Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 482727, 11 pages
http://dx.doi.org/10.1155/2013/482727

Research Article

The Influence of Communication Range on Connectivity for Resilient Wireless Sensor Networks Using a Probabilistic Approach

Yuanjiang Huang, José-Fernán Martínez, Juana Sendra, and Lourdes López

Centro de Investigación en Tecnologías Software y Sistemas Multimedia para la Sostenibilidad (CITSEM), Campus Sur Universidad Politécnica de Madrid (UPM), 28031 Madrid, Spain

Correspondence should be addressed to Yuanjiang Huang; yuanjiang@diatel.upm.es

Received 12 June 2013; Revised 6 August 2013; Accepted 8 August 2013

Academic Editor: Shengming Jiang

Wireless sensor networks (WSNs) consist of thousands of nodes that need to communicate with each other. However, it is possible that some nodes are isolated from other nodes due to limited communication range. This paper focuses on the influence of communication range on the probability that all nodes are connected under two conditions, respectively: (1) all nodes have the same communication range, and (2) communication range of each node is a random variable. In the former case, this work proves that, for $0 < \varepsilon < e^{-1}$, if the probability of the network being connected is 0.36ε, by means of increasing communication range by constant $C(\varepsilon)$, the probability of network being connected is at least $1 - \varepsilon$. Explicit function $C(\varepsilon)$ is given. It turns out that, once the network is connected, it also makes the WSNs resilient against nodes failure. In the latter case, this paper proposes that the network connection probability is modeled as Cox process. The change of network connection probability with respect to distribution parameters and resilience performance is presented. Finally, a method to decide the distribution parameters of node communication range in order to satisfy a given network connection probability is developed.

1. Introduction

Wireless sensor networks (WSNs) [1, 2] are a promising technology nowadays. The use of WSNs in numerous applications, such as forest monitoring, disaster management, space exploration, factory automation, secure installation, border protection, and battlefield surveillance, is emerging. WSNs technology is the basis of future network "Internet of Things" (IoT) [3], which offers a vision where anyone can interact with any addressable nodes (things or objects)—such as RFID tags, sensors, and mobile phones—anywhere and anytime. "Anywhere" suggests that any object is reachable from any location. From the network topology point of view, every node in WSNs should be able to, directly or through limited number of intermediate nodes, connect to any other nodes. This kind of network is called "connected network." If the network is still connected after removing at most $k-1$ nodes, it is called k-connected network, where $k = 1, 2, 3, \ldots$. A k-connected network guarantees that at least k different paths are available for transmitting signals from one node to any other nodes.

However, k-connected network is not always possible. In WSNs, sensor nodes are usually deployed in the areas of interest either randomly or according to a predefined distribution. In this case, it is likely that some nodes are isolated from other nodes. Therefore, the network connection is characterized by probability. On the other hand, the resilient problem, which indicates fault-tolerance capability in the presence of node failure, is also important in the probabilistic network. Our concern in this paper is the probability that the WSNs are a connected network and network resilience against the node failures.

Most of earlier studies focus on the model where each node in a network is the same and, for example, has the same communication range. However, WSNs nodes are usually heterogeneous. The communication range of the WSNs node

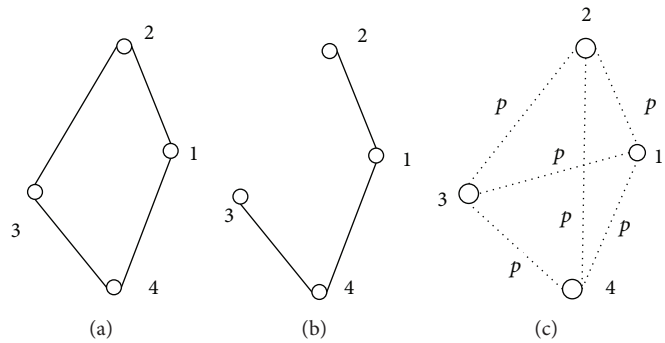


Figure 1: Different methods to connect nodes: (a) node connects 2 nearest neighbors; (b) node connects other nodes within its communication range; (c) node connects other nodes with same probability P.

may vary from one node to another, and even communication range of the same node may change over time. For instance, in a wireless network, the transmission power required for a node to reach another node is proportional to R^{α}, where R is the transmission radius and α is the loss constant depending on the wireless medium of which typical value is between 2 and 4 and may vary from devices to devices [4]. According to various wireless communication technologies, communication range may vary from tens to thousands meters, such as IEEE 802.11 (25–600 m), Bluetooth (10–100 m), ZigBee (10–75 m), HomeRF (50 m), UWB (10 m), and WiMAX (1–50 km). Depending on how long nodes work, residual energy of battery powered devices decreases over time, so a node may try to shorten communication range in order to save energy. Environments where nodes are deployed, for example, indoor or outdoor, with or without obstacle, result in communication range quite different due to the interference, shadowing, fading, and pass loss [5].

This work concentrates on WSNs connection probability for both heterogenous and homogenous networks in terms of communication range. Assuming WSNs nodes are randomly and uniformly distributed, two problems are addressed in this paper: given a network where all nodes have the same communication range, how does the connection probability change as communication range increases? In the case that communication range is a random variable, what is the network connection probability?

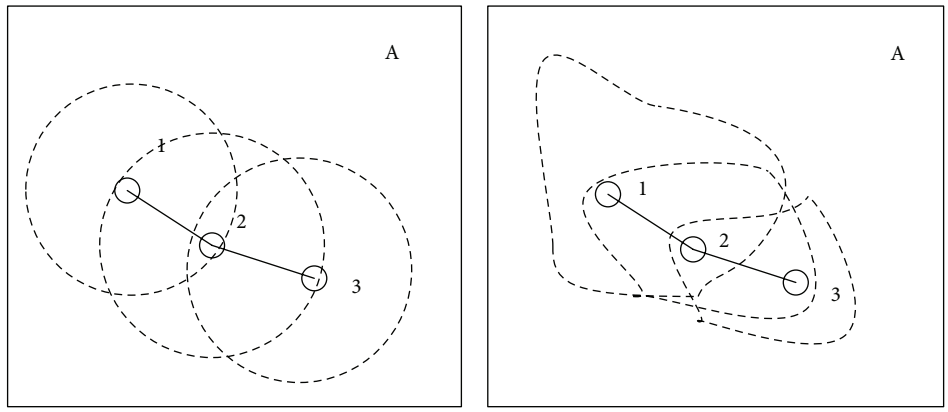
Through analysis, this work finds that for $0 < \varepsilon < e^{-1}$ and the number of nodes in the network is big enough and if the original network connection probability is 0.36ε, through increasing the communication range by constant $C(\varepsilon)$, the probability of a network being connected increases from 0.36ε to $1 - \varepsilon$. Explicit function $C(\varepsilon)$ is given in this paper. It turns out that, when a network is connected, it is also almost sure $\log(n) + b$-connected (where n is the total number of nodes deployed and b is a constant greater than 1), which is important for the WSNs resilient against the node failure. Afterwards, the connection probability problem with random communication range, which is often the real case in the WSNs, is studied. The model is reformulated as Cox process, and the connection probability is analyzed by simulation. A method for determining the distribution function parameters for a given connection probability is developed.

Our main contributions are as follows: first, this paper employs an effective and novel approach to obtain analytical results for homogenous WSNs connectivity, some of which have been validated by previous studies; second, we propose that the Cox process can be used to model heterogenous WSNs and the simulations are performed to reveal the relations between the network connection probability and its distribution parameters.

The rest of this paper is organized as follows. Section 2 introduces the basic concepts of network model and the problem to be addressed. In Section 3, derivation and verification in case that the network nodes have the same communication range are presented. In Section 4, communication range is modeled as a random variable. A brief introduction of related works is provided in Section 5 while Section 6 concludes our work.

2. Network Model and Problem Statement

Usually, there are three methods to create links between nodes, as presented in Figure 1. One is *k-nearest neighbor model*. In this model, the network is formed by each node connecting to k-nearest neighbors; for example, in Figure 1(a), each node has 2 neighbors. The second is *disc model*. Node is modeled as a disk with communication radius r. The node s is linked to node u if the Euclidean distance between s and u is less than r; for example, in Figure 1(b), node 3 cannot connect to node 2 and node 1 because they are out of communication range of node 3. The last one is *Erdös-Rényi random graph* that connects any two nodes by the same probability which is inappropriate in the WSNs; for example, in Figure 1(c), each node connects other nodes with the same probability P. The *k-nearest neighbor model* can be achieved by changing communication range of each node until the number of neighbors reaches k. *Disc model*, on the other hand, connects those nodes that fall into its communication range. *k-nearest neighbor model* and *disc model* are different. *k-nearest neighbor model* makes sure that there is no isolated node, but *disc model* is characterized by the probability that a network does not have isolated nodes. *Disc model* is more plausible in the WSNs in the case that obtaining k neighbors is not always feasible. For instance, in wireless environment, some nodes may be unable to connect



(a) (b)

Figure 2: (a) Disk communication model and (b) irregular communication model.

to a required number of neighbors due to the communication range limitation.

The notations and basic network definitions that will be used throughout the paper are now introduced. Additional terminologies are referred to [6]:

n: total number of nodes deployed in target field, and $n \gg 1$,

A: area of node deployed,

ρ: node density, defined as n/A,

t: expected number of neighbors of node.

Note that in this paper "log" means the logarithm to nature base e. Next, main definitions are introduced.

Definition 1. Node's communication range is defined as the area where other nodes can receive its signal.

For a disk, the communication range is the circle with radius r. However, communication range is not necessary modeled as a disk. The communication range of radio is highly probabilistic and irregular [7, 8]. Figures 2(a) and 2(b) illustrate the ideal disk communication and irregular communication model, respectively. More importantly, the communication range of each node may not be the same. Note that the analysis in this section is a disk, but it can also apply to the irregular communication model.

Definition 2. $S_{n,r}$ denotes a network following *disc model*. More specifically, the network is formed by n nodes randomly and uniformly deployed in area A. The node is modeled as a disk with radius r.

This paper focuses on the probability of network $S_{n,r}$ being k-connected. A k-connected network implies that there are still $k-1$ alternative path(s) if one path failed, therefore a higher k indicates that the network is more resilient against failures. In this paper, k is used to evaluate the WSNs resilience. This property depends on many factors, such as communication range, node density ρ, node processing capability, node energy, and deployment environment. This paper is only interested in the impact of communication range on the connection probability. The problem can be stated as follows.

"Given WSNs $S_{n,r}$ with fixed node density ρ, in the cases in which node communication range is the same and different, how network connection probability and resilience performance change as node communication ranges vary?"

3. Homogenous Node Deployment in WSNs

This section considers that, in the network $S_{n,r}$, each node has the same communication range. First, the mathematical model that will be used is presented. Based on this model, theoretical results are proved and validated by an example and simulations. In Section 4, the situation where communication range of each node is a random variable will be discussed.

3.1. Network Connection Probability Analysis. For uniformly distributed nodes with density ρ, the number of nodes in the area πr^2 has a *Poisson* distribution [9]; therefore the probability of a node having N neighbor nodes is

$$P(N) = \frac{\left(\rho\pi r^2\right)^N}{N!} e^{-\rho\pi r^2}. \tag{1}$$

Number of node neighbor is also called the node's degree. The minimal degree of all nodes is called the network degree. If the network has n nodes, the probability of network $S_{n,r}$ is k-connected given by following well-known formula [9]:

$$P(k) = \left(1 - \sum_{N=0}^{k-1} \frac{\left(\rho\pi r^2\right)^N}{N!} e^{-\rho\pi r^2}\right)^n. \tag{2}$$

Let

$$t = \pi r^2. \tag{3}$$

Note that t indicates the communication range of a node, but, if $\rho = 1$, t actually is the expected number of neighbors a node has.

Without loss of generality, assume that $\rho = 1$. For a real network the density of node $\rho = 1$ indicates that the average number of nodes in unit area is one. However, whether ρ is equal to 1 is irrelevant in this model, because if ρ is not 1, say ρ', then letting $r' = r/\sqrt{\rho'}$ the results will be the same. In order to simplify denotation, define

$$W(t,k) = \sum_{N=0}^{k-1} \frac{(t)^N}{N!} e^{-t}. \tag{4}$$

So, (2) can be rewritten as

$$P(t,k) = (1 - W(t,k))^n, \tag{5}$$

and its first derivation with respect to t is

$$\begin{aligned} \frac{dP(t,k)}{dt} &= -n(1 - W(t,k))^{n-1} \frac{dW(t,k)}{dt} \\ &= n\left(\frac{e^t - \sum_{N=0}^{k-1} \left(t^N/N!\right)}{e^t} \right)^{n-1} \left(e^{-t} \frac{t^{k-1}}{(k-1)!} \right). \end{aligned} \tag{6}$$

The function $P(t,k)$ can be written as $P(t) = (1 - e^{-t})^n$ when $k = 1$. In this section, the properties of $P(t)$ are analyzed, namely, 1-connected network. Two points are found out where $P(t)$ almost starts and stops growing in order to show that the connection probability increases from near 0 to reach 1.

Proposition 3. *Letting $P(t) = (1 - e^{-t})^n$ and $n \gg 1$, then following statements hold:*

(1) *for every $0 < \varepsilon < (1 - (1/n))^{(n-1)}$, there exists $0 < t_1 < t_2$ such that $P'(t_1) = P'(t_2) = \varepsilon$;*

(2) *$P(t)$ has a flex point at $(\log(n), (1 - (1/n))^n)$.*

Proof. (1) First, (5) is a monotonically increasing function for any $k \geq 1$ (note that $e^t = \sum_{N=0}^{+\infty} (t^N/N!)$ is the Taylor expansion of e^t, and $t \geq 0$); see Figure 3(a).

Now consider the first and second derivative functions of $P(t)$:

$$\begin{aligned} P'(t) &= n\left(1 - e^{-t}\right)^{n-1} e^{-t}, \\ P''(t) &= ne^{-t}\left(1 - e^{-t}\right)^{n-2} \left(ne^{-t} - 1\right). \end{aligned} \tag{7}$$

It is evident that $P''(t)$ vanishes at $t = 0$ and $t = \log(n)$. Furthermore, $P''(t) > 0$ for $t \in (0, \log(n))$ and $P''(t) < 0$ for $t \in (\log(n), \infty)$. Therefore, $P'(t)$ is an increasing function in the interval $(0, \log(n))$ and a decreasing function in $(\log(n), \infty)$. Hence, $P'(t)$ reaches the maximum value at $t = \log(n)$ (see Figure 3(a)). On the other hand, since $\lim_{t \to \infty} P'(t) = 0$, by applying *Bolzano* Theorem, for every $0 < \varepsilon < P'(\log(n)) = (1 - (1/n))^{(n-1)}$, there exists $0 < t_1 < t_2$, such that $P'(t_1) = P'(t_2) = \varepsilon$ (Figure 3(b)).

(2) It is derived from the proof of statement above. □

Remark 4. Note that $0 < \varepsilon < (1 - (1/n))^{(n-1)} < 1$ since $n \gg 1$. In Figure 3(b), $\varepsilon = (1 - (1/500))^{(500-1)} \approx 0.368$.

The proof of the previous proposition can be applied to obtain the following result.

Theorem 5. *Let $P(t) = (1 - e^{-t})^n$, $n \gg 1$, $b > 1$, and $\varepsilon = e^{-b}$. Then there exists a constant number of neighbors, $C(\varepsilon) = \log((1 - \log\varepsilon)/\varepsilon)$, for which the network becomes connected with probability increasing from $P(t_1 = \log(n/(b+1)))$ to $P(t_2 = \log(n) + b)$.*

Proof. First, it can be observed that $\varepsilon = e^{-b}$ satisfies the hypothesis in Proposition 3(1) since $\varepsilon = e^{-b} < e^{-1}$; therefore $0 < \varepsilon < (1 - (1/n))^{(n-1)}$. Let $x = e^{-t}$ and define $f(x)$ as

$$f(x) = n(1-x)^{n-1}x - \varepsilon. \tag{8}$$

Then, the goal is to find the roots of (8). For this purpose, consider the derivative function:

$$\begin{aligned} f'(x) &= n(1-x)^{n-1} - n(n-1)(1-x)^{n-2}x \\ &= n(1-x)^{n-2}(1 - nx). \end{aligned} \tag{9}$$

Note that $f(x)$ is the function $P'(t) - \varepsilon$ under the change of variable $x = e^{-t}$, and by applying the proof of Proposition 3 there exist only two roots x_1 and x_2 of $f(x)$ in the interval $(0, 1)$. *Newton* method can be used to find out the approximation of roots x_1 and x_2. However, its accuracy depends on the initial value, which should be close enough to the real root. Letting x_0 be the initial value, according to (8) and (9), yields

$$\begin{aligned} x &\approx x_0 - \frac{f(x_0)}{f'(x_0)} \\ &= x_0 - \frac{n(1-x_0)^{n-1}x_0 - \varepsilon}{n(1-x_0)^{n-2}(1 - nx_0)} \\ &= x_0 - \frac{n(1-x_0)x_0 - \left(\varepsilon/(1-x_0)^{n-2}\right)}{n(1 - nx_0)} \\ &= x_0 - \frac{(1-x_0)x_0}{1 - nx_0} + \frac{\varepsilon/(1-x_0)^{n-2}}{n(1 - nx_0)}. \end{aligned} \tag{10}$$

Additionally, the inequality $0 < x_2 < (1/n) < x_1 < 1$ holds from the proof of Proposition 3; then *Newton* method can be applied. Let $x_0 = 0$ as the initial value to approximate x_2 and $x_0 = (b/n)$ (where $b > 1$) as the initial value to find x_1:

$$\begin{aligned} x_2 &= \frac{\varepsilon}{n} = \frac{e^{-b}}{n}, \\ x_1 &= \frac{b}{n} - \frac{(1 - (b/n))(b/n)}{1-b} + \frac{\varepsilon/(1-(b/n))^{n-2}}{n(1-b)} \\ &= \frac{1}{n}\left(\frac{\left(b^2/n\right) - b^2 + \left(\varepsilon/(1-(b/n))^{n-2}\right)}{1-b} \right). \end{aligned} \tag{11}$$

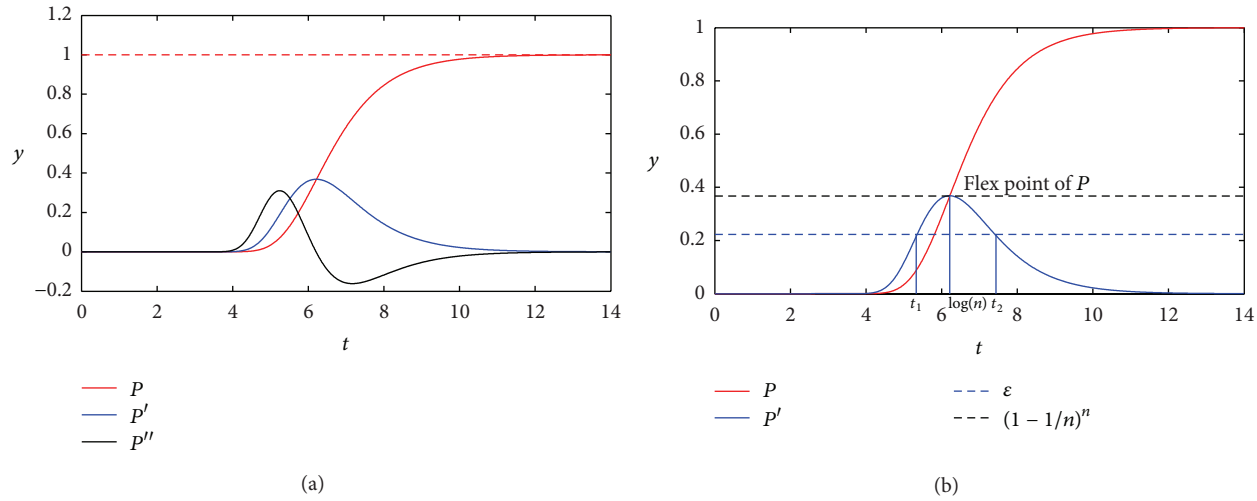


FIGURE 3: Network connection probability function $y = P(t)$ and corresponding $P'(t)$ and $P''(t)$ (a); values t_1 and t_2 for a given ε and $n = 500$ (b).

Taking into account that we have $(1 - (b/n))^{n-2} \to e^{-b}$ when $n \to +\infty$ and $\varepsilon = e^{-b}$, therefore

$$\frac{\left(b^2/n\right) - b^2 + \left(\varepsilon/(1-(b/n))^{n-2}\right)}{1-b} \longrightarrow 1 + b. \tag{12}$$

Note that $b = -\log(\varepsilon)$. Letting $t_2 = -\log x_2$, $t_1 = -\log x_1$, and defining $C(\varepsilon) = t_2 - t_1$, according to (12), we obtain

$$\begin{aligned} &C(\varepsilon) \\ &= \log\left(\frac{x_1}{x_2}\right) \\ &= \log\left(\frac{(1/n)\left(\left(\left(b^2/n\right) - b^2 + \left(\varepsilon/(1-(b/n))^{n-2}\right)\right)/(1-b)\right)}{\varepsilon/n}\right) \\ &\longrightarrow \log\left(\frac{1-\log(\varepsilon)}{\varepsilon}\right). \end{aligned} \tag{13}$$

Finally, taking into account that $t_2 = \log(n) + b$, we have $t_1 = t_2 - C(\varepsilon) = \log(n/(b+1))$. Since $P(t)$ is an increasing function, we conclude that the network becomes connected with probability increasing from $P(t_1 = \log(n/(b+1)))$ to $P(t_2 = \log(n) + b)$. □

Theorem 6. *Letting $\varepsilon = e^{-b}$ and $b > 1$, if $\pi r^2 = \log n + b$, then network connection probability $P(S_{n,r}$ connected$)$ is at least $1 - \varepsilon$ when $n \to +\infty$.*

Proof. Consider

$$P\left(S_{n,r} \text{ connected}\right) = \left(1 - \frac{e^{-b}}{n}\right)^n \geq 1 - n\frac{e^{-b}}{n} = 1 - \varepsilon. \tag{14}$$

□

Remark 7. This theorem shows that, as $b \to +\infty$, network connection probability tends to 1 and leads to the network that has degree $\log n + b$. The author in [10] proves that if a network does not have any links at the beginning, and later links are added to connect nodes, the resulting network becomes k-connected as soon as network degree is k. Therefore, this theorem shows that once network becomes connected, it turns out to be $\log n + b$-connected with high probability. This conclusion is consistent with the result in [11]: by increasing k network becomes s-connected very shortly after it becomes connected, for $s = O(\log n)$. $\log n + b$-connected network makes WSNs more resilient against node failure because there are $\log n + b$ distinct paths from one node to any other nodes.

Theorem 8. *Letting $\varepsilon = e^{-b}$ and $b > 1$, if $\pi r^2 = \log(n/(b+1))$, then the network connection probability $P(S_{n,r}$ connected$)$ is about 0.36ε when $n \to +\infty$.*

Proof. Consider

$$\begin{aligned} P\left(S_{n,r} \text{ connected}\right) &= \left(1 - \frac{b+1}{n}\right)^n \\ &\longrightarrow e^{-b-1} = 0.36e^{-b} = 0.36\varepsilon. \end{aligned} \tag{15}$$

□

Theorem 9. *Letting $\varepsilon = e^{-b}$ and $b > 1$, if $\pi r^2 = \log n + b$, then the network connection probability is $e^{-e^{-b}}$, when $n \to +\infty$.*

Proof. Consider

$$P\left(S_{n,r} \text{ connected}\right) = \left(1 - \frac{e^{-b}}{n}\right)^n \longrightarrow e^{-e^{-b}}. \tag{16}$$

□

Remark 10. This conclusion is the same as [12] and has similar form in the Erdös-Rényi random graph [13].

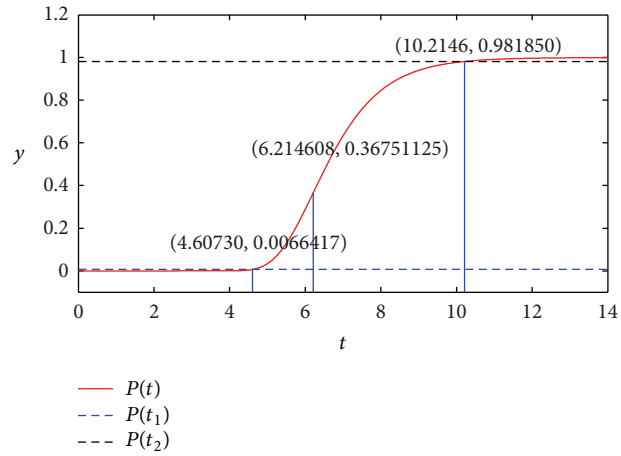


Figure 4: Network connection probability $y = P(t)$ increases from $P(t_1)$ to $P(t_2)$ for $n = 500$.

Corollary 11. *Letting* $0 < \varepsilon < e^{-1}$, *if the probability of network being connected is* 0.36ε *when communication range is* πr^2, *then, by increasing node communication range by constant* $C(\varepsilon) = \log((1 - \log\varepsilon)/\varepsilon)$, *namely,* $\pi r^2 + C(\varepsilon)$, *the probability of network being connected is at least* $1 - \varepsilon$.

Proof. It is obvious from the previous Theorems 5, 6, and 8. □

This section addresses one question. If a node current communication range is known, then the connection probability can be calculated by using (2). If the network connection probability is very low, maybe one wants to increase the node communication range to obtain a higher network connection probability. Equation (2) can be used again to calculate the required communication range, but surprisingly the corollary proved in this section shows that the incremental of communication range to obtain a high connection probability is a constant for any size of network.

3.2. Validation Results. This section validates the previous results by an example and simulations. In the example, 500 nodes with equal communication range are deployed in the field with $\sqrt{500} \cdot \sqrt{500}\, m^2$.

Example 12. The function $P(t) = (1-e^{-t})^n$ for $n = 500$ and $b = 4$ is studied. According to Theorem 5, there exists a constant number of neighbors $C(\varepsilon) = \log((4 + 1)/e^{-4}) = 5.60944$, for which the network becomes connected with probability increasing from $P(t_1) = 0.65705\%$ to $P(t_2) = 98.18507\%$ (as depicted in Figure 4).

First, it is observed that

$$\varepsilon = e^{-4} = 0.0183 < \left(1 - \frac{1}{500}\right)^{(500-1)} = 0.3682 \qquad (17)$$

which satisfies the hypothesis in Proposition 3(1). In our approach, the Newton's method is used to approximate the roots of

$$f(x) = 500(1 - x)^{500-1}x - 0.0183, \qquad (18)$$

obtaining

$$x_1 = 9.98 \cdot 10^{-3}, \qquad x_2 = 0.04 \cdot 10^{-3}. \qquad (19)$$

Observe that $0 < x_2 < 1/500 = 0.002 < x_1$. Hence,

$$t_2 = -\log(x_2) = 10.21461, \qquad t_1 = -\log(x_1) = 4.60517, \qquad (20)$$

which indicates $P(t_1 = 4.60517) = 0.65705\%$, $P(t_2 = 10.21461) = 98.18507\%$, and $C(\varepsilon) = t_2 - t_1 = 5.60944$.

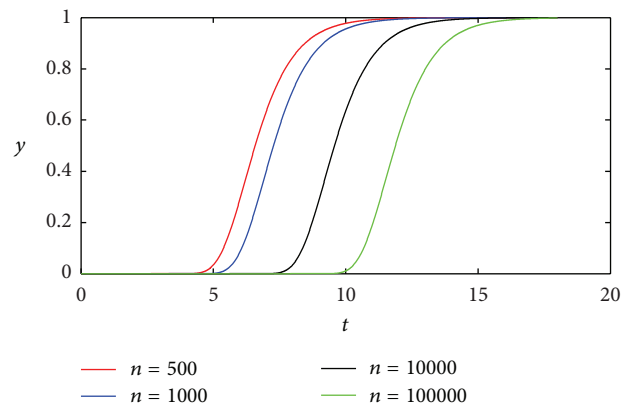


Figure 5: Network connection probability $y = P(t)$ when $n = 500, 1000, 10000$, and 100000.

Table 1: Connection probability with different sizes of network.

n	t_1	t_2	$t_2 - t_1$	$P(t_1)$	$P(t_2)$
500	4.60517	10.21461	5.60944	0.65705%	98.18507%
1000	5.29832	10.90776	5.60944	0.66540%	98.18509%
10000	7.60090	13.21034	5.60944	0.67295%	98.18511%
100000	9.90349	15.51293	5.60944	0.67371%	98.18511%

Figure 5 shows connection probability $y = P(t)$ when $n = 500, 1000, 10000, 100000$. Table 1 demonstrates the values of x_1, x_2, t_1, t_2, and $C(\varepsilon)$ and corresponding values of $P(t_1)$ and $P(t_2)$, for $n = 500, 1000, 10000, 100000$. For any n in the table, the obtained value $t_2 - t_1 \approx C(\varepsilon) = 5.60944$. Of course, in a real network, the number of neighbors is integer, so 6 neighbors are needed. This example implies that, regardless of network size (number of nodes should be big enough), if the network connection probability is 0.66%, by increasing the communication range until each node obtains 6 more neighbors (namely, increasing communication range by 6 m^2), the network connection probability reaches at least 98.17%. Meanwhile, the network will be at least 10-connected.

In order to validate Theorems 6 and 8, this paper calculates the error between theoretical results and approximation values with different , n and b, as shown in Figure 6. The error of Theorem 6 is defined as $(1 - (e^{-b}/n))^n - (1 - e^{-b})$, and the error of Theorem 8 is defined as $(1 - ((b + 1)/n))^n - 0.36e^{-b}$. The errors for both theorems are very small, which indicate that both have a good approximation.

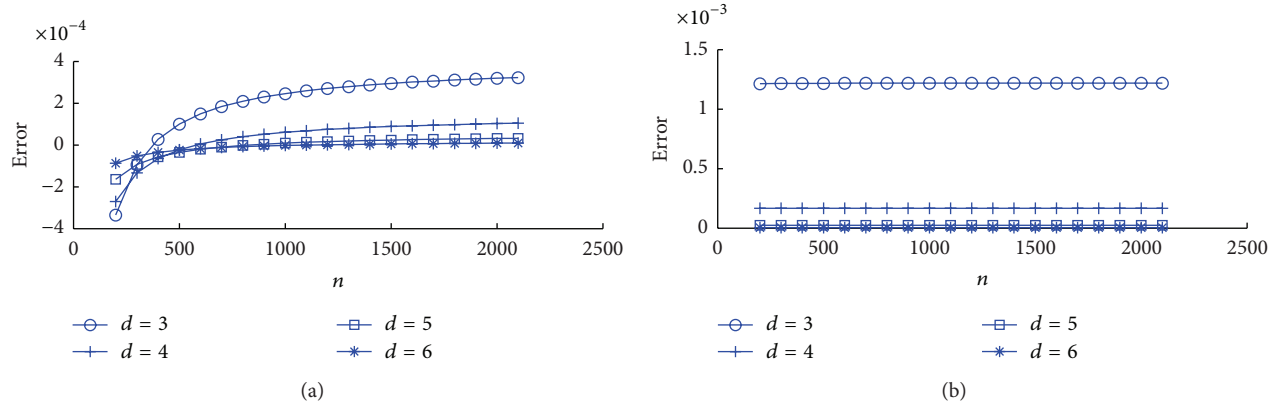


FIGURE 6: (a) The errors of Theorem 6; (b) the errors of Theorem 8.

4. Heterogenous Node Deployment in WSNs

In the last section, the obtained asymptotic results were based on the assumption that each node has the same communication range which is often not the case in practice. This section presents the connection probability when node communication range follows a normal distribution, that is, $t \sim N(\mu, \sigma^2)$.

Formally, network model is reformulated as follows: n nodes are randomly and uniformly deployed in area A with density $\rho = 1$. Communication range of node i, denoted as t_i, is i.i.d random variable and has normal distribution $t_i \sim N(\mu, \sigma^2)$. Hence, the number of neighbors of the node, denoted as k_t, is the Poisson random variable condition on parameter t, where $t \sim N(\mu, \sigma^2)$. This model is analog to the so-called Cox process in which random variable is Poisson process where density itself is a stochastic process. Cox process is widely used in economics, for example, [14].

4.1. Connection Probability for Random Communication Range. $E[V]$ denotes the expected value of a random variable V; therefore the expected neighbors of node are

$$E[k_t] = E[E[k_t \mid t]] = E[t] = \mu. \tag{21}$$

In what follows, connection probability itself is researched. For $k \geq 1$, probability of node i having at least k neighbors is given by

$$P_{i,k} = g_i(t_i) = 1 - \sum_{N=0}^{k-1} \frac{t_i^N}{N!} e^{-t_i}. \tag{22}$$

For $k = 1$, $P_{i,1} = 1 - (1/e^{t_i})$ is the probability that node i is not isolated. e^{t_i} has log normal distribution. Therefore, the expected value $E[P_{i,1}]$ and variance $\mathrm{Var}[P_{i,1}]$ can be obtained via standard method:

$$\begin{aligned} E[P_{i,1}] &= 1 - e^{-\mu + (1/2)\sigma^2}, \\ \mathrm{Var}[P_{i,1}] &= \left(e^{\sigma^2} - 1\right)\left(E[P_{i,1}]\right)^2. \end{aligned} \tag{23}$$

For $k > 1$ neighbors, the distribution of $P_{i,k}$ does not have a closed-form expression.

If n is big enough, the probability of network being k-connected is

$$P_k = \prod_{i=1}^{n} P_{i,k}. \tag{24}$$

Since parameter t is a random variable, P_k is a random variable as well. Letting $P_{\min} = \min\{P_{1,k}, P_{2,k}, \ldots, P_{n,k}\}$, because $0 \leq P_{i,k} \leq 1$, so

$$P_{\min}^n \leq P_k \leq P_{\min}. \tag{25}$$

Therefore the obstruction of connection probability of entire network is the node which has the minimal communication range.

P_k is affected by several parameters: k, n, μ, and σ. Theorem 6 is used to decide μ. According to Theorem 6, the probability of the network being connected is at least 99.33% when $b = 5$. Let $\rho = 1$ and take $\log n + 5$ as average μ of communication range; for instance, if $n = 500$, then $\mu = 11.2$. In other words, 500 nodes with node communication range following normal distribution $t \sim N(11.2, \sigma^2)$ are deployed.

Our major concerns are the parameter σ which indicats communication range difference and k which shows the resilience capability. In order to study the changes of connection probability P_k as parameters vary, the following simulations are performed: (1) cumulative distribution function (CDF) of P_k is calculated after 500 runs with various σ and k, as shown in Figures 7 and 8; (2) given μ and σ, what is the probability of network being k-connected as the number of nodes deployed grows? This is done by computing average of $P_{n,k}$ after 500 runs for a given number of nodes, as illustrated in Figures 9 and 10; (3) how to choose the parameters in order to get the required connection probability. This is discussed in Section 4.3.

Figures 7 and 8 show the CDF of $P_{n,k}$ when σ and k change. The network probability is sensitive to standard deviation. As mentioned earlier, a single node that has small communication range can cause the whole network connection probability to be low. For instance, in Figure 8 when $\sigma = 3$ and $k = 2$, the probability of network being connected is almost sure less than 40%.

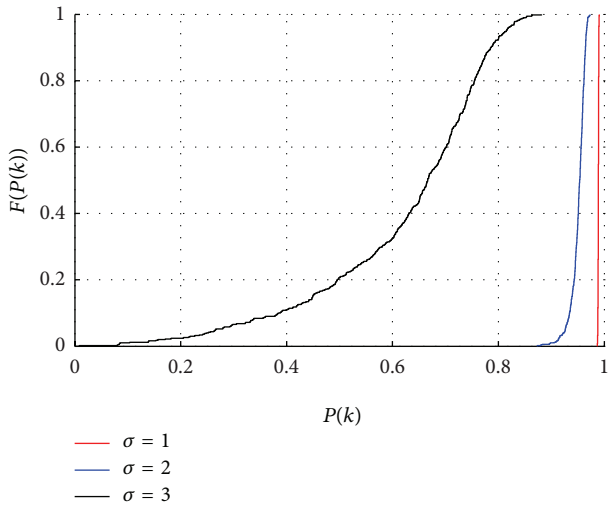


Figure 7: Cumulative distribution function (CDF) of P_k when $\sigma = 1, 2$, and 3, $n = 500$, and $k = 1$.

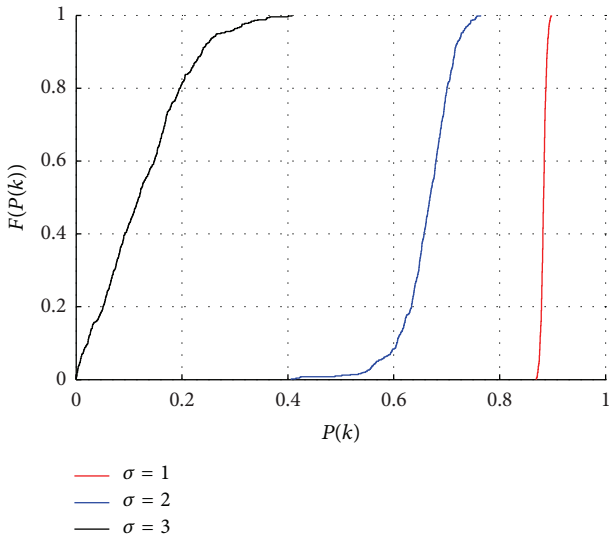


Figure 8: Cumulative distribution function (CDF) of P_k when $\sigma = 1, 2$, and 3, $n = 500$, and $k = 2$.

Figure 9 illustrates the connection probability as N nodes were deployed in network when $\sigma = 2$ and $k = 1, 2, 3, 4$. Figure 10 shows the changes when $\sigma = 1, 2, 3$ and $k = 1$. Both figures show that the average of P_k is the decreasing function of k, N, and σ. Network connection probability as network size growing is predictable. For instance, Figure 10 shows that the network average connection probability for $\sigma = 3$ is about 73% when the network has 250 nodes, but the probability falls to 55% if the network size is doubled. Figure 9 shows how the resilience performance decreases when network size grows or the probability decreases if higher resilience performance is required. For example, for networks which have 200 nodes, the probability that this network can tolerate 1, 2, and 3 (i.e., $k = 2, 3, 4$) nodes failure are about 83%, 50%, and 10%, respectively.

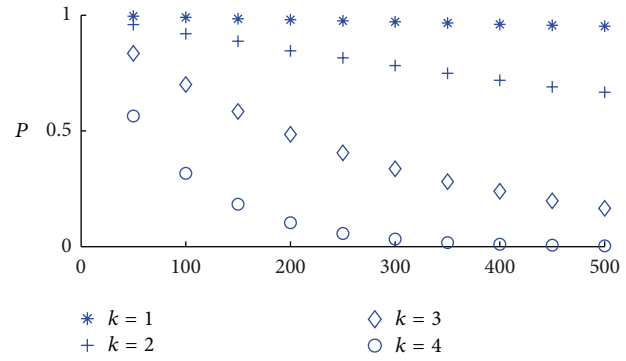


Figure 9: k-connected probability P when $\sigma = 2$ and $k = 1, 2, 3$, and 4.

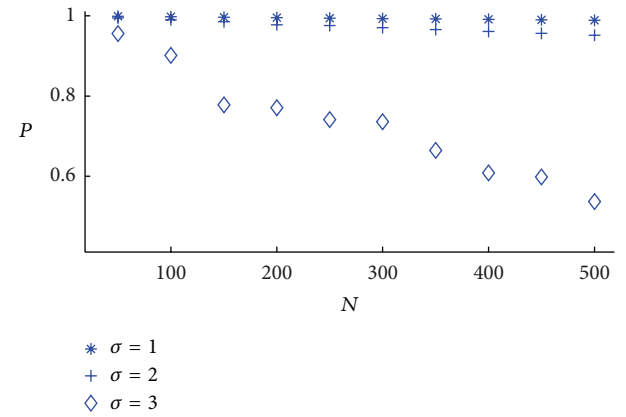


Figure 10: Connection probability P when $\sigma = 1, 2$, and 3 and $k = 1$.

4.2. Choose Distribution Parameter. The simulations in Section 4.2 show P_k with different parameters. In this section, it is addressed which distribution parameter(s) can maintain the given P_k. This is helpful to choose appropriate parameters when network simulator is used to simulate real networks. According to (6), (22) is a monotonically increasing function of $t_i \in [0, +\infty)$, and its inverse function is written as

$$t_{i,k} = g_i^{-1}(p_{i,k}). \tag{26}$$

Letting $p_{i,k}^{(0)}$ be an instance of $p_{i,k}$, thus $t_{i,k}^{(0)} = g_i^{-1}(p_{i,k}^{(0)})$. The probability $p_{i,k}$ being greater than $p_{i,k}^{(0)}$ is given by

$$\int_{t_{i,k}^{(0)}}^{+\infty} f_T(t)\, dt, \tag{27}$$

where $f_T(t)$ is the probability density function of t. If the probability of a network required to keep network k-connected is at least P_0, the corresponding probability for each node is at least

$$p_{\min} \geq P_0^{1/n}. \tag{28}$$

With formula (26)–(28), the required density function parameter of communication range for given P_0 can be calculated.

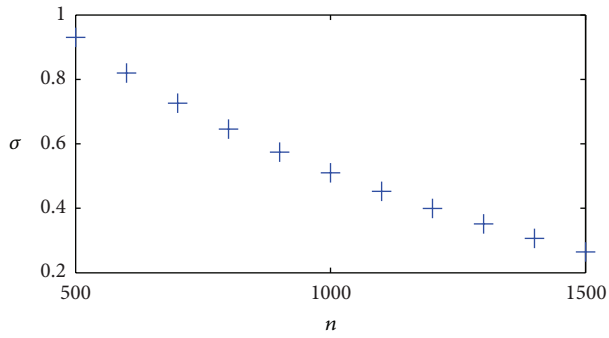


Figure 11: Required σ for 90% connection probability when $\mu = 10$.

For example, 500 nodes are deployed in $\sqrt{500} \cdot \sqrt{500}\,\mathrm{m}^2$; communication area is $t \sim N(\mu, \sigma^2)$ with mean $\mu = 10$. If the desired probability of the whole network being connected, that is, k-connected, is at least 90%. Standard deviation of this distribution is evaluated. For $k = 1$, according to (26) and (28), corresponding minimal range is $t_{i,1}^{(0)} = 8.46$. In order to make probability of $t_{i,1}$ greater than $t_{i,1}^{(0)}$ is high, for example, at least 95%, according to (27), $t_{i,1} = \mu - 1.65\sigma$. Therefore corresponding standard deviation σ should be no more than 0.93. This is useful in the case of using network simulator to choose appropriate parameters to design high probability connected networks. Figure 11 shows the required σ in order to make the network connection probability at least 90% when the number of nodes are different. Note that the node density is always 1.

5. Related Works

Extensive studies have been done on the connection problem of networks. Many of them focus on how many neighbors or network density is needed so that a network connects with high probability, such as [15]; some construct network to satisfy connectivity [16, 17]; some works try to develop algorithms to preserve network connectivity or coverage, for example, [18–20], while some other works study other aspects of network connectivity, such as [21] which evaluates the quality of connectivity by measuring the reliability of link; it shows that the largest eigenvalue of the probabilistic connectivity matrix can serve as a good measure of the quality of network connectivity. When all the nodes of a region fail, [22] measures the number of connected components. This paper studies the connection probability when the network nodes are randomly deployed.

When nodes are randomly deployed, asymptotic upper and lower bounds of connection probability for both *k-nearest neighbor* and *disk model* have been studied [12]. For *k-nearest neighbor*, [23] concludes that, as $n \to \infty$, if each node is connected to less than $0.074 \log n$ neighbors, the network is disconnected with probability one, while, if neighbors are more than $5.1774 \log n$, the network is connected with probability one. Reference [24] finds that if $k \leq 0.3043 \log n$, the network is not connected with high probability and if $k \geq 0.5139 \log n$, then network is connected with high probability as $n \to \infty$. But for the directed network the upper and lower bounds are $0.7209 \log n$ and $0.9967 \log n$, respectively. Reference [25] improves the upper bound to be $0.4125 \log n$. For *disk model* [26] states that 6 to 10 average numbers of neighbors almost make sure that network will be fully connected no matter how many nodes there are totally in the network. In [27], if communication range $\pi r^2 = \log n + b$, then the network connection probability tends to be $e^{-e^{-b}}$. Compared with [26], Table 1 in this paper shows that, when $n = 10000$, at least 13 neighbors are needed in order to make sure that network is connected with high probability. Besides, a result (Theorem 9) presented in our paper is the same as [27] but uses a totally different approach.

Reference [11] shows that, in *k-nearest neighbor model* by increasing k, network becomes s-connected very shortly after it becomes connected, where $s = O(\log n)$. Reference [28] proves one conjecture in [24] that, in *k-nearest neighbor model* for every $0 < \varepsilon < 1$ and n sufficiently large, there exists $C = C(\varepsilon)$ such that, if the network has k-connected probability ε, then $(k + C)$-connected probability is bigger than $1 - \varepsilon$. This paper improves the results in [11], obtaining an explicit expression for *disk model*, that is, $s = \log n + b$, where $b > 1$. The corollary in this paper proves that the result for *disc model* has a similar form presented in [28].

Nodes having the same communication range usually are not true in reality. In order to make the model more accurate, [8, 29] utilize irregular radio to model real nodes. The connectivity for heterogenous networks has been well studied; for example, [16, 30] investigate the relay node placement problem such that network is the k-connected. The authors in [31] assumes that node communication radius r_i of node i is i.i.d. random variable with normal probability density $r_i \sim N(\mu, \sigma^2)$. Reference [32] adopts the model that Poisson intensity is given by a normal distribution; then it obtains the asymptotic bound of range that all nodes in this area are connected to the origin. Reference [33] considers nodes are placed according to a shot-noise Cox process rather than uniform deployment. This paper employs the stochastic methods to characterize heterogenous network. In this paper the density is maintained constant, but the node communication range is normal distribution.

6. Conclusion and Future Works

When deploying many WSNs nodes, one of the key problems is whether all nodes in the network are connected to other nodes. Isolated nodes will be useless for applications. This paper presents the results on how the network connection probability changes as the communication range varies in randomly and uniformly distributed homogenous and heterogenous WSNs. In case of network with all nodes having the same communication range, through theory derivation and validation, this paper proves that, regardless of network size, the network connection probability increases from 0.36ε to $1 - \varepsilon$ by increasing constant communication range of each node. As the example shows in Section 3.2, regardless of network size, if the network connection probability is 0.66%, by increasing the communication range until each

node obtains 6 more neighbors, the network connection probability reaches at least 98.17%. On the other hand, this paper shows that, once network is connected, it also becomes $\log n + b$-connected with high probability, which makes the network resilient against node failures because there are $\log n + b$ alternative paths between any two distinct nodes.

In case each node communication range is i.i.d random variable which has normal distribution, this paper analyzes the connection probability by simulation. This paper shows that network connection probability is determined by the distribution parameters and the network size, especially sensitive to standard deviation σ. The reason is that the network connection probability is dependent on the node that has minimal communication range. It implies that it needs to take care of the node which has minimal communication range because it is the bottleneck of the whole network. The network will become disconnected if they fail. With the same configuration, the resilience capability decreases when network size grows. Besides, given the required connection probability, this paper develops one method to decide the distribution parameter of communication range. This method can be used to choose appropriate distribution parameter of communication range for network simulators or real deployments.

In some circumstances, a full connected network is impractical and not necessary. One would be more interested in the giant connected component which contains most nodes of entire network are connected. More specifically, the relation between the giant connected component and the communication range distribution is what is wanted to be learnt. It is a percolation problem with random communication range. Percolation occurs when a node belongs to infinite component with none-zero possibility. The critical intensity λ_c is defined as the minimum intensity in which percolation occurs. For *disk model*, the bound for critical intensity is known (e.g., [34]) but for variable radius is unknown. Therefore, studying the percolation problem with i.i.d communication range (or radius) will be our future work. On the other hand, the degree of the node obeys Poisson distribution in this paper. It has been found that many networks, such as the World Wide Web, the Internet, airplanes connection networks, some biological systems, and international ownership network, have power-law degree distribution with an exponent that ranges between 2 and 3 [35]. Our future work will center on connection probability with a more accurate model.

References

[1] A. A. Aziz, Y. A. Sekercioglu, P. Fitzpatrick, and M. Ivanovich, "A survey on distributed topology control techniques for extending the lifetime of battery powered wireless sensor networks," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 1, pp. 121–144, 2013.

[2] C. Zhu, C. Zheng, L. Shu, and G. Han, "A survey on coverage and connectivity issues in wireless sensor networks," *Journal of Network and Computer Applications*, vol. 35, no. 2, pp. 619–632, 2012.

[3] L. Atzori, A. Iera, and G. Morabito, "The internet of things: a survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[4] I. Saha, L. K. Sambasivan, S. K. Ghosh, and R. K. Patro, "Distributed fault-tolerant topology control in wireless multihop networks," *Wireless Networks*, vol. 16, no. 6, pp. 1511–1524, 2010.

[5] P. C. Pinto and M. Z. Win, "Communication in a poisson field of interferers—part I: interference distribution and error probability," *IEEE Transactions on Wireless Communications*, vol. 9, no. 7, pp. 2176–2186, 2010.

[6] Adrian Bondy and U. S. R. Murty, *Graph Theory*, Springer, Berlin, Germany, 2008.

[7] P. K. Chong and D. Kim, "Surface-level path loss modeling for sensor networks in flat and irregular terrain," *ACM Transactions on Sensor Networks*, vol. 9, no. 2, pp. 1–32, 2013.

[8] K. Zhu, "Ad hoc network deployment accommodating short and uncertain transmission range," in *Proceedings of the 9th ACM International Symposium on Mobility Management and Wireless Access (MobiWac '11)*, pp. 101–108, New York, NY, USA, November 2011.

[9] C. Bettstetter, "On the minimum node degree and connectivity of a wireless multihop network," in *Proceedings of the 3rd ACM International Symposium On Mobile Ad Hoc Networking & Computing (MOBIHOC '02)*, pp. 80–91, Lausanne, Switzerland, June 2002.

[10] M. D. Penrose, "On k-connectivity for a geometric random graph," *Random Structures and Algorithms*, vol. 15, no. 2, pp. 145–164, 1999.

[11] P. Balister, B. Bollobás, A. Sarkar, and M. Walters, "Highly connected random geometric graphs," *Discrete Applied Mathematics*, vol. 157, no. 2, pp. 309–320, 2009.

[12] P. Balister, A. Sarkar, and B. Bollobs, "Percolation, connectivity, coverage and colouring of random geometric graphs," in *Handbook of Large-Scale Random Networks*, pp. 117–142, Springer, Berlin, Germany, 2008.

[13] R. Van Der Hofstad, "Random graphs and complex networks," http://www.win.tue.nl/~rhofstad/NotesRGCN2010.pdf.

[14] S. Delattre, C. Y. Robert, and M. Rosenbaum, "Estimating the efficient price from the order flow: a Brownian Cox process approach," *Stochastic Processes and Their Applications*, vol. 123, no. 7, pp. 2603–2619, 2013.

[15] H. Cai, X. Jia, and M. Sha, "Critical sensor density for partial connectivity in large area wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 7, no. 4, article 35, 2011.

[16] X. Han, X. Cao, E. L. Lloyd, and C.-C. Shen, "Fault-tolerant relay node placement in heterogeneous wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 5, pp. 643–656, 2010.

[17] J. L. Bredin, E. D. Demaine, M. Hajiaghayi, and D. Rus, "Deploying sensor networks with guaranteed capacity and fault tolerance," in *Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC '05)*, pp. 309–319, Urbana-Champaign, III, USA, May 2005.

[18] J. Yu, S. Ren, S. Wan, D. Yu, and G. Wang, "A stochastic coverage scheduling algorithm in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 746501, 11 pages, 2012.

[19] Z. Mi, Y. Yang, and H. Ding, "Self-organized connectivity control and optimization subjected to dispersion of mobile ad hoc sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 672436, 15 pages, 2012.

[20] M. Bishop, M. Carvalho, R. Ford, and L. M. Mayron, "Resilience is more than availability," in *Proceedings of the New Security Paradigms Workshop (NSPW '11)*, pp. 95–104, Marin County, Calif, USA, September 2011.

[21] S. Dasgupta and G. Mao, "On the quality of wireless network connectivity," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '12)*, vol. 500, no. 505, pp. 3–7, December 2012.

[22] S. Banerjee, S. Shirazipourazad, P. Ghosh, and A. Sen, "Beyond connectivity—new metrics to evaluate robustness of networks," in *Proceedings of the IEEE 12th International Conference on High Performance Switching and Routing (HPSR '11)*, pp. 171–177, July 2011.

[23] F. Xue and P. R. Kumar, "The number of neighbors needed for connectivity of wireless networks," *Wireless Networks*, vol. 10, no. 2, pp. 169–181, 2004.

[24] P. Balister, B. Bollobás, A. Sarkar, and M. Walters, "Connectivity of random k-nearest-neighbour graphs," *Advances in Applied Probability*, vol. 37, no. 1, pp. 1–24, 2005.

[25] M. Walters, "Small components in k-nearest neighbour graphs," *Discrete Applied Mathematics*, vol. 160, no. 1314, pp. 2037–2047, 2012.

[26] J. Ni and S. A. G. Chandler, "Connectivity properties of a random radio network," *IEE Proceedings*, vol. 141, no. 4, pp. 289–296, 1994.

[27] M. D. Penrose, *Random Geometric Graphs*, Oxford University Press, Oxford, UK, 2003.

[28] V. Falgas-Ravry and M. Walters, "Sharpness in the k-nearest neighbors ranmdom geometric graph model," *Advances in Applied Probability*, vol. 44, no. 3, pp. 617–634, 2012.

[29] H. Chenji and R. Stoleru, "Mobile sensor network localization in harsh environments," in *Proceedings of the 6th IEEE International Conference on Distributed Computing in Sensor Systems*, pp. 244–257, Springer, Santa Barbara, Calif, USA, 2010.

[30] D. R. Dandekar and P. R. Deshmukh, "Relay node placement for multi-path connectivity in heterogeneous wireless sensor networks," *Procedia Technology*, vol. 4, pp. 732–736, 2012.

[31] Z.-H. Guan, L. Ding, and Z.-M. Kong, "Multi-radius geographical spatial networks: statistical characteristics and application to wireless sensor networks," *Physica A*, vol. 389, no. 1, pp. 198–204, 2010.

[32] P. Balister, B. Bollobas, A. Sarkar, and M. Walters, "Connectivity of a Gaussian network," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 3, no. 3, pp. 204–213, 2008.

[33] G. Alfano, M. Garetto, E. Leonardi, and V. Martina, "Capacity scaling of wireless networks with inhomogeneous node density: lower bounds," *IEEE/ACM Transactions on Networking*, vol. 18, no. 5, pp. 1624–1636, 2010.

[34] S. C. Ng, G. Mao, and B. D. O. Anderson, "Analytical bounds on the critical density for percolation in wireless multi-hop networks," in *Proceedings of the 54th Annual IEEE Global Telecommunications Conference Energizing Global Communications (GLOBECOM '11)*, pp. 1–6, Houston, Tex, USA, December 2011.

[35] S. Vitali, J. B. Glattfelder, and S. Battiston, "The network of Global corporate control," *PLoS One*, vol. 6, no. 10, Article ID e25995, 2011.