Wireless Netw (2009) 15:945–964
DOI 10.1007/s11276-009-0179-9

Connectivity properties of large-scale sensor networks

Hossein Pishro-Nik · Kevin Chan · Faramarz Fekri

Published online: 11 April 2009

Abstract In wireless sensor networks, both nodes and links are prone to failures. In this paper we study connectivity properties of large-scale wireless sensor networks and discuss their implicit effect on routing algorithms and network reliability. We assume a network model of n sensors which are distributed randomly over a field based on a given distribution function. The sensors may be unreliable with a probability distribution, which possibly depends on n and the location of sensors. Two active sensor nodes are connected with probability $p_e(n)$ if they are within communication range of each other. We prove a general result relating unreliable sensor networks to reliable networks. We investigate different graph theoretic properties of sensor networks such as k-connectivity and the existence of the giant component. While connectivity (i.e. $k = 1$) insures that all nodes can communicate with each other, k-connectivity for $k > 1$ is required for multi-path routing. We analyze the average shortest path of the k paths from a node in the sensing field back to a base station. It is found that the lengths of these multiple paths in a k-connected network are all close to the shortest path. These results are shown through graph theoretical derivations and are also verified through simulations.

Keywords Wireless sensor networks · Connectivity · Multi-path routing · Unreliable sensors

H. Pishro-Nik (✉)
School of Electrical and Computer Engineering,
University of Massachusetts, 100 Natural Resources Road,
Amherst, MA 01003-9292, USA
e-mail: pishro@ecs.umass.edu

K. Chan · F. Fekri
School of Electrical and Computer Engineering,
Georgia Institute of Technology, Atlanta, GA
30332-0250, USA

K. Chan
e-mail: kschan@ece.gatech.edu

F. Fekri
e-mail: fekri@ece.gatech.edu

1 Introduction

Wireless sensor networks have received a great deal of interest lately. They have benefited from advances in both MEMS technology and networking. Potential applications include a diversified range in military and civilian surveillance and sensing tasks and services that would enhance the ability of the growing domain of wireless technologies [1].

Wireless sensor networks consists of a large number (in the order of thousands) of identical nodes which are constrained in available energy, computational power, memory, and communication range. In potential sensing applications, the sensor nodes may be randomly deployed in a hazardous or dangerous environment where the nodes are physically inaccessible after deployment. Hence, the design of the network needs to consider energy conserving schemes to account for a limited energy supply, low memory/computation and resilient networking schemes to account for the hostile environment.

The primary task of wireless sensor networks is to have the sensors relay information back to one or more base stations. This is accomplished without globally known network addressing (i.e. IP addresses); therefore, the sensor nodes rely on broadcasting techniques to deliver information in possibly a multihop fashion. Information is either sent from the base stations to the sensor nodes or from the sensor nodes to the base stations. The flow of information in wireless sensor networks distinguishes itself from ad hoc networking and other varieties of wireless networking. The characteristics of the other wireless communications is less defined, as communications may occur in node-to-node fashion which potentially requires communication schemes and algorithms that are quite different than those of wireless sensor networks.

The design of networks must consider routing protocols and communication schemes to best fit the intended sensing task at hand [2]. Furthermore, these schemes must observe the restrictions of the sensor network such as the conservation of energy while still maintaining a certain level of resiliency against node failure or capture. Resiliency in large-scale sensor networks is often linked to the connectivity of the network. That is, every node in the network should be able to communicate with the base stations in the network. Without such connectivity, the network is unable to provide proper functionality. Moreover, redundancy that is added through sending information through multiple paths is another characteristic within sensor networks that is utilized.

Graph theoretic properties of wireless networks have been studied extensively. In this paper we consider the effect of node and link failures, which are common in sensor networks, on different network properties. Our model of sensor networks assumes n sensors are distributed randomly over a field based on a given distribution function. We include link failures in our model, that is two active sensor nodes are connected with probability $p_e(n)$ if they are within the communication range of each other. The parameter $p_e(n)$ represents the effect of link failures, that is a link fails with probability $1 - p_e(n)$. In sensor networks, different factors may contribute to link failures. In some scenarios, the link failures are not black and white as assumed here. For example, each transmission, rather than each link, may fail with a certain probability. In these cases, our analysis can give an instantaneous status of the network. Nevertheless, in some other scenarios, a link may fail permanently with probability $1 - p_e(n)$. In these cases, our analysis provides a steady state status of the network. For example, consider key management schemes used for security of wireless sensor networks. In random key management schemes [3, 4], two neighbor nodes can establish a link only if they share a key. In these schemes, we choose a random key pool from the key space. Each key has an identifier. Before deployment, each sensor node is given a random subset of keys along with their identifiers from the key pool. If two nodes are in the communication range of each other and share a common key identifier, then they can use the corresponding key as their shared secret to initiate communication. In [4], authors gave a modified version of the above scheme which they called q-composite key predistribution scheme. If $s(n)$ is the number of keys in the key pool and $k(n)$ is the number of keys stored in each sensor, then we have

$$p_e(n) = 1 - \frac{\binom{s(n)}{k(n)}\binom{s(n)-k(n)}{k(n)}}{\binom{s(n)}{k(n)}^2}. \tag{1}$$

Usually, $k(n)$ and $s(n)$ are chosen such that $p_e(n)$ is bounded away from zero as n grows [3, 4]. Node failure is also a common phenomenon in sensor networks. Sensor nodes may fail due to lack of power, physical damage or environmental interference [1]. It is very important that the network can still continue to work properly even after some nodes have failed. In our model any sensor node may fail with probability $1 - p_{sf}(x, y, n)$, where (x, y) is the location of the node in the plane. For simplicity we study the link failures and node failures separately. First, we study the effect of link failures on the network. While some properties of link-reliable networks (networks with reliable links) can be easily extended to networks with unreliable links, some other properties require more complicated analysis. We then study the effect of node failures. We prove general statements relating node-reliable networks to unreliable ones. Using this general theorems we study the properties of networks with unreliable sensor nodes. Finally, we show that the two results can be combined for the analysis of networks with unreliable nodes and links.

The focus of this paper is to provide analysis of some network properties that affect network functionality. We study k-connectivity of large-scale sensor networks. We derive necessary and sufficient conditions for k-connectivity of the network graph. We study the minimum communication radius of sensor nodes to provide k-connectivity within the network. We analyze the average shortest path of the k paths from a node in the sensing field back to a base station. We also study the existence of the giant component (a large subset of nodes that are connected). These results have been shown through graph theoretical derivations and also have been verified through simulations. For clarity of exposition, we provide the lengthy proofs in the Appendix. However, it should be noted that a major contribution of this paper is to provide the mathematical methodology for dealing with large-scale sensor networks. Thus, an important part of the paper lies in the proofs of the results given in Appendix.

Formally, we say that a graph is connected if there is a path between every pair of vertices. A graph is said to be k-vertex-connected or simply k-connected if there does not exist a set of $k - 1$ vertices whose removal disconnects the graph. For $k \geq 2$, we say a graph is k-edge-connected if it has at least two vertices and no set of at most $k - 1$ edges separates it.

The k-connectivity property is important from the network reliability perspective. In particular, a k-connected network remains connected if less than k nodes are removed from the network as a result of node failures or an attack by an enemy. Moreover, k-connectivity is necessary for multi-path routing. The concept of k-connectivity considers a random graph and infers that there exists k disjoint paths between each pair of nodes. Thus, there exist k disjoint paths between any two nodes if an only if the associated random graph is k-connected. In terms of wireless networks, this implies that, on the link level, there exists k disjoint paths from each pair of nodes by hopping through unique sets of intermediate nodes. In the case of sensor networks, it is important to show the k-connectivity between the base station(s) and each sensor node in the field. However it is still up to the route discovery mechanism to find these k disjoint paths. The existence of the giant component is important when the network loses connectivity. In some applications, it is sufficient for the operation of the network to have a large subset of active nodes connected to each other (i.e., the network possesses a giant component).

Throughout the paper we assume $\mathcal{B}(\mathbb{R}^2)$ is the Borel σ-algebra on $\mathbb{R}^2$ and m is the Lebesgue measure on $\mathcal{B}(\mathbb{R}^2)$. $\overline{B(\overline{X},R)}$ is the closed ball with radius R centered at $\overline{X}$ in $\mathbb{R}^2$. $\overline{S(\overline{X},L)}$ is the closed square with side L centered at $\overline{X}$ in $\mathbb{R}^2$. In particular $S_0 = \overline{S(\overline{O},1)}$ is the closed square with unit area centered at the origin. For any $E \in \mathcal{B}(\mathbb{R}^2)$ we define $v(E) = m(E \cap S_0)$. Clearly v defines a measure on $\mathcal{B}(\mathbb{R}^2)$. For an integer $n, (n)_k = n(n-1)\ldots(n-k+1)$. For a random variable Y, $E(Y)_k$ shows the kth factorial moment. That is $E(Y)_k = E[Y(Y-1)\ldots(Y-k+1)]$. Let ε_n be an event depending on a parameter n. We say that ε_n holds asymptotically almost surely, or ε_n holds with high probability, if $\Pr\{\varepsilon_n\}$ tends to 1 as $n \to \infty$. For two sequences a_n and b_n, $a_n \sim b_n$ means $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

The remainder of the paper is structured into several parts. The next section provides an overview of the work related to our study. Section 3 establishes the formulation and preliminaries of the problem we have considered. Section 4 studies sensor networks with unreliable links and establishes proofs pertaining to connectivity and k-connectivity. Section 5 considers unreliable sensors and establishes a general connection between reliable and unreliable networks. We study some properties of unreliable sensor networks such as connectivity and the existence of the giant component. Section 6 contains simulations of these graph theoretic properties, in particular k-connectivity and average path lengths for networks with unreliable links and giant component analysis for networks with unreliable sensors. We propose some extensions of our work by applying them to specific routing algorithms in Sect. 7. Finally Sect. 8 concludes the paper.

2 Related work

Related problems to graph theoretic results in this paper have been studied in the context of random graph theory [5], continuum percolation and geometric probability [6–10] and the study of wireless network graphs [11–16]. In random graph theory, the model $G(n, p)$ is extensively studied, in which edges appear in a graph of n vertices with probability p independently of each other. In continuum percolation theory, usually infinite graphs on $\mathbb{R}^d$ are studied. Finally, in geometric probability and the study of graphs of wireless networks, the graphs in which nodes and links are reliable are usually studied.

Previously, k-connectivity of wireless networks has been studied in [17] and [18]. In [17] k-connectivity is studied in the context of fault-tolerant networks. The authors find lower bounds for the probability that the network is k-connected. They also present a method to control the network topology given that the network is k-tolerant (k-connected). In [18], authors study the asymptotic critical transmission radius for k-connectivity and asymptotic critical neighbor number for k-connectivity in wireless networks. The connectivity in ad-hoc and hybrid networks is studied in [19]. In [19], authors specifically consider the effect of base stations. They show that the introduction of a sparse network of base stations significantly increases the connectivity. In [20], trade-off between connectivity and capacity of dense networks is studied. In particular, the effect of the attenuation function on network properties is studied. In [21], authors consider a model in which two nodes can communicate if and only if the signal to noise ratio at the receiver is higher than some threshold. Thus, in this way they study the impact of interferences on the connectivity of ad hoc networks.

In this paper, we consider the connectivity properties of large-scale sensor networks. Thus, we consider the effects of the specific parameters of sensor networks on network properties. In particular, we consider unreliable links, unreliable nodes, and non-uniform distribution of nodes. However, in the papers mentioned above, it is assumed that links and nodes do not experience failures and nodes are distributed uniformly at random over the region. It is sometimes trivial to extend the previous results to include sensor networks (with node and link failures and non-uniform distribution). However, in many cases these new properties of sensor networks introduce new challenges. Thus, in this paper we need to use new methods for analyzing network properties. In particular, this paper provides the following contributions:

- It provides a detailed analysis on how the link failures affect connectivity and k-connectivity of sensor networks. The new challenges that arise from boundary conditions and the different regions for failure probabilities are addressed.

- It obtains the distributions for isolated nodes and discusses the effect of the sensor deployment density on the connectivity and the required transmission power.
- It introduces a general result that relates reliable networks and unreliable networks. This result provides a tool to extend the previous results on reliable networks to the case of networks with unreliable links and nodes.
- Using the general framework, it studies the emergence of the giant component in wireless sensor networks.
- The length of the shortest paths in the network is studied and the implications on multi-path routing algorithms are discussed.

It is worth noting that the node failure has been studied in [16]. However, the sensor deployment is confined to a grid and the random distribution of nodes is unexplored. A similar issue to link failures has been studied in [22] in the context of gossip-based routing. They introduce a gossiping-based routing, where each node forwards a message with some probability. However, [22] only provides empirical results. Moreover, in this paper, we introduce new results about the path lengths and latency in k-connected networks. In particular, we show that multi-path routing can be done efficiently (in a certain sense) in sensor networks.

In this paper we also consider multi-path routing with its implicit connection to k-connectivity. Several works have been published in sensor network multiple-path routing. Ganesan et al. [23] introduces multi-path routing in wireless sensor networks and considers disjoint and braided multi-paths. Our work provides the underlying mathematical foundations for which these algorithms may be applied. Multi-path routing in ad hoc networks has also been studied in [24]. Ayanoglu et al. [25] study coding diversity in multiple paths. The results in the following sections, particularly the study of k-connectivity, help to formalize the connectivity and availability of multiple paths in large-scale sensor networks.

Finally, there are some other papers that have empirically studied node failures and the lifetime of wireless sensor networks. Lifetimes of networks have also been considered in terms of energy usage of proposed communications and routing algorithms. Ganesan et al. [23] study the presence of patterned and isolated failures as it relates to multi-path routing. In [26], the lifetime of the network is measured in terms of the number of alive nodes as a function of time for a specific routing algorithm in LEACH. There are also comparisons the energy usage over time for several multicast and flooding schemes against proposed algorithm [27, 28]. Other common studies consider the packet delivery ratio [29, 30], but this work considers properties of the network on the link level. Our paper focuses on the broader scope of properties of wireless sensor networks as a whole, including connectivity, average path length, and the presence of a giant component.

3 Formulation and preliminaries

In this section we provide some definitions and preliminary lemmas that are needed throughout the paper. Wireless networks are sometimes modeled by the probability space of graphs that we represent with $g(n, r(n))$. The properties of this model have been studied previously [11, 12, 31]. In this model, it is assumed that n nodes are uniformly and randomly distributed over $S_0 = \overline{S(\overline{O}, 1)}$. If two nodes u and v satisfy $d(u, v) \leq r(n)$ ($d(u, v)$ is the Euclidean distance between u and v), then the edge $\{u, v\}$ belongs to edges of the graph. A more general model is the model $g(n, r(n), f_{XY})$ that is defined as follows. Let X and Y be absolutely continuous random variables with continuous joint density function $f_{XY}(x, y)$ satisfying $f_{XY}(x, y) > 0$ for all $(x, y) \in S_0 = \overline{S(\overline{O}, 1)}$, and $f_{XY}(x, y) = 0$ otherwise. A graph in $g(n, r, f)$ has n nodes and is generated as follows. For any node v, its position (X, Y) is chosen according to $f_{XY}(x, y)$ independently of other nodes. If two nodes u and v satisfy $d(u, v) \leq r(n)$, then the edge $\{u, v\}$ belongs to edges of the graph.

However, to study sensor networks, we now introduce two new parameters, link failure probability $1 - p_e(n)$ and node failure probability $1 - p_{sf}(x, y, n) = 1 - p_{sf}(x, y) p_{sf}(n)$. We first consider networks experiencing link failures. We introduce the probability space $g(n, r(n), f_{XY}(x, y), p_e(n))$ that we use to model graphs of sensor networks with possibly unreliable links. Let X and Y be absolutely continuous random variables with continuous joint density function $f_{XY}(x, y)$ satisfying $f_{XY}(x, y) > 0$ for all $(x, y) \in S_0 = \overline{S(\overline{O}, 1)}$, and $f_{XY}(x, y) = 0$ otherwise. A graph in $g(n, r(n), f_{XY}(x, y), p_e(n))$ has n nodes and is generated as follows. For any node v, its position (X, Y) is chosen according to $f_{XY}(x, y)$ independently of other nodes. If two nodes u and v satisfy $d(u, v) \leq r(n)$, then with probability $p_e(n)$ the edge $\{u, v\}$ belongs to the edges of the graph. Note that in the above model sensors are assumed to be reliable. Similar to reliable networks, if $f_{XY}(x, y) = 1_{\{(x,y) \in S_0\}}$ (i.e., nodes are distributed uniformly over the square S_0), we show the corresponding random graph by $g(n, r(n), p_e(n))$.

We then consider node failures. To study sensor networks with unreliable nodes, we define the probability space $g(n, r(n), p_{sf}(x, y, n))$, where $p_{sf}(x, y, n) = p_{sf}(x, y) p_{sf}(n)$. In this model n nodes are uniformly and randomly distributed over S_0; however, a sensor node at the point (x, y) is active with probability $p_{sf}(x, y) p_{sf}(n)$ and fails with probability $1 - p_{sf}(x, y) p_{sf}(n)$. The function $p_{sf}(x, y)$

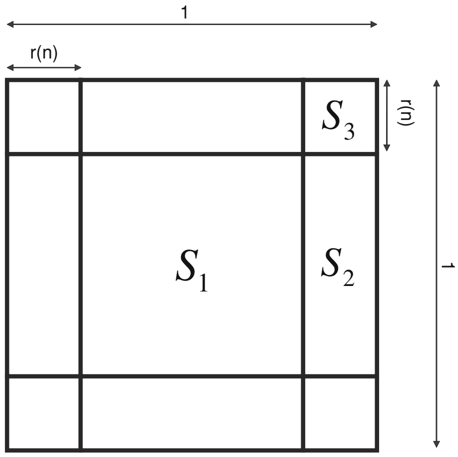

Fig. 1 The field S_0 and its divisions S_1, S_2, and S_3

models the possible spatial dependency of failure probability and $p_{sf}(n)$ models possible dependency on n. The nodes that are not active are assumed to be removed from the graph. If two active nodes u and v satisfy $d(u, v) \leq r(n)$, then the edge $\{u, v\}$ belongs to edges of the graph. The generalized model $g(n, r(n), f_{XY}(x, y), p_{sf}(x, y, n))$ is defined similarly. Finally we will consider the combined model $g(n, r(n), f_{XY}(x, y), p_e(n), p_{sf}(x, y, n))$. For simplicity, when there is no danger of confusion, we may drop the arguments, for example we may use $g(n, r, f_{XY})$ instead of $g(n, r(n), f_{XY}(x, y))$. For the purpose of analysis, we divide the square S_0 to different regions shown in Fig. 1.

The following lemma is useful when working on large-scale wireless sensor networks. It can be proved using direct computations and taking limits.

Lemma 1

$$\lim_{\frac{x}{r}\to o}\left[\frac{\pi r^2 - m\left(\overline{B(\overline{0}, r)} \cap \overline{B((0, x), r)}\right)}{2rx}\right] = 1. \tag{2}$$

$$\lim_{\frac{x}{r}\to o}\left[\frac{v\left(\overline{B((\frac{1}{2} - x, 0), r)}\right) - \frac{\pi r^2}{2}}{2rx}\right] = 1. \tag{3}$$

We frequently need to find asymptotic behavior of integrals of the form

$$\int_{-\infty}^{+\infty} \varphi(x, n)dx \quad n \to \infty, \tag{4}$$

in which $\varphi(x, n)$ has a sharp peak. These integrals can usually be approximated by the contribution of some neighborhood of the peak. This method is usually called the Laplace method for integrals.

We now quickly review some definitions and results from continuum percolation that we will need later. For a point process χ on $\mathbb{R}^2$ and a Borel set A, let $\chi(A)$ be the number of points of the process in A. The point process is said to be a Poisson process with density $\lambda > 0$ if [31]

- For mutually disjoint Borel sets $A_1, A_2, \ldots, A_k$, the random variables $\chi(A_1), \ldots, \chi(A_k)$ are mutually independent.
- For any bounded Borel set $A \in \mathcal{B}(\mathbb{R}^2)$ and for every $k \geq 0$, we have

$$\Pr\{\chi(A) = k\} = e^{\lambda m(A)} \frac{\lambda^k (m(A))^k}{k!}. \tag{5}$$

The model for continuum percolation that we use in this paper is obtained from a Poisson process that is conditioned to have a point at the origin $\chi_\lambda \cup \{\overline{O}\}$ and a connection radius d. In this model two points are connected to each other by an edge if their distance is less than or equal to d. We denote this model by (χ_λ, d) and show the corresponding graph by $g(\chi_\lambda, d)$. Let $p_k(\lambda)$ be the probability that the component of $g(\chi_\lambda, 1)$ containing the origin has k vertices. Then the percolation probability $p_\infty(\lambda)$ is the probability that $\overline{O}$ lies in an infinite component of the graph $g(\chi_\lambda, 1)$, and is defined by [6, 31]

$$p_\infty(\lambda) = 1 - \sum_{k=1}^{\infty} p_k(\lambda). \tag{6}$$

The critical value λ_c which is called the continuum percolation threshold is defined by

$$\lambda_c = \inf\{\lambda > 0 : p_\infty(\lambda) > 0\}. \tag{7}$$

It is well-known that $0 < \lambda_c < \infty$. In particular, we know that $.696 < \lambda_c < 3.372$ [6, 31].

4 Networks with unreliable links

In this section, we study the random graph $g(n, r(n), f_{XY}(x, y), p_e(n))$. Here is the summary of the results. In Sect. 1, we study connectivity (1-connectivity). Theorem 1 relates k-connectivity to minimum vertex degree. This simplifies the study of connectivity. In particular, Theorem 1 relates the connectivity to isolated vertices. Theorem 2 gives conditions for having isolated vertices. By combining Theorems 1 and 2, the connectivity of $g(n, r(n), p_e(n))$ is characterized in Theorem 3. Lemma 2 is used to show the existence of isolated vertices in the proof of Theorem 3. Theorem 4 takes the study one step further and provides the distribution of the isolated vertices for the unconnected networks. Finally, Theorem 5 generalizes the results for any continuous distribution of the sensor node deployment. In Sect. 2, we generalize the results for k-connectivity.

4.1 Connectivity

We first consider the case where $f_{XY}(x,y) = 1_{\{(x,y)\in S_0\}}$ (i.e., nodes are distributed uniformly over the square S_0). As we discussed, in this case we show the random graph by $g(n,\ r(n),\ p_e(n))$. Similar results for general $f_{XY}(x,\ y)$ will be given later. We first need to prove a lemma. Let $A_{n,1}, A_{n,2}, \ldots, A_{n,n}$ be a sequence of events in the probability space $(\Omega_n, \mathcal{F}_n, P_n)$. Let $X_{n,j}$ be the random variable defined to be one when $A_{n,j}$ occurs and zero otherwise for $j = 1, 2, \ldots, n$. Let also $X_n = \sum_{j=1}^{n} X_{n,j}$ be the number of events that occur from the set $\{A_{n,1}, A_{n,2}, \ldots, A_{n,n}\}$. Define

$$\mu_n = E[X_n] = \sum_{j=1}^{n} \Pr\{A_{n,j}\}, \tag{8}$$

$$\Delta_n = \sum_{i=1}^{n} \sum_{j\neq i} \Pr\{A_{n,i} \cap A_{n,j}\}. \tag{9}$$

We now state the following lemma which is similar to Janson's inequality; however, it is applicable to a more general case.

Lemma 2 *Let* $A_{n,\ j}, \Delta_n, \mu_n$ *be as defined. Assume* $\lim_{n\to\infty} \mu_n = \mu, 0 \le \mu \le \infty$ and $\lim_{n\to\infty} \Delta_n = \Delta, 0 \le \Delta \le \infty$.

Then

$$\limsup_{n\to\infty} \Pr\left\{\bigcap_{i=1}^{n} \overline{A_{n,i}}\right\} \le 1 - \mu + \frac{\Delta}{2}. \tag{10}$$

If $\Delta \ge \mu$, then

$$\limsup_{n\to\infty} \Pr\left\{\bigcap_{i=1}^{n} \overline{A_{n,i}}\right\} \le 1 - \frac{\mu^2}{2\Delta}. \tag{11}$$

Proof We have

$$\begin{aligned} 1 - \Pr\left\{\bigcap_{i=1}^{n} \overline{A_{n,i}}\right\} &\ge \sum_{j=1}^{n} \Pr\{A_{n,j}\} - \sum_{i=1}^{n} \sum_{j\le i} \Pr\{A_{n,i} \cap A_{n,j}\} \\ &= \mu_n - \frac{\Delta_n}{2} \end{aligned} \tag{12}$$

Thus,

$$\limsup_{n\to\infty} \Pr\left\{\bigcap_{i=1}^{n} \overline{A_{n,i}}\right\} \le 1 - \mu + \frac{\Delta}{2}. \tag{13}$$

Now, if $\Delta \ge \mu$, Then $\frac{\mu_n}{\Delta_n} \le 1 + o(1)$. Let $J \subseteq \{1, 2, \ldots, n\}$ be chosen in the following way. For any $i \in \{1, 2, \ldots, n\}$, we have $i \in J$ with probability $\frac{\mu_n}{\Delta_n}(1 - o(1)) \le 1$ independently. Then, using (12) we have

$$\begin{aligned} \Pr\left\{\bigcap_{i\in J} \overline{A_{n,i}}\right\} &\le 1 - \sum_{i\in J} \Pr\{A_{n,i}\} \\ &\quad + \frac{1}{2} \sum_{i\in J} \sum_{j\in J, j\neq i} \Pr\{A_{n,i} \cap A_{n,j}\}. \end{aligned} \tag{14}$$

By taking expectation, we get

$$\begin{aligned} E\left[\Pr\left\{\bigcap_{i\in J} \overline{A_{n,i}}\right\}\right] &\le 1 - E\left[\sum_{i\in J} \Pr\{A_{n,i}\}\right] \\ &\quad + E\left[\frac{1}{2} \sum_{i\in J}^{n} \sum_{j\in J, j\neq i} \Pr\{A_{n,i} \cap A_{n,j}\}\right] \\ &= 1 - \frac{\mu_n}{\Delta_n}\mu_n - \left(\frac{\mu_n}{\Delta_n}\right)^2 \frac{\Delta_n}{2} + o(1) \\ &= 1 - \frac{\mu_n^2}{2\Delta_n} + o(1). \end{aligned} \tag{15}$$

In particular, there exists $J \subseteq \{1, 2, \ldots, n\}$ such that

$$\Pr\left\{\bigcap_{i\in J} \overline{A_{n,i}}\right\} \le 1 - \frac{\mu_n^2}{2\Delta_n} + o(1). \tag{16}$$

Therefore, we obtain

$$\Pr\left\{\bigcap_{i=1}^{n} \overline{A_{n,i}}\right\} \le \Pr\left\{\bigcap_{i\in J} \overline{A_{n,i}}\right\} \le 1 - \frac{\mu_n^2}{2\Delta_n} + o(1). \tag{17}$$

Taking limits we obtain

$$\limsup_{n\to\infty} \Pr\left\{\bigcap_{i=1}^{n} \overline{A_{n,i}}\right\} \le 1 - \frac{\mu^2}{2\Delta}. \tag{18}$$

Consider the class of graphs $g(r) = g(n,\ r,\ f,\ p_e)$ in which the radius r is variable and all other parameters are fixed. In other words, to generate a class of graphs from the ensemble, we place n nodes randomly and independently on S_0. For any two nodes v and w, we assign the number x_{vw} which is zero with probability $1 - p_e(n)$ and is 1 with probability $p_e(n)$. Now for a given r, the vertices v and w are connected by an edge if and only if $x_{vw} = 1$ and $d(u,\ v) \le r$. Let Q be a property of graphs and let

$$r(Q) = \inf\{r\text{: g(r) has Q}\}. \tag{19}$$

Let $Q_{c,k}$ be the property of being k-connected and let $Q_{\delta,k}$ be the property that the minimum degree of the graph is at least k. The following result is very similar to the one for the $g(n,p)$ model. It can be shown by using arguments similar to [32] and [10] and we omit the proof due to the space limitation.

Theorem 1 *Given a positive integer k, for almost all g(r) in* $g(n,\ r,\ f,\ p_e)$ *we have*

$$r(Q_{c,k}) = r(Q_{\delta,k}). \tag{20}$$

We note that if $r(Q_{ce,k})$ is the corresponding threshold for k-edge-connectivity, we have $r(Q_{c,k}) \geq r(Q_{ce,k}) \geq r(Q_{\delta,k})$. Thus Theorem 1 implies that $r(Q_{c,k}) = r(Q_{ce,k}) = r(Q_{\delta,k})$.

Discussion This theorem states that for large enough networks, the graph is *k*-connected if and only if the minimum vertex degree is at least *k*. This is very useful because studying the minimum degree is much simpler than studying *k*-connectivity. This can also be useful in practice when we want to check the connectivity number of the networks. A simple algorithm is to look at the minimum vertex degree in the graph.

We now consider the connectivity of the random graph $g(n, r, p_e)$. Let $V = \{v_1, v_2, \ldots, v_n\}$ be the set of vertices of a random graph $g_n = g(n, r, p_e)$ that are uniformly placed on $S_0 = \overline{S(\overline{O}, 1)}$. Suppose $\overline{X}_i = (x_i, y_i)$ is the position of v_i for $i = 1, 2, \ldots, n$ and $B_i = B(\overline{X}_i, r(n))$ is the coverage area of v_i. For any node v_i, if we know the location of the node $\overline{X}_i = (x_i, y_i)$, then the probability that the node is isolated (i.e., the node is not connected to any other node in the graph) is given by

$$(1 - v(B_i)p_e(n))^{n-1}. \tag{21}$$

Since $\overline{X}_i = (x_i, y_i)$ is uniformly distributed over S_0, the probability that a certain node in the graph is isolated is given by

$$n \int_{S_0} \left(1 - v(B(\overline{X}, r(n)))p_e(n)\right)^{n-1} dm(\overline{X}). \tag{22}$$

Let Z_n be the number of isolated vertices in g_n. Then

$$\begin{aligned} EZ_n &= EZ_n(r(n)) \\ &= n \int_{S_0} \left(1 - v(B(\overline{X}, r(n)))p_e(n)\right)^{n-1} dm(\overline{X}). \end{aligned} \tag{23}$$

It is easy to prove that EZ_n is a decreasing function of n and there exists $r^*(n)$ satisfying $0 \leq \lim_{n\to\infty} EZ_n(r^*(n)) \leq \infty$. We call $r^*(n)$ a threshold of $g_n = g(n, r, p_e)$ for isolated vertices. In fact, as we will see, $r^*(n)$ is a threshold for the property of having isolated vertices in the graph. Thus by Theorem 1, $r^*(n)$ is the connectivity threshold.

Theorem 2 *Let $p_e(n) \geq \frac{c}{\ln n}$, for some constant c. Then $r(n) = r^*(n)$ is a threshold of $g_n = g(n, r, p_e)$ for isolated vertices if and only if*

$$0 < \lim_{n\to\infty} \left[n\pi r^2(n)p_e(n) - \ln(n)\right] \leq \infty. \tag{24}$$

More specifically, $\lim_{n\to\infty} EZ_n(r(n)) = 0$ if and only if $\lim_{n\to\infty} \left[n\pi r^2(n)p_e(n) - \ln(n)\right] = \infty$ and $\lim_{n\to\infty} EZ_n(r(n)) = \infty$ if and only if $\lim_{n\to\infty} \left[n\pi r^2(n)p_e(n) - \ln(n)\right] = -\infty$.

Discussion Theorem 2 gives us the threshold for isolated vertices. As we will see asymptotically, this determines the threshold for connectivity. This theorem also reveals an important difference between reliable networks like $g(n, r)$ and unreliable networks such as $g(n, r, p_e)$. To see this, let us examine the condition $p_e(n) \geq \frac{c}{\ln n}$. It is worth noting that the condition $p_e(n) \geq \frac{c}{\ln n}$ is not crucial for our proofs. We can still prove the existence of connectivity thresholds without assuming this condition. However, without this assumption, the results would not have closed form representation. Instead, they would include integrals over the region. Hence, the results would depend on the field shape and boundary. As we will see, by assuming $p_e(n) \geq \frac{c}{\ln n}$, we will obtain very simple conditions for connectivity and the results would not depend on the shape of the sensor field. In fact, although we prove the theorems for S_0, they can be extended to all regions with smooth boundary. Thus unlike reliable networks, in unreliable networks, if $p_e(n)$ is small, the connectivity properties of the networks may depend on the shape of the deployment field. In these networks, unlike the reliable networks, the boundary effects are important. Nevertheless, in most practical applications such as random key distribution schemes, the condition $p_e(n) \geq \frac{c}{\ln n}$ is usually satisfied. This theorem is proved in the Appendix.

The connectivity of $g(n, r, p_e)$ can be characterized by the following theorem.

Theorem 3 *Consider the random graph $g = g(n, r, p_e)$. Let $p_e(n) \geq \frac{c}{\ln n}$, for some constant c. Then g is connected asymptotically almost surely if and only if $\lim_{n\to\infty} \left[n\pi r^2(n) p_e(n) - \ln(n)\right] = \infty$.*

Discussion Theorem 3 gives a necessary and sufficient condition for connectivity of $g(n, r, p_e)$. In particular, we can observe the effect of link failures on the connectivity of the network. Under the condition $p_e(n) \geq \frac{c}{\ln n}$, the effect of p_e can be modeled by defining an *effective radius* $r_{eff}(n) = \sqrt{p_e(n)}r(n)$. That is, the random graph $g(n,r,p_e)$ is asymptotically almost surely connected if and only if $g(n,r_{eff})$ is connected asymptotically almost surely. However, if the condition $p_e(n) \geq \frac{c}{\ln n}$ does not hold, such an easy interpretation is not possible. The theorem is proved in the Appendix.

Moreover, we can find the distribution of the isolated vertices as follows.

Theorem 4 *Consider the random graph $g = g(n, r, p_e)$ for which $p_e(n) \geq \frac{c}{\ln n}$. Let*

$$r(n) = \sqrt{\frac{\ln n + c}{\pi n p_e(n)}}. \tag{25}$$

Let I_n be the number of isolated vertices in g, which are in $\overline{S(0, 1-2r(n))}$. Let $I \in Po(e^{-c})$ (i.e., I has Poisson distribution with mean e^{-c}). Then I_n converges in distribution to I.

Discussion Theorem 4 gives the distribution of the number of isolated vertices in $g(n, r, p_e)$. First of all, if the condition of Theorem 3 is satisfied, then we should have $c \to \infty$ and thus $e^{-c} \to 0$, which implies that there is no isolated vertices in the network with high probability. This is obviously predictable because the network should be connected in this case. On the other hand, when $c < \infty$ the network is not connected because of some isolated vertices. One way to solve this problem is to increase the communication coverage of the isolated vertices such that they get connected to the rest of the graph. Theorem 4 provides the number of isolated vertices in the network in these situations. Thus, we can estimate the amount of extra transmission power needed for connectivity.

Proof We use the method of factorial moments to prove the theorem. It suffices to show

$$E(I_n)_k \to e^{-kc} \quad \text{as} \quad n \to \infty \text{ for } \quad k = 1, 2, \ldots \tag{26}$$

In fact, for $k = 1$ and 2, this has been shown in the proof of Theorem 3 and it is easily extendable to higher values of k. Let $Is(\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k)$ be the probability that the nodes which are located at the locations $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k$ are isolated in $g = g(n, r, p_e)$. Then

$$E(I_n)_k = (n)_k \int\limits_{(S_1)^k} Is(\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k) dm(\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k) \tag{27}$$

Thus, for example by (46) in the Appendix, $E(I_n)_1 \to e^{-c}$ as $n \to \infty$. For $k = 2$, we note that $E(I_n)_2 = (1 - o(1))\Delta_n^1$, where Δ_n^1 defined in (54). Thus, using (55), we conclude that $E(I_n)_2 \to e^{-2c}$ as $n \to \infty$. Finally, we note that this argument can be generalized for an arbitrary k.□

In summary, Theorem 3 gives the necessary and sufficient condition for connectivity for $g = g(n, r, p_e)$. We now generalize this result to any other continuous density function $f_{XY}(x, y)$ as follows. First, note that since $\overline{S_0}$ is a compact set in $\mathbb{R}^2$ and $f_{XY}(x, y) > 0$ for all $(x, y) \in \overline{B(0, R)}$, the function $f_{XY}(x, y)$ has a strictly positive minimum on $\overline{B(0, R)}$. We call this minimum f_{min}. The following theorem gives the the necessary and sufficient condition for connectivity of $g = g(n, r, f, p_e)$.

Theorem 5 *Consider the random graph $g = g(n, r, f, p_e)$ for which $p_e(n) \geq \frac{c}{\ln n}$, and $f_{min} = \min\{f_{XY}(x, y), (x, y) \in S_0\}$. Then g is connected asymptotically almost surely if and only if there exists ω(n) satisfying $\omega(n) \to \infty$ as $n \to \infty$ and $n_0 > 0$ such that*

$$r(n) \geq \sqrt{\frac{\ln n + \omega(n)}{n p_e(n) \pi f_{min}}} \quad \text{for} \quad n \geq n_0. \tag{28}$$

Discussion The main message here is that the connectivity condition is completely determined by the area in the field that has the lowest density f_{min}. Thus, if we have a non-uniform distribution of nodes, assuming the same communication radius, we will need more nodes to obtain a connected network.

Proof (Sketch) If $r(n) \geq \sqrt{\frac{\ln n + \omega(n)}{n p_e(n) \pi f_{min}}}$, then the expected number of isolated vertices in S_0 tends to zero by direct calculation and by comparison with (23). Thus, there are no isolated vertices with high probability. On the other hand, if $\lim\limits_{n\to\infty} \sup \omega(n)$ in (28) is finite, then for a small enough ε, we consider a square S' in S_0 such that $f_{XY}(x,y) < (1 + \varepsilon) f_{min}$ for all $(x, y) \in S'$. Then, similar to the proof of Theorem 3, we can show that with a strictly positive probability independent of n, there exists an isolated vertex in S'. □

This theorem implies an interesting property of the uniform distribution:

Corollary 1 *The uniform distribution $f_{XY}(x, y) = 1_{\{(x,y) \in S_0\}}$ requires the lowest amount of transmission power for connectivity.*

If we let $p_e(n)$ be the probability of having a shared secret key between two nodes, then Theorem 5 gives a necessary and sufficient condition for the connectivity of the graph in the general key distribution schemes.

4.2 *K*-connectivity

In this section we study the k-connectivity property of $g(n, r, f, p_e)$. In summary, the k-connectivity transitions are very sharp. In fact, similar to the situation in $G(n, p)$ model, it can be shown that increasing $\pi r^2(n) p_e(n) \ln n$ by an additive factor $O(\ln \ln n)$ will change the probability of k-connectivity from $o(1)$ to $1 - o(1)$. Although, this can be proved using similar arguments to the previous section, for analyzing sensor networks, we might be interested in a coarser view of the k-connectivity threshold. Again, for simplicity we prove the result for the case $f_{XY}(x, y) = 1_{\{(x,y) \in S_0\}}$, and then state the general result for other densities by considering the minimum value of the density function f_{min}.

Theorem 6 *Consider the random graph $g = g(n, r, p_e)$. Let $p_e(n) \geq \frac{c}{\ln n}$, for some constant c. Assume*

$$\lim_{n \to \infty} \left(\frac{n \pi r^2(n) p_e(n)}{\ln n} \right) = \alpha. \tag{29}$$

Let k be a positive integer. If $\alpha > 1$, then g is k-connected asymptotically almost surely. On the other hand, if $\alpha < 1$, then g is not k-connected asymptotically almost surely.

Discussion Note that the condition given here for k-connectivity does not depend on k. We can actually give a more refined condition for k-connectivity, and show that increasing $\pi r^2(n)p_e(n)\ln n$ by an additive factor $O(\ln\ln n)$ will change the probability of k-connectivity from $o(1)$ to $1 - o(1)$. However, in practice the condition given here is sufficient to show the behavior of k-connectivity. An important conclusion that we obtain here is that, the transition from a disconnected graph to a fully k-connected graph is very sharp in large-scale sensor networks. Thus, for example, the graph is actually disconnected with high probability when $\alpha = .99$. On the other hand choosing $\alpha = 1.01$, the graph suddenly becomes k-connected. However, in practice, depending on the network size, we may need to choose a larger α to ensure k-connectivity.

It is also worth noting that for the special case of reliable networks in which $p_e(n) = 1$, the result of Theorem 6 is consistent with [17]. In particular, for reliable networks, it is shown in [17] that if $n\pi r^2(n) \geq \ln n + (2k-1)\ln\ln n - 2\ln(k!) + 2\beta$, then the probability that the network is $(k+1)$-connected is at least $e^{-e^{-\beta}}$. Now for $p_e(n) = 1$, our condition reduces to $\lim_{n\to\infty}\left(\frac{n\pi r^2(n)}{\ln n}\right) > 1$. It easy to see that under this condition we should have $\lim_{n\to\infty}\beta = +\infty$. Thus, the probability of $(k+1)$-connectivity, $e^{-e^{-\beta}}$, converges to one when n approaches infinity, as suggested by Theorem 6. This theorem is proved in the Appendix.

Similar to Theorem 5, we can generalize Theorem 6 to other density functions.

Theorem 7 *Consider the random graph* $g = g(n, r, f, p_e)$ *for which* $p_e(n) \geq \frac{c}{\ln n}$, *and* $f_{min} = \min\{f_{XY}(x,y), (x,y) \in S_0\}$. *Assume*

$$\lim_{n\to\infty}\left(\frac{nf_{min}\pi r^2(n)p_e(n)}{\ln n}\right) = \alpha. \qquad (30)$$

Let k be a positive integer. If $\alpha > 1$, then g is k-connected asymptotically almost surely. On the other hand, if $\alpha < 1$, then g is not k-connected asymptotically almost surely.

As we mentioned previously, k-connectivity is a necessary and sufficient condition for the existence of at least k-disjoint paths between every two vertices in the graph. In sensor networks, we may only need k disjoint paths between the sink and other nodes. However, in large scale sensor networks, this requirement is also equivalent to k-connectivity. The reason is as follows. If the graph is k-connected then obviously there are at least k disjoint paths between the sink and any other node in the graph. On the other hand, if the graph is not k-connected, there is a node in the graph with degree lower than k with high probability by Theorem 1. Thus, there cannot be k disjoint paths between this node and the sink.

5 Networks with unreliable sensors

Here, we consider sensor failures. The summary of the results is as follows. In Sect. 1 we provide fundamental results that relate unreliable networks to the reliable ones. First, using Lemma 3, we simplify the analysis by reducing the graph model $g(n, r(n), f_{XY}(x, y), p_{sf}(x, y, n))$ to $g(n, r(n), f_{XY}(x, y), p_{sf}(n))$. We then prove Theorem 8 which shows a general method to obtain the properties of the unreliable networks using the previously known results concerning the reliable networks. Theorem 9 is a converse to Theorem 8. Section 2 provides two important applications of these results. Theorem 10 provides a necessary and sufficient condition for k-connectivity. Theorem 12 provides conditions for having a giant component. Finally, in Sect. 3, we combine the results on link failures and node failures.

5.1 Connection between reliable and unreliable networks

In continuum percolation, unreliable nodes are handled easily by using the Thinning Theorem, which states that an unreliable (with the above definition of reliability) Poisson process is equivalent to a reliable one. For instance, if in the process χ_λ, each node is accepted with probability p and rejected with probability $1-p$, then the resulting process is equivalent to $\chi_{\lambda\ p}$, that is a Poisson point process with density λp. However, the relation between reliable graphs $(g(n, r(n), f_{XY}))$ and unreliable graphs $(g(n, r(n), f_{XY}, p_{sf}))$ is more complicated. In this section, we prove a general result about this relation. This results allows us to find properties of unreliable sensor networks from the well studied model for reliable networks.

Note that a common choice for $p_{sf}(x, y, n)$ is a spatially uniform distribution of unreliability, that is $p_{sf}(x, y, n) = p_{sf}(n)$ for all $(x,y) \in S_0$. However, in some scenarios, sensor nodes at some part of the field may be more prone to failure than other parts. For these situations a spatially non-uniform $p_{sf}(x, y, n)$ is more suitable. We first prove that it suffices to study the uniform $p_{sf}(x, y, n) = p_{sf}(n)$. This is because any $g(n, r(n), f_{XY}(x, y), p_{sf}(x, y, n))$ is equivalent to $g(n, r(n), f'_{XY}(x, y), p'_{sf}(n))$ for some $f'_{XY}(x, y)$ and $p'_{sf}(n)$ as shown in below. Remember we always assume $p_{sf}(x, y, n) = p_{sf}(x, y)\, p_{sf}(n)$.

Lemma 3 *The two models $g(n, r(n), f_{XY}(x, y), p_{sf}(x, y, n))$ and $g(n, r(n), f_{XY}'(x, y), p_{sf}'(n))$ are equivalent if*

$$f'_{XY}(x,y) = \frac{f_{XY}(x,y)p_{sf}(x,y)}{\int_{S_0} f_{XY}(x,y)p_{sf}(x,y)dxdy}$$
$$p'_{sf}(n) = \int_{S_0} f_{XY}(x,y)p_{sf}(x,y,n)dxdy. \tag{31}$$

Discussion Note that as we mentioned before, in these models we always assume that the failed sensors are removed from the graph. Otherwise, obviously the two models will not be equivalent. The importance of this lemma is its implication that we only need to study $g(n, r(n), f_{XY}(x, y), p_{sf}(n))$. That is, we do not need to consider the dependency of p_{sf} on the location (x, y) because it can be absorbed in $f'_{XY}(x, y)$ as stated in the lemma. This significantly simplifies the analysis.

Proof Since in both models the location of a sensor nodes and its failure is independent of the other sensor nodes, it suffices to prove that in both models each sensor fails with the same probability, and if it does not fail its location has the same probability distribution in both models. First, we note that in $g(n, r(n), f_{XY}(x, y), p_{sf}(x, y, n))$ each sensor is active with probability

$$\int_{S_0} f_{XY}(x,y)p_{sf}(x,y,n)dxdy = p'_{sf}(n), \tag{32}$$

which is the corresponding probability in $g(n, r(n), f'_{XY}, p'_{sf}(n))$. Now, if a sensor does not fail, in $g(n, r(n), f'_{XY}, p'_{sf}(n))$ its location has the density function $f'_{XY}(x, y)$. In $g(n, r(n), f_{XY}(x, y), p_{sf}(x, y, n))$, if a node does not fail its location has the density function

$$\begin{aligned} f'_{XY}(x,y) &= \frac{f_{XY}(x,y)p_{sf}(n)p_{sf}(x,y)}{\int_{S_0} f_{XY}(x,y)p_{sf}(x,y)p_{sf}(n)dxdy} \\ &= \frac{f_{XY}(x,y)p_{sf}(x,y)}{\int_{S_0} f_{XY}(x,y)p_{sf}(x,y)dxdy} \\ &= f'_{XY}(x,y). \end{aligned}$$

□

Thus, from now on we study $g(n, r(n), f_{XY}(x, y), p_{sf}(n))$. We also note that the model $g(n, r(n), f_{XY}(x, y), p_{sf}(n))$ is similar to the $G(\mathcal{P}_\lambda; r)$ defined in [31] in the sense that both have a random number of nodes. However, there is an important distinction between them. The model $G(\mathcal{P}_\lambda; r)$ is simpler to work with because of the spatial independency in the Poisson process. However, we do not have such spatial independency property in $g(n,r(n),f_{XY}(x,y),p_{sf}(n))$. Thus, in [31] the model $G(\mathcal{P}_\lambda; r)$ is used to prove some properties of $g(n, r(n), f_{XY})$ but here we use $g(n, r(n), f_{XY})$ to prove properties of $g(n, r(n), f_{XY}(x, y), p_{sf}(n))$.

Let Q be a property of graphs. Then, $g \in Q$ means the graph g has property Q. The following result establishes a connection between reliable and unreliable networks. It is in some sense similar to the relation between $G(n, p)$ and $G(n, M)$ given in [5, 33] and in fact it is proved using a similar argument. We say that almost every graph in $g(n, r(n), f_{XY}(x, y), p_{sf}(n))$ has Q if $g(n, r(n), f_{XY}(x, y), p_{sf}(n))$ has Q asymptotically almost surely.

Theorem 8 *Let Q be a graph property and let $p_{sf}(n)(1 - p_{sf}(n))n \to \infty$ as $n \to \infty$. If for every sequence $m = m(n)$ satisfying $m = np_{sf}(n) + O\left(\sqrt{np_{sf}(n)(1-p_{sf}(n))}\right)$, we have $Pr\{g(m(n), r(n), f_{XY}(x, y))$ has Q$\} \to 1$ as $n \to \infty$, then almost every graph in $g(n, r(n), f_{XY}(x, y), p_{sf}(n))$ has Q.*

Discussion This is a fundamental theorem that relates unreliable networks to reliable ones. In particular, it shows how to apply any previously known result for reliable networks, to prove the same result for unreliable networks. Note that the theorem is quite general and can be applied to any properties of the networks, not just the connectivity properties.

Proof Let $N(g)$ be the number of vertices of the graph g and $q(n) = 1 - p_{sf}(n)$. For any positive real number β, let $A_n(\beta)$ be the set of integers m satisfying $|m - p_{sf}(n)n| \leq \beta\sqrt{p_{sf}(n)q(n)n}$. Let $E_n(\beta)$ be the event that $N(g(n, r(n), f_{XY}(x, y), p_{sf}(n))) \in A_n(\beta)$ and let $E_n^c(\beta)$ be its compliment. Then by Chebyshev's inequality

$$\begin{aligned} \Pr\{E_n^c(\beta)\} = \Pr\{&N(g(n, r(n), f_{XY}(x,y), p_{sf}(n))) \\ &\in A_n(\beta)\} \leq \frac{1}{\beta^2}. \end{aligned}$$

Let also $m_{min}(n)$ be an element of $A_n(\beta)$ with the lowest $\Pr\{g(m(n),r(n),f_{XY}(x,y))\}$ and define m_{max} similarly. Then we have

$$\begin{aligned} &\Pr\{g(n,r,f_{XY},p)) \text{ has Q}\} \\ &\quad \geq \Pr\{g(n,r,f_{XY},p)) \text{ has Q given } \mathrm{E_n}(\beta)\}\Pr\{E_n(\beta)\} \\ &\quad \geq \Pr\{g(m_{min}(n), r(n), f_{XY}(x,y)) \text{ has Q}\}(1 - \frac{1}{\beta^2}) \\ &\quad \geq (1 - o(1))(1 - \frac{1}{\beta^2}). \end{aligned}$$

If we let β tend to infinity, then $1 - \frac{1}{\beta^2}$ tends to one, thus we conclude that $\Pr\{g(n, r, f_{XY}, p))$has Q$\}$ is greater than any fixed real number less than one. Thus $\Pr\{g(n, r, f_{XY}, p)$has Q$\}$ tends to one as n goes to infinity. Therefore, almost every graph in $g(n, r, f_{XY}, p)$ has Q. □

Theorem 8 shows how to apply previously proven results for reliable networks to prove the same results for unreliable networks. The converse is also possible for certain properties, although it is less interesting in this paper. To show the converse we first need some definitions. For two graphs g, g' on $\mathbb{R}^2$, we write $g' \subset_v g$ if g' is obtained by deleting a subset of vertices of g. We say that property Q is increasing if whenever $g' \in Q$ and $g' \subset_v g$

then $g \in Q$. Similarly, we say that property Q is decreasing if whenever $g \in Q$ and $g' \subset_v g$ then $g' \in Q$. Finally Q is said to be convex if $g' \subset_v g \subset_v g''$ and $g' \in Q, g'' \in Q$ imply that $g \in Q$. Note that the above definitions are slightly different from the usual definitions of increasing, decreasing, and convex properties in graph theory. Note also that if Q is either increasing or decreasing then it is convex. For example, if Q is the property of having a specific subgraph, then obviously Q is increasing. Therefore, it is also convex.

Suppose Q is an increasing property. Let $p^1_{sf}(n) < p^2_{sf}(n)$ and $p(p^i_{sf}, Q)$ be the probability that $g(n, r, f_{XY}, p^i_{sf})$ has Q for $i = 1, 2$. Using a coupling argument we can easily show that $p(p^1_{sf}, Q) \le p(p^2_{sf}, Q)$. Thus, if $g(n, r, f_{XY}, p^1_{sf}) \in Q$ with high probability, then $g(n, r, f_{XY}, p^2_{sf}) \in Q$ with high probability, as well. Similarly if Q is decreasing then $p(p^1_{sf}, Q) \ge p(p^2_{sf}, Q)$. Finally if Q is a convex property and we have $g(n, r, f_{XY}, p^1_{sf}) \in Q$ and $g(n, r, f_{XY}, p^2_{sf}) \in Q$ with high probability, then we can conclude for $p^1_{sf}(n) < p^3(n)_{sf} < p^2_{sf}(n), g(n, r, f_{XY}, p^3_{sf}) \in Q$ with high probability. Using this fact, we can prove the following theorem. It states that for convex properties we can use unreliable networks to prove the similar properties for reliable networks. Here, we just state the theorem and omit the proof.

Theorem 9 *Let Q be a convex property and let $p_{sf}(n)(1 - p_{sf}(n))n \to \infty$ as $n \to \infty$. If almost every graph in $g(n, r(n), f_{XY}(x, y), p_{sf}(n))$ has Q, then for fixed real number β, $Pr\{g(m_x(n), r(n), f_{XY}(x, y))$ has $Q\} \to 1$ as $n \to \infty$, where $m_\beta(n) = \lfloor p_{sf}(n)n + \beta\sqrt{p_{sf}(n)q(n)n} \rfloor$.*

Discussion This is the converse to Theorem 8. In other words, if a result has been previously proven for $g(n, r(n), f_{XY}(x, y), p_{sf}(n))$, for $p_{sf}(n) \neq 0$, we can use this theorem to conclude the same result for $g(n, r(n), f_{XY}(x, y))$.

Finally, we end the section by noting that the number of active nodes has a Gaussian distribution. Let $N(g)$ be the number of (active) vertices of the graph $g = g(n, r(n), f_{XY}(x, y), p_{sf}(n))$ and $q(n) = 1 - p_{sf}(n)$. Define

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt. \tag{33}$$

If $p_{sf}(n)(1 - p_{sf}(n))\ n \to \infty$, then by the Laplace-Demoivre Theorem we have

$$\begin{aligned} &\Pr\{|N(g(n, r, f_{XY}, p)) - p_{sf}(n)n| \le x\sqrt{p_{sf}(n)q(n)n}\} \\ &\quad = (1 + o(1))[\Phi(x) - \Phi(-x)]. \end{aligned} \tag{34}$$

5.2 Some properties of unreliable sensor networks

In this section, we specifically study some important graph theoretic properties of node-unreliable sensor networks. We employ the results in the previous section relating reliable and unreliable networks. These results can be proved directly. However, using the previous work on $g(n, r(n), f_{XY}(x, y))$ and the previous section results, they can be proved in a much simpler way. First, we find the necessary and sufficient condition for k-connectivity. Then, we study another important property in the existence of a giant component. For simplicity we only consider the case that nodes are distributed uniformly over the field. That is $f_{XY}(x, y) = 1_{\{(x,y) \in S_0\}}$. Thus we may use $g(n, r(n), p_{sf}(n))$ and $g(n, r(n))$ to represent $g(n, r(n), f_{XY}(x, y), p_{sf}(n))$ and $g(n, r(n), f_{XY}(x, y))$ respectively.

We now study k-connectivity of $g(n, r(n), f_{XY}(x, y), p_{sf}(n))$. As a special case of Theorem 6 if we let $p_e(n) = 1$, then we obtain the following result.

Corollary 2 *Consider the random graph $g = g(n, r, f_{XY})$ with $f_{XY}(x, y) = 1_{\{(x,y) \in S_0\}}$. Assume*

$$\lim_{n \to \infty} \left(\frac{n\pi r^2(n)}{\ln n} \right) = \alpha. \tag{35}$$

Let k be a positive integer. If $\alpha > 1$, then g is k-connected asymptotically almost surely. On the other hand, if $\alpha < 1$, then g is not k-connected asymptotically almost surely.

We now prove the following theorem on k-connectivity of unreliable networks.

Theorem 10 *Consider $g = g(n, r(n), f_{XY}(x,y), p_{sf}(n))$ with* $f_{XY}(x, y) = 1_{\{(x,y) \in S_0\}}$ and assume $np_{sf}(n) \to \infty$ as $n \to \infty$. Assume

$$\lim_{n \to \infty} \left(\frac{n\pi p_{sf}(n) r^2(n)}{\ln p_{sf}(n) + \ln n} \right) = \alpha. \tag{36}$$

Let k be a positive integer. If $\alpha > 1$, then g is k-connected asymptotically almost surely. On the other hand, if $\alpha < 1$, then g is not k-connected asymptotically almost surely.

Discussion Note that this is very similar to Theorem 6. Thus, one way to prove this, is to use similar proofs given for the previous section. However, as we see applying Theorem 8 makes the proof much simpler.

Proof We use Theorem 8. Consider a sequence $m = m(n)$ satisfying $m = np_{sf}(n) + O(\sqrt{np_{sf}(n)(1 - p_{sf}(n))})$, then

$$\lim_{m(n) \to \infty} \left(\frac{m(n)\pi r^2(n)}{\ln m(n)} \right) = \lim_{n \to \infty} \left(\frac{np(1 + o(1))\pi r^2(n)}{\ln(np(1 + o(1)))} \right) = \alpha.$$

Thus, by Theorem 8 and Corollary 2, if $\alpha > 1$, then g is k-connected asymptotically almost surely. On the other hand, if $\alpha < 1$, then g is not k-connected asymptotically almost surely. $\square$

So far, we have studied conditions for connectivity of unreliable sensor networks. On the other hand, if a graph is not connected, it can be divided into connected components (disjoint connected subgraphs). In these situations, the sensor network may continue to operate if it has one large component. For the graph $g = g(n, r(n))$ it has been shown in [31] that there exists a threshold $r^*(n)$ such that when $r(n)/r^*(n) < 1$, all components are small (logarithmic in n) with high probability. On the other hand, if $r(n)/r^*(n) > 1$, there exists one giant component (with size linear in n), and other components are small. Note that if the density function is not uniform, there may be more than one giant component. We now generalize these results to unreliable sensors. Again, for simplicity we only consider a uniform distribution of nodes the field, that is $f_{XY}(x,y) = 1_{\{(x,y)\in S_0\}}$. Thus we drop the density function from the notation. The general case of non-uniform distribution can be proved similarly. Let L_j denote the size of the jth largest component in a graph. We recall that the critical value λ_c is the continuum percolation threshold. The following theorem is proved in [31].

Theorem 11 *Consider the random graph $g(n, r(n))$ and suppose $nr^2(n) \to \lambda$ as $n \to \infty$. Then, if $0 < \lambda < \lambda_c$, there exists a positive constant δ independent of n such that the size of the largest component satisfies $L_1 < \delta \ln n$ with high probability. On the other hand if $\lambda > \lambda_c$, there exists a positive constant α independent of n such that the size of the largest component satisfies $L_1 > \alpha n$ with high probability. Moreover, the size of other components is sublinear. That is, for $j > 1$, $L_j/n \to 0$ as $n \to \infty$ with high probability.*

We now state and prove the corresponding result for unreliable sensor networks, $g(n, r(n), p_{sf}(n))$.

Theorem 12 *Consider the random graph $g(n, r(n), p_{sf}(n))$ and suppose $np_{sf}(n) \to \infty$ and $n\, p_{sf}(n)\, r^2(n) \to \lambda$ as $n \to \infty$. Then if $0 < \lambda < \lambda_c$, there exists a positive constant δ independent of n such that the size of the largest component satisfies $L_1 < \delta \ln n$. On the other hand if $\lambda > \lambda_c$, there exists a positive constant α independent of n such that the size of the largest component satisfies $L_1 > \alpha np_{sf}(n)$ with high probability. Moreover the size of other components is sublinear. That is for $j > 1$, $L_j/(np_{sf}(n)) \to 0$ as $n \to \infty$ with high probability.*

Discussion Note that direct proof of this theorem is very involved and cumbersome. However, as we see by using Theorem 8, the proof is almost trivial.

Proof Again we use Theorem 8. Consider a sequence $m = m(n)$ satisfying $m = np_{sf}(n) + O(\sqrt{np_{sf}(n)(1-p_{sf}(n))})$, then

$$\lim_{m(n)\to\infty} m(n)r^2(n) = \lim_{n\to\infty}\left[np_{sf}(n) + O(\sqrt{np_{sf}(n)(1-p_{sf}(n))})\right]r^2(n) = \lim_{n\to\infty} np(1+o(1))r^2(n) = \lambda.$$

Thus, if $0 < \lambda < \lambda_c$, by Theorem 11, there exists a positive constant δ independent of n such that the size of the largest component satisfies $L_1 < \delta \ln m(n)$ with high probability. Thus, we conclude that there exists a positive constant δ' independent of n such that $L_1 \leq \delta' \ln(np_{sf}(n))$. On the other hand if $\lambda > \lambda_c$, there exists a positive constant α independent of n such that the size of the largest component satisfies $L_1 > \alpha m(n)$ with high probability. Thus we conclude that there exists a positive constant α' independent of n such that $L_1 > \alpha' np_{sf}(n)$ with high probability. Moreover, the size of other components is sublinear. Thus, by Theorem 8 we conclude the proof of this theorem. □

5.3 Networks with Unreliable Links and Nodes

We can easily combine the results in previous sections to analyze $g(n, r, f, p_e, p_{sf})$. Here we state the results for connectivity.

Corollary 3 *Let Z_n be the the number of isolated vertices in $g_n = g(n, r, p_e, p_{sf})$ and assume $p_e(n) \geq \frac{c}{\ln n}$, for some constant c Then $r(n) = r^*(n)$ is a threshold of g for the existence of isolated vertices if and only if*

$$0 < \lim_{n\to\infty}\left[n\pi r^2(n)p_{sf}p_e(n) - \ln(n)\right] < \infty. \quad (37)$$

More specifically, $\lim_{n\to\infty} EZ_n(r(n)) = 0$ if and only if $\lim_{n\to\infty}\left[n\pi r^2(n)p_{sf}(n)p_e(n) - \ln(n)\right] = \infty$ and $\lim_{n\to\infty} EZ_n(r(n)) = \infty$ if and only if $\lim_{n\to\infty}\left[n\pi r^2(n)p_{sf}\, p_e(n) - \ln(n)\right] = -\infty$.

Corollary 4 *Consider the random graph $g = g(n, r, f, p_e, p_{sf})$ for which $p_e(n) \geq \frac{c}{\ln n}$, and $f_{min} = \min\{f_{XY}(x,y), (x,y) \in S_0\}$. Then g is connected asymptotically almost surely if and only if there exists $\omega(n)$ satisfying $\omega(n) \to \infty$ as $n \to \infty$ and $n_0 > 0$ such that*

$$r(n) \geq \sqrt{\frac{\ln n + \omega(n)}{np_{sf}(n)p_e(n)\pi f_{min}}} \quad \text{for } n \geq n_0. \quad (38)$$

Corollary 5 *Consider the random graph $g = g(n, r, f, p_e, p_{sf})$ for which $p_e(n) \geq \frac{c}{\ln n}$, and $f_{min} = \min\{f_{XY}(x,y), (x,y) \in S_0\}$. Assume*

$$\lim_{n\to\infty}\left(\frac{nf_{min}\pi r^2(n)p_{sf}(n)p_e(n)}{\ln n}\right) = \alpha. \quad (39)$$

Let k be a positive integer. If $\alpha > 1$, Then g is k-connected asymptotically almost surely. On the other hand, if $\alpha < 1$, Then g is not k-connected asymptotically almost surely.

6 Simulation results

Simulations were run to verify the theoretical development found in the previous sections. The implication of the preceding development is a threshold effect on connectivity under certain network parameters. We show this in networks of varying communication radii ($r(n)$), containing unreliable sensors (sensor failure with probability $1 - p_{sf}(n)$), unreliable links (link failure with probability $1 - p_e(n)$) and of varying distributions (f_{min}). Furthermore, the developments have shown that k-connectivity is achieved rapidly at this threshold. We provide the results of simulations to validate these claims.

The results for networks of size $n = \{1000, 2000, 5000\}$ are provided. They have been deployed into a field S_0 of unit dimensions with finite communication radius, $r(n)$. We look at the probability of disconnectivity, p_{disc}, as a function of varying each of the network parameters that we have considered in this work such as $r(n)$, $p_{sf}(n)$, $p_e(n)$, and f_{min}. Looking at the various sizes of networks verifies that the claims are asymptotically valid, as the behavior of connectivity around a threshold is increasingly tighter. This threshold effect is such that for values lower than this threshold, the graph is disconnected with high probability. For values above this threshold, the graph is connected with high probability. Therefore, in the following figures, we represent the theoretical threshold as a step function, where $p_{disc} = 1$ for values below the threshold and $p_{disc} = 0$ past this threshold. The simulations show that the threshold effect occurs situations where each of the network parameters are varied.

We also provide two related characteristics of network connectivity in looking at shortest paths between nodes and the presence of giant components. We look at the relationship between k-connectivity and average shortest path lengths. We show an important result in that the first k shortest paths in a k-connected graph have almost the same length (by the length of a path, we mean the number of hops). Also, we examine the size of the giant component is simulated for networks with unreliable sensor nodes.

6.1 Connectivity versus communications radius

The threshold for the radius required to provide connectivity has been derived in previous sections. In this section, we provide simulation results to validate the theoretical development of this property (38). As we consider networks of different size, we show that the minimum transmission radius occurs at

$$r(n) \geq (1+\varepsilon)\sqrt{\frac{\ln n}{np_{sf}(n)p_e(n)\pi f_{min}}} \tag{40}$$

where ε is some small fixed constant. Here we set $\varepsilon = .1$.

We assume a fixed, uniform communications radius for each node in the network. Additionally, the nodes are distributed uniformly, $f_{XY}(x,y) = 1_{\{(x,y)\in S_0\}}$, where also $p_{sf}(n) = p_e(n) = 1$. We see that (40) determines the value of $r(n)$ at which this threshold for connectivity should occur. From Figs. 2, 3 and 4, we see a threshold effect in p_{disc} as $r(n)$ increases. The effect grows tighter to bound as the size of the network increases. This was expected since the theoretical results are asymptotic and apply to very large networks.

6.2 Networks with unreliable links and sensors

With (38), we can also derive the requirement for $p_{sf}(n)$ and $p_e(n)$ to achieve connectivity within the network. In this section, we provide simulation results to validate the

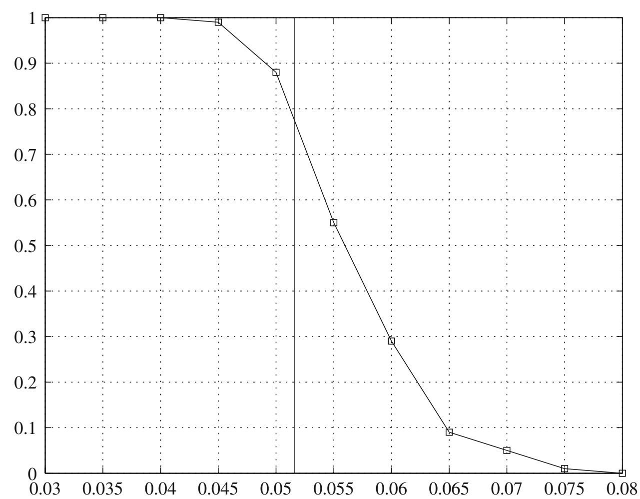

Fig. 2 The minimum radius to provide connectivity for a network of size $n = 1000$

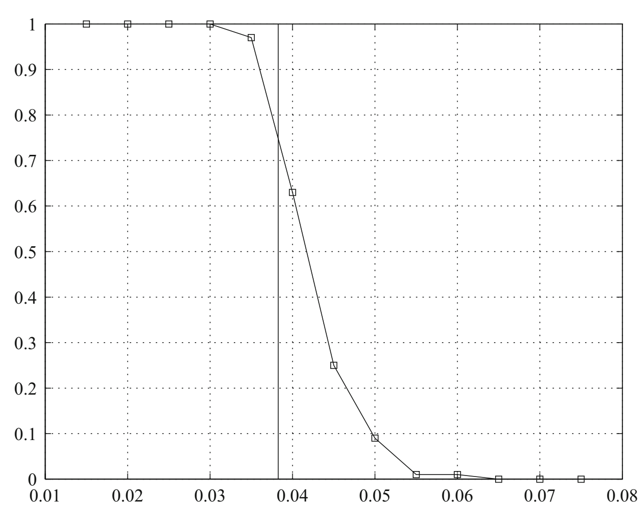

Fig. 3 The minimum radius to provide connectivity for a network of size n = 2000

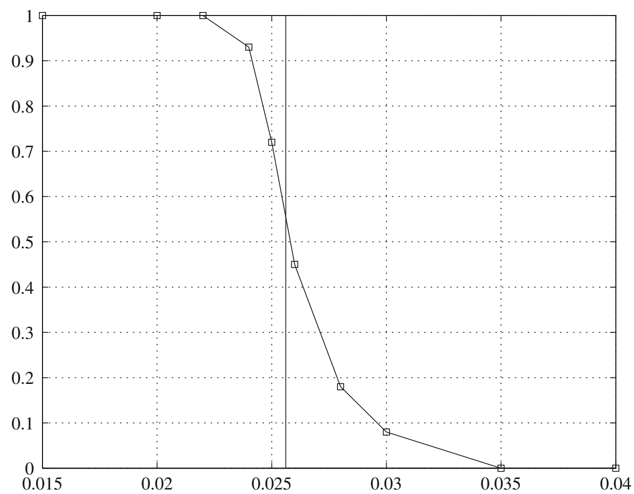

Fig. 4 The minimum radius to provide k-connectivity for a network of size n = 5000

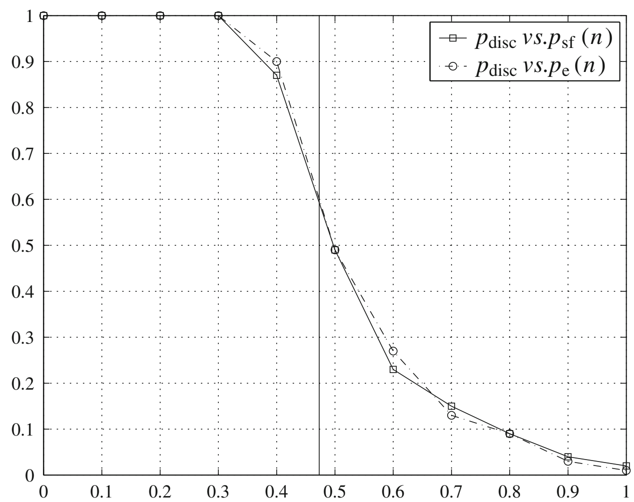

Fig. 5 Plot of p_{disc} versus both p_e and p_{sf} for n = 1000

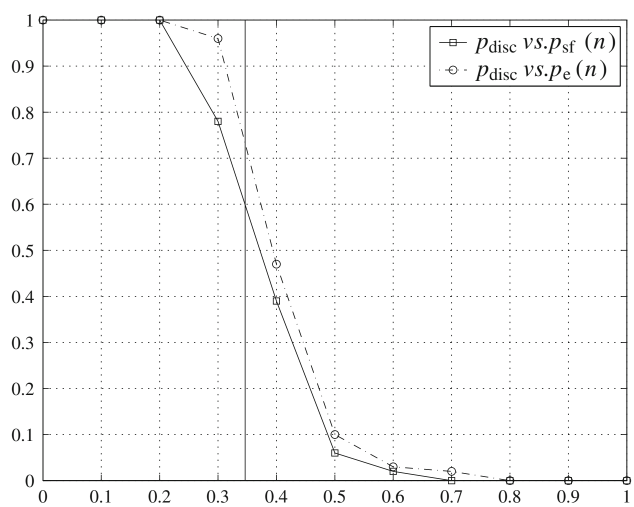

Fig. 6 Plot of p_{disc} versus both p_e and p_{sf} for n = 2000

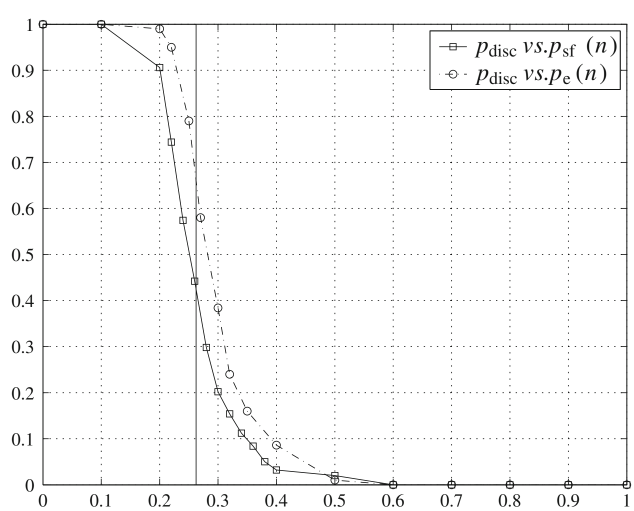

Fig. 7 Plot of p_{disc} versus both p_e and p_{sf} for n = 5000

threshold effect for networks with unreliable links and sensors. We consider the two cases separately, but it also is easy to consider them simultaneously. The sensor failure occurs when after deployment, the node fails to communicate with any device with probability $1 - p_{sf}(n)$. For the link failure, we assume that any link between two nodes within the communication range of each other is formed with probability $p_e(n)$. As we consider networks with failures in either links or sensors, we show that connectivity is achieved at

$$p_{sf}(n) \geq \frac{\ln n}{np_e(n)\pi f_{min} r^2(n)}(1+\varepsilon') \tag{41}$$

and

$$p_e(n) \geq \frac{\ln n}{np_{sf}(n)\pi f_{min} r^2(n)}(1+\varepsilon') \tag{42}$$

In experiment, we assume a fixed, uniform communications radius for each node in the network. Additionally, the nodes are distributed uniformly, $f_{XY}(x,y) = 1_{\{(x,y)\in S_0\}}$, where we have fixed $r(n)$ greater than the threshold of connectivity for n = {1000, 2000, 5000}, respectively. For instance, for n = 5000, we have chosen $r(5000)$ = .05, where the threshold value is $r(n) \geq$ .0256. Figures 5, 6 and 7 show the probability of disconnectivity versus the values of $p_e(n)$ or p_{sf}. For the plot where we vary $p_e(n)$ we set $p_{sf}(n) = 1$ and where we vary $p_{sf}(n)$ we set $p_e(n) = 1$. We have provided the results for networks of size n = {1000, 2000, 5000}, respectively. We see a threshold effect in p_{disc} as $p_e(n)$ and $p_{sf}(n)$ increase. The threshold effect is increasingly drastic as the size of the network increases. We note that as the network size increases, the simulation results approach the theoretical threshold.

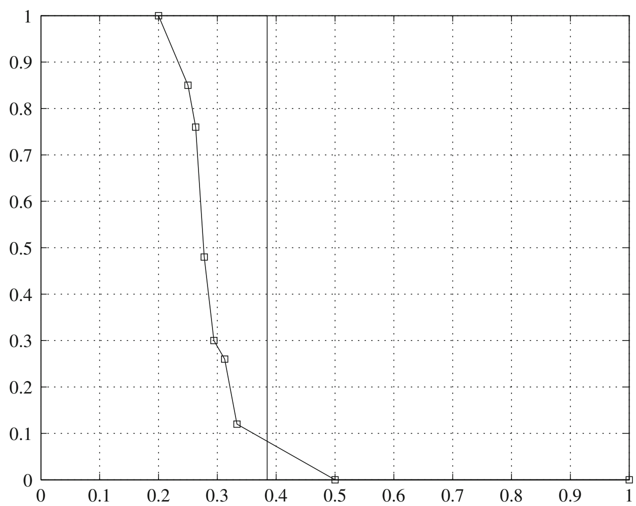

Fig. 8 P_{disc} versus σ for n = 5000

6.3 Connectivity versus the distribution of nodes within networks

Thus far, we have considered the case where the distribution of the nodes is uniform in S_0. Recall the distribution function $f_{XY}(x, y) = 1_{\{(x,y) \in S_0\}}$. We have also stated that the results are valid for any distribution, where the requirement is dependent on f_{min} the minimum density in S_0.

Therefore, we choose to look at the normal distribution, with truncation. That is, we consider a bivariate normal distribution of nodes on the unit area S_0, only choosing nodes whose coordinates were within S_0. The relationship between σ and f_{min} is determined by (43) and (44).

The distribution of nodes in this case

$$f_{XY}(xy) = \alpha e^{-(x^2+y^2)/2\sigma^2}_{\{(x,y) \in S_0\}}. \quad (43)$$

where

$$\alpha = \left(\int_{-.5}^{.5} \int_{-.5}^{.5} e^{-(x^2+y^2)/2\sigma^2} dy dx \right)^{-1} \quad (44)$$

We are able to observe various values of f_{min} by varying the value of σ. In Fig. 8, we see that the threshold of p_{disc} and observe that p_{disc} for the truncated bivariate normal distributions follows the general threshold for connectivity. Distributions were generated from several values in $\sigma = [.2, \ 1]$. Note that we considered reliable sensor networks.

6.4 Average shortest path in k-connected networks

In this section we consider k-connectivity. Maintaining a network with several paths when failures in links may occur is important. Furthermore, in routing protocols, multiple paths are used to add redundancy to packet transmission through diversity [25]. Here, we show that for a k-connected sensor network, the first k shortest paths between two nodes in the network have almost the same length. Therefore, when using multiple paths for transmission, the latency between using different paths does not deviate considerably.

In our simulation, we considered a network of $n = 5000$ nodes with a fixed uniform communication radius of $r(n)$ = .05 that ensures k-connectivity for $k < 6$. We also set $p_e(n) = p_{sf}(n) = 1$. We select two nodes in extremal areas of the region S_0. The simulation finds the shortest path between the extremal nodes. Then, the intermediate nodes, those nodes which were used to traverse between the two nodes, are eliminated from the network and the new shortest path is found again. The experiment is repeated to achieve the average shortest path for k-connectivity for k = {1, 2, 3, 4, 5}. The result of this simulation shows that the average shortest path for k = {1, 2, 3, 4, 5} varies by only one hop. This shows confidence that latency among multiple shortest paths does not vary greatly and also demonstrates a great potential for routing algorithms that consider multiple paths. This is a desirable property for algorithms of large-scale sensor networks that employ multiple paths for robust routing and networking schemes. It is also a desirable property for networks with sleeping sensors because it suggests that only a small penalty may be paid if the first shortest path is not used for packet transmission due to sleeping nodes (Fig. 9).

6.5 Giant component within networks

In some instances, it may be acceptable to not have full connectivity with all nodes. Instead, a certain proportion of

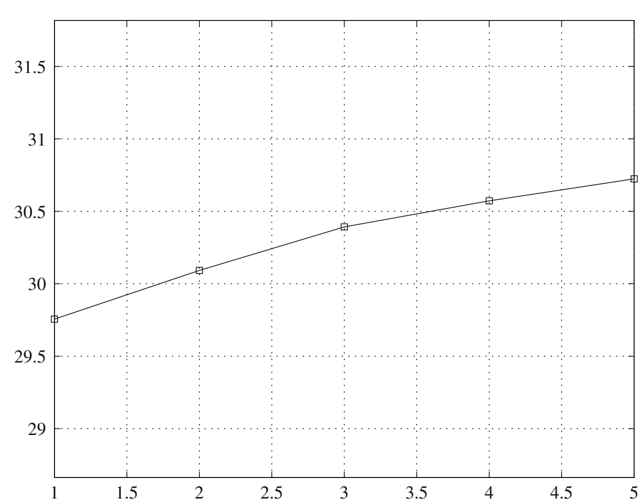

Fig. 9 Average Shortest path for k = {1, 2, 3, 4, 5}, n = 5000, r = .05

the nodes may be connected and be able to function adequately. In this simulation, we look to the presence of a giant component within these networks, the largest subset of the active nodes that is connected. We have considered this problem in the case where unreliable nodes exist by considering different values of the number of active nodes in the network by varying $p_{sf}(n)$ in a network of size $n = 10{,}000$ with communication radius $r(n) = .025$.

In the Theorem 12, the threshold for the network to possess a giant component defined

$$p_{sf} n r^2 = \lambda_c \cong 1.44 \Rightarrow p_{sf} \cong .2304 \tag{45}$$

Therefore the size of the giant component will decrease sharply as $p_{sf}(n)$ decreases below .2304. We have identified this characteristic to be important in the potentially capability of sensor networks. Certainly, it is not desirable for large portions of the network not to be able to communicate with the majority of the nodes. This affects the ability of the nodes to relay information back to the base station.

Figure 10 shows the size of the giant component and the number of active sensors as a function of $p_{sf}(n)$. The solid line represents the average number of active sensors in the network for the specified value of the probability that a node is active, $p_{sf}(n)$, and the dashed lines with boxes is the average size of the giant component in the network. This additionally provides justification of the threshold effect of wireless networks that we have described in this work. The giant component has a threshold effect along with the connectivity.

Collectively, in this section simulation results have verified the theoretical exposition in the preceding sections. We have considered connectivity properties of large-scale networks of varying size. These simulations have confirmed the theoretical developments of unreliable networks with sensor failures and link failures. We have also shown that these claims are valid for other distributions of nodes. Additionally, we have shown that the first shortest paths, on average, are not drastically different in length for the k-connected networks.

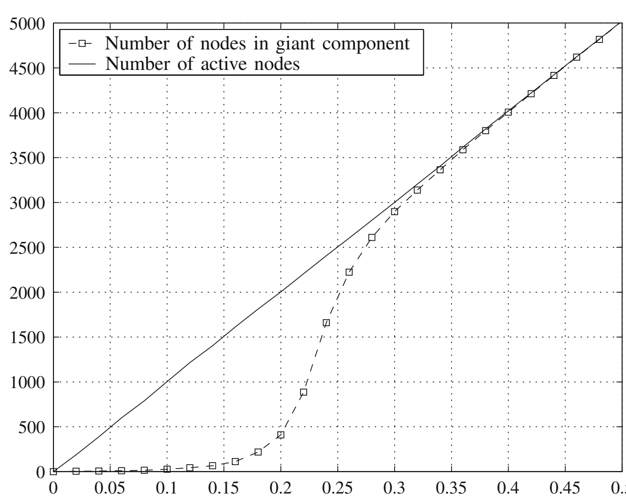

Fig. 10 The size of the giant component and the number of active nodes versus $p_{sf}(n)$, the probability that a node is active

7 Possible applications

We now explore application to which this work may be applied. The graph theoretic derivations and simulations were completed independent of the consideration of specific routing protocols. This work addresses general connectivity properties of large-scale sensor networks. The results of this work would be practically useful to study sensor networks with unreliable links and sensor nodes. However, applying the concepts of connectivity that are studied in this paper may provide further insight into the resiliency of the multi-path elements of particular routing algorithms or sleeping sensor networks. We present preliminary analysis of several routing algorithms and how this work may be extended to these specific cases. Three algorithms have been selected, one from each of the three classifications that was noted in [2].

(1) Directed diffusion [27] is a routing algorithm that establishes a multi-hop network where the base station broadcasts interests to the sensor network, and the sensor nodes forward and respond to the interests, forming interest gradients so that the flow of information is established. Several paths from nodes to the base stations, and based on a positive and negative reinforcement scheme paths are utilized according to their quality.

The reinforcement scheme benefits from the analysis provided in this work. The number of hops in the k paths from the nodes to the base stations is one component in determining path quality. This work, particularly the average shortest path of the k paths, offers an indication of the lengths (number of hops) in each of the possible paths. Therefore, it is possible to measure the effects of having one or more paths in the interest gradient to undergo negative reinforcement and to determine the number and quality of the remaining paths. There is a large potential of analysis with Directed Diffusion and its k-connectivity properties. Generally, the data-centric routing protocols may benefit greatly from the study of the k-connectivity in large-scale sensor networks.

2) Low-energy adaptive clustering hierarchy (LEACH) [34] is a hierarchical routing protocol that establishes clusters of nodes based who send their information to a local clusterhead, which is chosen by received signal

strength and available energy. Each node rotates on a random round schedule to become a clusterhead. In this way, the network is partitioned into clusters, where the clusterheads in each cluster communicate directly with the base station.

LEACH does not employ a multi-hop strategy, and it is apparent that this study does not offer analytical insight into the general classification of hierarchical routing protocols. However, it may be possible to relate the study of k-connectivity to the approximate neighborhood density of each node. With a k-connected network this provides a lower bound on the number of neighbors of each potential clusterhead. However, this is not a direct application of the results of this study.

3) Greedy perimeter stateless routing (GPSR) [30] is a location-based routing protocol where knowledge of the location of the sensors is available to sensor nodes; therefore, the flow of information is done by forwarding packets to nodes that are closer to the intended destination. For instances where there does not exist such a node, a perimeter routing scheme is employed. This routing protocol may benefit from a study on k-connectivity. For this protocol, a planar graph is required. In this way, a larger communication radius would be required since potential links would not be established to preserve the planar property of the graph. Additionally, the study of the average path length offers an idea of the increase in path length for the paths which include the perimeter routing. Other, location-based routing protocols figure to benefit from similar analyses.

8 Conclusion

We studied several properties of large-scale sensor networks. We have investigated different graph theoretic properties of sensor networks such as k-connectivity, giant component and disjoint paths. We considered a model for these networks that includes node and link failures. We proved a general result connecting reliable and unreliable networks. For any positive integer k, we derived the necessary and sufficient conditions for k-connectivity of the sensor network. If $k = 1$, the corresponding condition is the necessary and sufficient condition for connectivity which is clearly an important property of the network. Moreover, k-connectivity is investigated for potential application in multi-path routing or networks with sleeping sensors. The giant component is also studied. We also verified our results by simulation. In particular, we showed that multiple disjoint paths can be found with length very close to the length of the shortest path in a k-connected sensor network. This shows the potential efficiency of multi-path routing in large-scale sensor networks.

Acknowledgements This work is supported by National Science Foundation under grants CCF-0728970 and CCF-0728772.

Appendix

Proofs

Proof of Theorem 2

Proof Define $\omega(n) := n\pi r^2(n)p_e(n) - \ln(n)$, thus $\pi r^2(n) = \frac{\ln n + \omega(n)}{np_e(n)}$. Let $S_1 = \overline{S(\overline{O}, 1 - 2r(n))}$. We now obtain

$$\begin{aligned} EZ_n &= n\int_{S_0} \left(1 - v(B(\overline{X}, r(n)))p_e(n)\right)^{n-1} dm(\overline{X}) \\ &\geq n\int_{S_1} \left(1 - v(B(\overline{X}, r(n)))p_e(n)\right)^{n-1} dm(\overline{X}) \\ &= n\int_{S_1} \left(1 - \frac{\ln n + \omega(n)}{n}\right)^{n-1} dm(\overline{X}) \\ &= n\left(1 - \frac{\ln n + \omega(n)}{n}\right)^{n-1} m(S_1) \\ &= e^{-\omega(n)}(1 + o(1)). \end{aligned} \tag{46}$$

Therefore, we conclude that $\lim_{n\to\infty} EZ_n(r(n)) = \infty$ if $\lim_{n\to\infty} \omega(n) = -\infty$. Moreover, $\lim_{n\to\infty} EZ_n(r(n)) > 0$ if $\lim_{n\to\infty} \omega(n) \leq \infty$. Now assume that $\lim_{n\to\infty} \omega(n) > -\infty$. Let $Y_{3,n}$ be the number of isolated vertices in S_3. Then we get

$$\begin{aligned} EY_{3,n} &\leq nr^2(n)\left(1 - \frac{\pi r^2(n)}{4}p_e(n)\right)^{n-1} \\ &\leq nr^2(n)e^{-\frac{\pi r^2(n)}{4}p_e(n)(n-1)}. \end{aligned} \tag{47}$$

Using $p_e(n) \geq \frac{c}{\ln n}$ and $\pi r^2(n) = \frac{\ln n + \omega(n)}{np_e(n)}$, we conclude

$$EY_{3,n} = O\left(\frac{\ln n(\ln n + \omega(n))e^{-\omega(n)/4}}{n^{\frac{1}{4}}}\right) = o(1). \tag{48}$$

Therefore, there is no isolated vertex in S_3 with high probability. Next, let $Y_{2,n}$ be the number of isolated vertices in S_2. Then

$$EY_{2,n} = n\int_{S_2} \left(1 - v(B(\overline{X}, r(n)))p_e(n)\right)^{n-1} dm(\overline{X}) \tag{49}$$

Using the Laplace method for integrals and Lemma 1, it can be shown that

$$EY_{2,n} = O\left(\frac{e^{-\frac{\omega(n)}{2}}}{r(n)p_e(n)\sqrt{n}}\right) \tag{50}$$

Using $p_e(n) \geq \frac{c}{\ln n}$ and $\pi r^2(n) = \frac{\ln n + \omega(n)}{np_e(n)}$, we conclude

$$EY_{2,n} = O\left(\frac{e^{-\frac{\omega(n)}{2}}}{\sqrt{\left(c + \frac{c\omega(n)}{\ln(n)}\right)}}\right). \tag{51}$$

Thus if $\lim_{n\to\infty} \omega(n) = \infty$ then $Y_{2,n} = 0$ asymptotically almost surely. Moreover, if $0 \leq \lim_{n\to\infty} \omega(n) \leq \infty$ then $Y_{2,n}$ is finite asymptotically almost surely. Combining with (46) we conclude the theorem. □

Proof of Theorem 3

Proof By Theorem 2, when $\lim_{n\to\infty}\left[n\pi r^2(n)p_e(n) - \ln(n)\right] = \infty$, we have $\lim_{n\to\infty} EZ_n(r(n)) = 0$. Thus, by Markov's inequality there is no isolated vertex with high probability. Then, by Theorem 1 the graph is connected asymptotically almost surely. Hence, we focus on the proof of the other direction. That is if $0 < \lim_{n\to\infty}\left[n\pi r^2(n)p_e(n) - \ln(n)\right] < \infty$ (or equivalently $0 < \lim_{n\to\infty} EZ_n(r(n)) < \infty$), then there exists $\delta > 0$ such that $\liminf_{n\to\infty} p_n^{disc} > \delta > 0$, where p_n^{disc} is the probability that g_n is disconnected. The proof is as follows. Let $A_{n,j}$ be the event that the vertex v_j is isolated. Then we want to prove

$$\limsup_{n\to\infty} \Pr\left\{\bigcap_{i=1}^{n} \overline{A_{n,i}}\right\} \leq 1. \tag{52}$$

To prove the above, we use Lemma 2. Let $\Delta_n = \sum_{i=1}^{n}\sum_{j\neq i} \Pr\{A_{n,i} \cap A_{n,j}\}$. We show that under the condition $0 \leq \mu < \infty$, we have $\lim_{n\to\infty} \Delta_n = \Delta < \infty$. Thus by applying Lemma 2 we conclude the theorem. It remains to prove $\Delta < \infty$. We note that

$$\begin{aligned}\Delta_n \leq n(n-1) \int_{S_0\times S_0} &\Big(1 - v(B(\overline{X}, r(n)))p_e(n)\\ &- v(B(\overline{X}, r(n)))p_e(n)\\ &+ v(B(\overline{X}, r(n))) \cap B(\overline{Y}, r(n)))p_e^2(n)\Big)^{n-2} dm(\overline{X}) \times m(\overline{Y})\end{aligned} \tag{53}$$

We have $S_0 \times S_0 = (S_1 \times S_1) \cup (S_0 \times S_0 \setminus S_1 \times S_1)$. It suffices to show that the integral over the set $S_1 \times S_1$ and $S_0 \times S_0 \setminus S_1 \times S_1$ is finite. Let Δ_n^1 and Δ_n^2 be the two integrals respectively. For example, for $S_1 \times S_1$ we have

$$\begin{aligned}\Delta_n^1 &= n(n-1) \int_{S_1\times S_1} \Big(1 - v(B(\overline{X}, r(n)))p_e(n)\\ &\quad - v(B(\overline{X}, r(n)))p_e(n)\\ &\quad + v(B(\overline{X}, r(n))) \cap B(\overline{Y}, r(n)))p_e^2(n)\Big)^{n-1} d(m\times m)\\ &= n(n-1) \int_{S_1\times S_1} \Big(1 - \frac{\ln n + \omega(n)}{n}\\ &\quad - \frac{\ln n + \omega(n)}{n} + v(B(\overline{X}, r(n))) \cap B(\overline{Y}, r(n)))p_e^2(n)\Big)^{n-1}\\ &\quad d(m\times m) \leq e^{-2\omega(n)} \int_{S_1\times S_1} e^{v(B(\overline{X},r(n)))\cap B(\overline{Y},r(n)))p_e^2(n)(n-1)}\\ &\quad d(m\times m) = e^{-2\omega(n)} \int_{S_1} e^{v(B(\overline{O},r(n)))\cap B(\overline{Y},r(n)))p_e^2(n)(n-1)} dm(Y)\\ &= e^{-2\omega(n)} \int_{S_1\setminus\ B(\overline{O},2r(n))} e^{v(B(\overline{O},r(n)))\cap B(\overline{Y},r(n)))p_e^2(n)(n-1)} dm(Y)\\ &\quad + e^{-2\omega(n)} \int_{B(\overline{O},2r(n))} e^{v(B(\overline{O},r(n)))\cap B(\overline{Y},r(n)))p_e^2(n)(n-1)} dm(Y)\\ &= e^{-2\omega(n)}\\ &\quad + e^{-2\omega(n)} \int_{B(\overline{O},2r(n))} e^{v(B(\overline{O},r(n)))\cap B(\overline{Y},r(n)))p_e^2(n)(n-1)} dm(Y).\end{aligned} \tag{54}$$

Using the Laplace method for integrals and Lemma 1 we obtain

$$\Delta_n^1 = e^{-2\omega(n)} + O\left(\frac{e^{-(2-p_e(n))\omega(n)}}{n^{(2-p_e(n))}p_e(n)^4 r^2(n)}\right) \tag{55}$$

Using $p_e(n) \geq \frac{c}{\ln n}$ and $0 < \lim_{n\to\infty} \omega(n) < \infty$, we conclude

$$\lim_{n\to\infty} \Delta_n^1 < \infty. \tag{56}$$

Similarly, we can show $\lim_{n\to\infty} \Delta_n^2 < \infty$. Therefore, $\lim_{n\to\infty} \Delta_n = \Delta < \infty$, which concludes the theorem. □

Proof of Theorem 6

By a simple coupling argument, we find that the probability of having at least one isolated vertex is a decreasing function of $r(n)$. If $\alpha < 1$, then for any constant c and large enough n, we have

$$r(n) < \sqrt{\frac{\ln n + c}{\pi n p_e(n)}}. \tag{57}$$

Thus, by Theorem 4, the probability that $g = g(n, r, p_e)$ has at least one isolated vertex is asymptotically greater than or equal to $e^{-e^{-c}}$ for any real number c. Thus, if $\alpha < 1$, the graph $g = g(n, r, p_e)$ has an isolated vertex with high probability, and thus it is not k-connected for any positive integer k.

Now, by Theorem 1, it suffices to prove that if $\alpha > 1$, for any fixed $k \in \{0, 1, 2, \ldots\}$,$g(n, r, p_e)$ does not have any vertices of degree k with high probability. Let $\alpha > 1$ and $Y_{j,k,n}$ be the number of vertices of degree k in S_j, for $j = 1$, 2, 3. It suffices to show $Y_{j,k,n} = 0$ asymptotically almost surely for $j = 1$, 2, 3.

We first consider $Y_{1,k,n}$. We have

$$EY_{1,k,n} = n \int_{S_1} nk[v(B(\overline{X}, r(n)))p_e(n)]^k \times \left(1 - v(B(\overline{X}, r(n)))p_e(n)\right)^{n-k-1} dm(\overline{X}). \tag{58}$$

But for $\overline{X} \in S_1$, we have $v(B(\overline{X}, r(n))) = \pi r^2(n)$. Thus

$$EY_{1,k,n} = O\left(\frac{(\ln n)^k}{n^{\alpha-1}}\right) = o(1). \tag{59}$$

Therefore, $Y_{1,k,n} = 0$ asymptotically almost surely. We now consider $Y_{2,k,n}$. We have

$$EY_{2,k,n} = n \int_{S_2} nk[v(B(\overline{X}, r(n)))p_e(n)]^k \times \left(1 - v(B(\overline{X}, r(n)))p_e(n)\right)^{n-k-1} dm(\overline{X}). \tag{60}$$

Using the Laplace method for integrals, Lemma 1, and $p_e(n) \geq \frac{c}{\ln n}$ we can write

$$EY_{2,k,n} = O\left(\frac{(\ln n)^{2k+1}}{n^{\frac{\alpha}{2}-\frac{1}{2}-o(1)}}\right) = o(1) \tag{61}$$

This implies $Y_{2,k,n} = 0$ asymptotically almost surely. We now prove $Y_{3,k,n} = 0$ asymptotically almost surely. We note that

$$\begin{aligned} EY_{3,k,n} &\leq nr^2(n)nk(\pi r^2(n)p_e(n))^k \left(1 - \frac{\pi r^2(n)}{4} p_e(n)\right)^{n-k-1} \\ &\leq n^{k+1} r^{2k+2}(n) p_e(n) e^{-\frac{\pi r^2(n)}{4} p_e(n)(n-k-1)}. \end{aligned} \tag{62}$$

Using $p_e(n) \geq \frac{c}{\ln n}$ and $\lim_{n\to\infty} \left(\frac{n\pi r^2(n)p_e(n)}{\ln n}\right) = \alpha$, we conclude

$$EY_{3,k,n} = O\left(\frac{(\ln n)^{k+2}}{n^{\frac{1}{4}-o(1)}}\right) = o(1) \tag{63}$$

This implies that $Y_{3,k,n} = 0$ asymptotically almost surely.□

References

1. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002, August). A survey on sensor networks. *IEEE Communications Magazine, 40*, 102–114.
2. Akkaya, K., & Younis, M. (2004). A survey of routing protocols in wireless sensor networks. *Elsevier Ad Hoc Network Journal* (to appear).
3. Eschenauer, L., & Gligor, V. D. (2002, November). A key management scheme for distributed sensor networks. In *The 9th ACM conference on computer and communication security* (pp. 41–47).
4. Chan, H., Perrig, A., & Song, D. (2003). random key predistribution schemes for sensor networks. In *2003 IEEE symposium on research in security and privacy* (pp. 197–213).
5. Bollobás, B. (2001). *Random graphs* 2nd (ed.). Cambridge: Cambridge University Press.
6. Meester, R., & Roy, R. (1996). *Continuum percolation*. Cambridge: Cambridge University Press.
7. Penrose, M. D., & Pistztora, A. (1996). Large deviations for discrete and continous percolation. *Advances in Applied Probability, 28*, 29–52.
8. Penrose, M. D. (1993). On the spread-out limit for bond and continuum percolation. *Annals of Applied Probability, 3*(1), 253–276.
9. Penrose, M. D. (1997). The longest edge of the random minimal spanning tree. *The Annals of Applied Probability, 6*, 340–361.
10. Penrose, M. D. (1999). On k-connectivity for a geometric random graph. *Random Structures and Algorithms, 15*, 145–164.
11. Gupta, P. & Kumar, P. R. (1998). Critical power for asymptotic connectivity in wireless networks. In W. M. McEneany, G. Yin, & Q. Zhang (Eds.), *Stochastic analysis, control, optimization and applications: A volume in honor of W. H. Fleming* (pp. 547–566). Boston, MA: Birkhauser.
12. Gupta, P. & Kumar, P. R. (2000). The capacity of wireless networks. *IEEE Transactions on Information Theory, 46*(2), 388–404.
13. Xue, F., & Kumar, P. R. (2004). The number of neighbors needed for connectivity of wireless networks. *Wireless Networks, 10*(2),169–181.
14. Booth, L., Bruck, J., Franceschetti, M., & Meester, R. (2003, May). Covering algorithms, continuum percolation and the geometry of wireless networks. *Annals of Applied Probability, 13*.
15. Franceschetti, M., Booth, L., Cook, M., Bruck, J., & Meester, R. (2005). Continuum percolation with unreliable and spread out connections. *Journal of Statistical Physics, 118*(3/4), 721–734.
16. Shakkottai, S., Srikant, R., & Shroff, N. (2003, April). Unreliable sensor grids: Coverage, connectivity and diameter. In *The proceedings of IEEE INFOCOM'03*, San Francisco, CA.
17. Li, X. Y., Wan, P., Wang, Y., & Yi, C. W. (2003). Fault tolerant deployment and topology control in wireless networks. *ACM symposium on mobile ad hoc networking and computing, MOBIHOC*.
18. Wan P., & Yi, C. W. (2004). Asymptotic critical transmission radius and critical neighbor number for k-connectivity in wireless ad hoc networks. *ACM symposium on mobile ad hoc networking and computing, mobiHoc*.
19. Dousse, O., Thiran, P., & Hasler, M. (2002). Connectivity in ad-hoc and hybrid networks. *IEEE Infocom*.
20. Dousse, O. & Thiran, P. (2004). Connectivity vs capacity in dense ad hoc networks. *IEEE Infocom*.
21. Dousse, O., Bacelli, F., & Thiran, P. (2005). Impact of interferences on connectivity in ad hoc networks. *IEEE/ACM Transactions on Networking, 13*, 425–436.
22. Haas, Z., Halpern, J., & Li, L. (2002). Gossip-based ad hoc routing. *IEEE Infocom*.
23. Ganesan, D., Govindan, R., Shenker, S., & Estrin, D. (2002). Highly-resilient, energy-efficient multipath routing in wireless sensor networks. *Mobile computing and communications review (MC2R '02)*.
24. Nasipuri, A., & Das, S. R. (1999). On-demand multipath routing for mobile ad hoc networks. In *Proceedings of the IEEE international conference on computer communication and networks (ICCCN'99)*.

25. Ayanoglu, E., Gitlin, C., & Mazo, J. (1993). Diversity coding for transparent self-healing and fault-tolerant communication networks. *IEEE Transactions on Communications, 41*(11), 1677–1686.
26. Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2002, October). An application specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications, 1*, 660–670.
27. Intanagonwiwat, C. Govindan, R., & Estrin, D. (2000). Directed diffusion: A scalable and robust communication paradigm for sensor networks. *Mobile Computing and Networking*, 56–67
28. Manjeshwar, A. & Agrawal, D. P. (2002). APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. In *Proceedings of the international parallel and distributed processing symposium*.
29. Braginsky, D., & Estrin, D. (2002). Rumor routing algorithm for sensor networks. *ACM international workshop on wireless sensor networks and applications*.
30. Karp, B., & Hung, H. T. (2000). Gpsr: Greedy perimenter stateless routing for wireless network. In *Proceedings of the 6th annual ACM/IEEE international conference on mobile computing and networking*.
31. Penrose, M. (2003). *Random geometric graphs*. Oxford University Press.
32. Penrose, M. D. (1999). A strong law for the longest edge of the random minimal spanning tree. *The Annals of Applied Probability, 27*, 246–260.
33. Janson, S., Lucszak, T., & Rucinski, A. (2000). *Random graphs*. Wiley.
34. Heinzelman, W., Chandrakasan, A., & Balakrishnan, H. (2000, January). Energy-efficient communication protocol for wireless sensor networks. In *Proceedings of the Hawaii international conference system sciences*.

Author Biographies

Hossein Pishro-Nik received the M.S. and Ph.D degrees in Electrical and Computer Engineering from the Georgia Institute of Technology in 2003 and 2005 respectively. Currently, he is an assistant professor in the Electrical and Computer Engineering Department, University of Massachusetts Amherst. His research interests are in the general area of communications, in particular, error control coding and communication networks. Dr. Pishro-Nik received the National Science Foundation CAREER Award in 2009. He received ECE Graduate Research Assistant Excellence Award and CSIP Outstanding Research Award from the Georgia Institute of Technology in 2005.

Kevin Chan received the B.S. degree in Electrical and Computer Engineering and Engineering and Public Policy from Carnegie Mellon University, Pittsburgh, PA, in 2001, and the M.S. and Ph.D. degrees from the Georgia Institute of Technology, Atlanta, GA, in 2003 and 2008, respectively. He is currently an ORAU postdoctoral fellow at the United States Army Research Laboratory, Adelphi, MD. His research interests include cryptography and security services for wireless networks.

Faramarz Fekri received the National Science Foundation CAREER Award (2001), and Southern Center for Electrical Engineering Education (SCEEE) Research Initiation Award (2003), Outstanding Young faculty Award of the School of ECE (2006). Fekri's current research interests are in the area of communications and signal processing, in particular coding and information theory, information processing for wireless and sensor networks, and communication security. He is a Senior Member of the IEEE.