_{0.95}/l$	Energy sav.	$R_{0.90}/l$	Energy sav.
20	0.4454	0.3988	19.8%	0.3830	26.1%
40	0.3209	0.2859	20.6%	0.2688	29.8%
60	0.2625	0.2286	24.1%	0.2145	33.2%
80	0.2263	0.1947	25.9%	0.1826	34.9%
100	0.2009	0.1725	26.2%	0.1611	35.7%
120	0.1817	0.1556	26.6%	0.1459	36.1%

comparison is shown in Table III. We can see that the energy saving is significant and increases with n .

VI. CONCLUSION AND FUTURE WORK

In this paper, we investigated the probability that a network has a giant component with size above a given threshold in wireless multi-hop networks. In the one dimensional case, we derived a closed-form analytical formula for calculating this probability. In the two dimensional case, we proposed an empirical formula for estimating this probability by fitting the simulation results. The empirical formula is shown to be very accurate for $n \leq 200$.

Part of our future work will be investigating the probability of having a giant component analytically in two and three dimensional networks, and in extending the results of this paper to other channel models, e.g. log-normal shadowing model [16]. In addition, investigating the asymptotic behavior of the giant component using Theorem 1 for one-dimensional networks will also be an interesting problem for our future work.

REFERENCES

- [1] P. Santi and D. M. Blough, "The Critical Transmitting Range for Connectivity in Sparse Wireless Ad Hoc Networks," *IEEE Transactions on Mobile Computing*, vol. 2, no. 1, pp. 25–39, 2003.
- [2] P. Gupta and P. Kumar, "Critical power for asymptotic connectivity in wireless networks," *Stochastic Analysis, Control, Optimization and Applications*, pp. 547–566, 1998.
- [3] A. Tang, C. Florens, and S. H. Low, "An Empirical Study on the Connectivity of Ad Hoc Networks," in *IEEE Aerospace Conference*, vol. 3, March, 2003, pp. 1333–1338.
- [4] R. Hekmat and P. V. Mieghem, "Connectivity in Wireless Ad-hoc Networks with a Log-normal Radio Model," *Mobile Networks and Applications*, vol. 11, no. 3, pp. 351–360, June, 2006.
- [5] M. Penrose, "On k-connectivity for a geometric random graph," *Random Structures and Algorithms*, no. 15, pp. 145–164, 1999.
- [6] C. Bettstetter, "On the Minimum Node Degree and Connectivity of a Wireless Multihop Network," in *3rd ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Lausanne, 2002, pp. 80–91.
- [7] U. N. Raghavan, H. P. Thadakamalla, and S. Kumara, "Phase Transition and Connectivity in Distributed Wireless Sensor Networks," in *13th International Conference on Advanced Computing and Communications*, Coimbatore, India, December, 2005.
- [8] D. M. Blough, M. Leoncini, G. Resta, and P. Santi, "The k-Neighbors Approach to Interference Bounded and Symmetric Topology Control in Ad Hoc Networks," *IEEE Transactions on Mobile Computing*, vol. 5, no. 9, pp. 1267–1282, 2006.
- [9] M. Desai and D. Manjunath, "On the Connectivity in Finite Ad Hoc Networks," *IEEE Communications Letters*, vol. 6, no. 10, pp. 437–439, October, 2002.
- [10] B. Bollobas, *Random Graphs*. Academic Press, 1985.
- [11] G. Németh and G. Vattay, "Giant Clusters in Random Ad Hoc Networks," *Physical Review E*, vol. 67, no. 3, March, 2003.
- [12] E. Godehardt and J. Jaworski, "On the Connectivity of a Random Interval Graph," *Random Structures and Algorithms*, vol. 9, no. 1 and 2, pp. 137–161, 1996.
- [13] P. Santi, D. M. Blough, and F. Vainstein, "A Probabilistic Analysis for the Range assignment Problem in Ad Hoc Networks," in *MOBIHOC*, Long Beach, CA, 2001.
- [14] Y. Xu, J. Heidemann, and D. Estrin, "Geography-Informed Energy Conservation for Ad Hoc Routing," in *ACM Mobicom'01*, 2001, pp. 70–84.
- [15] C. H. Foh and B. S. Lee, "A closed form network connectivity formula for one-dimensional MANETs," in *IEEE International Conference on Communications (ICC)*, vol. 6, Paris, France, 2004, pp. 3739–3742.
- [16] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 1st ed. New Jersey: Prentice Hall PTR, 1996.