Threshold Behavior of Multi-Path Random Key Pre-Distribution for Sparse Wireless Sensor Network

Wei-Shuo Li\textsuperscript{a}, Chun-Wei Tsai\textsuperscript{a,b}, Min Chen\textsuperscript{c}, Wen-Shyong Hsieh\textsuperscript{a}, Chu-Sing Yang\textsuperscript{d}

Abstract

Wireless sensor network (WSN) has been an active research topic because its application encompasses various domains. In particular, a lot of attentions have been paid to the common feature of WSN to show that every node in a large enough network contains certain properties. Real-world applications of random key pre-distribution naturally involve geometric and combinatorial techniques and are even more challenging technically. This paper presents an efficient scheme, which can approximate a complex network by a much simpler object such that the approximation is “regular” between most pairs of partition of this network. Once a more traceable network is obtained, bounds for the probability of the property that a random key pre-distribution subgraph satisfies that each node has a path of length $\ell$ to its $\ell$-th-hop neighbors are established. Then, by using the sparse version of the Szemerédi’s regularity lemma and letting $C$ be a constant, $n$ the number of vertices, $p$ the probability of any two nodes sharing at least one common key, a sharp threshold $p \geq Cn^{-(\ell-1)/\ell}$ that satisfies this property is shown. Moreover, computer simulations are also given to show the performance of the proposed scheme.

Key words: Sensor network, security, random key pre-distribution.

1. Introduction

Wireless sensor network (WSN) has been a hot research topic over the past ten years because it provides solutions for various applications from traditional ones, such as temperature and humidity monitoring, analysis of the motion of tornadoes, fire detection, military defense [2, 24, 27, 28]. Due to the intrinsic features of small size, cheap price, and energy efficiency, the sensors of WSN are used in some hostile environments. To better understand whether traditional technologies can be applied to WSN or not, the differences between the WSN and the ad hoc network have been presented in [2]. For example, the topology of a sensor network are changing very frequently compared with ad hoc network. In the same study, Akyildiz et al. [2] also pointed out that the sensors of WSN are limited in power, computational capacity, memory, and storage. As a consequence, more efficient algorithms, applications, and policies for WSN are needed.

Recently, several studies [1, 2, 27, 30] have tried to discuss and review the research on WSN from different perspectives. In [30], Yick et al. divided WSN into five types: terrestrial WSN, underground WSN, underwater WSN, multimedia WSN, and mobile WSN, each of which can be adopted to the particular approach in question. Another approach to classifying the WSN is by using the sensor network protocol stack, such as physical layer, data link layer, network layer, transport layer, and application layer [2]. This model helps us differentiate the research issues of WSN. A good example in the physical layer is energy-efficient, which is somehow related to low-power studies. In [6], in order to analyze the security issue of WSN, Chen et al. further suggested to use the open system interconnect (OSI) model to analyze the WSN, which is similar to the above protocol stack model except that the OSI model adds the middleware layer.

In order to enhance the performance of WSN, previous works pointed out that the sensor network design is influenced by many factors such as fault tolerance, scalability and production cost [5]. Tilak et al. [27] presented several metrics to evaluate the performance of WSN that consists of energy-efficiency (system lifetime), latency, accuracy, fault-tolerance, and scalability. Two promising research directions that may affect the performance and reliability of WSN have been addressed. One is the routing protocols [1] while the other is security [6, 8, 29]. In [1], Akkaya and Younis raised the issue that the traditional routing protocols cannot be used for the WSN because they lack global addressing scheme (e.g., IP-based protocols). Also, the traditional routing protocols do not consider the energy cost. This kind of problems makes traditional wireless network technologies not suitable for the WSN.

Du and Xiao [8] have done a good survey on the security of the sensor network. They first introduced several attacks on the sensor network in terms of layers. Two well known attacks on the physical layer introduced by this study are jamming and...
tampering. The jamming attack uses a small number of randomly distributed jamming nodes to disrupt the network in such a way that all of the nodes in the sensor network cannot be served. The tampering attack means that the adversary (i.e., the attacker) can capture the sensor node and extract some important information by tampering with the node they captured. In [29], Wang et al. considered most of todays WSN nodes are, by design, not tamper-resistant because the cost of tamper-resistant sensor nodes is very high. For the same reason, when the adversary compromises a node, the whole sensor network will be compromised [8][29]. The random key pre-distribution scheme described herein avoids such a problem, by reducing the number of nodes compromised. Nevertheless, this may somehow affect the connectivity. The question is, how to balance such a tradeoff, i.e., the number of nodes compromised and the connectivity?

In this paper, we analyze the question of the threshold probability function for the property “every pair of vertices lies in a secure path” with respect to a security parameter \( p \). The goal is to obtain a threshold function \( p_0(n) \) such that if \( p \gg p_0 \), then the probability of a security subgraph having the above property is close to 1 whereas if \( p \) is slightly smaller than \( p_0 \), then the probability is close to 0. More precisely, we first obtain an \((\varepsilon)\)-regular partition of the sensor network by using the version of Szemerédi regularity lemma for the sparse graphs. Then, we cover each pair of nodes that is regular by a secure path \( P' \) of length \( \ell \). Finally, a threshold is obtained for almost all pairs of nodes (the so-called regularity nodes). Nevertheless, the main difference is in that in a random graph, the dependence and disjointness of certain events come naturally, whereas in an \((\varepsilon)\)-regular graph, one needs to choose an appropriate parameter. We show that in almost all security graphs, the number of secure paths of length \( \ell \) is close to its expectation provided that \( \varepsilon \) is sufficiently small, \( p \geq Cn^{(\ell-1)/\ell} \), and \( n \) is sufficiently large.

The remainder of the paper is organized as follows. Section 2 gives a brief introduction to key management and random key pre-distribution based schemes. Section 3 discusses and defines the problem we aim to solve in this paper. Section 4 gives the notation and definition of the Szemerédi’s regularity lemma. The main results are given in section 5. In Section 6, we give the detailed proof of our theorem described in this paper. The conclusion and further work are drawn in Section 8.

2. Related Works

Key management is apparently one of the most important security mechanisms to be addressed. In [6], Chen et al. categorized the key management protocols into key pre-distribution schemes, hybrid cryptography schemes, one way hash schemes, key injection schemes, and key management in hierarchy networks. Of them, key pre-distribution schemes is probably the most promising option because it requires to store only one master key in the sensor node. In other words, the communication overhead between the base station and the sensor nodes is much lower compared to all the centralized approaches [6].

Random key pre-distribution (RKP) based schemes [4][7][14][22] have been developed to guarantee the existence of specific properties, such as disjoint secure paths and disjoint secure cliques, to achieve a secure cooperation between nodes [14][22]. In the RKP WSN, such subgraph count problems are difficult to attack because there are several levels of dependencies among the problems.

Example 1. (Random Key Pre-distribution) Suppose we have a set of key identifiers \( I = \{id_i \mid 1 \leq i \leq K\} \) each of which is defined as \( id_i = f(k_i) \) where \( k_i \) is a key chosen at random from a key pool \( \mathcal{P} \) with \( |\mathcal{P}| = K \), and \( f \) is an arbitrary bijective function. Let \( \mathcal{K} \) and \( \mathcal{I} \) be collections of subsets of key pool \( \mathcal{P} \) and \( I \), respectively. Let \( \mathcal{R} = \{(R, I) \mid R \in \mathcal{K} \land I \subseteq I\} \) denote the set of collection of key rings. Then, node \( u \) will choose an \( R_u \in \mathcal{K} \) s.t. \( |R_u| = k \), randomly and uniformly (there may exist \( R_u = R_v \) for some \( u \neq v \)). Thus, each node contains a set of keys \( R_u \) and is identified by \( I_u = \{id_i \mid k_i \in R_u\} \).

In [8], Du et al. introduced several key pre-distribution schemes. Of them are fully pairwise key scheme, \( \lambda \)-secure key scheme, \( p \)-probabilistic key pre-distribution scheme, \( q \)-composite key pre-distribution scheme, random pairwise key scheme, and multi-space key scheme. Key pre-distribution schemes can also be classified according to the security model, be it deterministic or probabilistic, and the properties of the group keys. The primary threat of a security model is the ability of a compromised node to obtain a group key. The deterministic key establishment schemes can guarantee that a communication group \( C \subseteq \mathcal{I} \) is able to establish a common key (or pairwise key). In contrast, the probabilistic key establishment schemes of a communication group has a certain probability of being secure. This kind of group key schemes is dependent on the combinatorial methods. For instance,

Example 2. (Key Distribution Patterns) Mitchell et al. [20] generalized a proposal of Blom and considered a scheme in which the key distribution center (KDC) stores some global key pool \( \mathcal{P} \) with \( |\mathcal{P}| = K \) and issues to each node \( u \) a subset (block) \( B_u \subset \mathcal{P} \) of these keys. Now, if node \( u \) wishes to establish a secure link with node \( v \), the keys to be used are constructed from the set of keys contained in \( B_u \cap B_v \). The key idea in KDP is the property that no other nodes hold these keys. In symbols,

\[
\bigcap_{u \in A} B_u \not\subseteq \bigcup_{v \in A} B_v
\]  

(1)

On the other hand, the required keys are available to both nodes (or group \( A \)) but to no other single node (or other nodes outside \( A \)). Their proposed scheme is highly related to classical combinatorial problems in design theory or intersection set system such as Spener system and random covering designs. A known difficult problem in KDP is the collision problem by which attacking some members of some group compromises the secure link between nodes \( u \) and \( v \). A first probabilistic analysis can be seen in [9].

In this paper, our main concern is random key pre-distribution (RKP) based scheme. By comparing with [17][18], where we proposed a scheme to enhance the connectivity by using path-key, one can see here that we also presented a rather precise probabilistic model. The major difference is that in this
paper, we present a constructive approach, based on the algorithmic aspects of the regularity lemma, to solving the problem. For the other similar aspect using Szemerédi’s regularity lemma for the dense case, see [19].

3. The Problem Definition

In this section, we will discuss problems concerning subgraph count of large WSN. Many real-world applications, such as random key pre-distribution and multichannel assignment, naturally involve dependence between nodes. Thus, analyzing such a structure or combinatorial problem is complicated using classical wireless network models, such as percolation theories and random geometric graphs. Particularly, proofs of results in geometric setting models often blend stochastic geometric and combinatorial techniques and are more technically challenging. Graph models of RKP on wireless network usually consist of the following two sorts of graphs: random geometric graph and the intersection graph. We will discuss each of them in turn below.

Random Geometric Graphs (RGG). RGG has been a very influential and well-studied model for large wireless ad hoc and sensor networks [11][13][21][23]. For an increasing property $Q$ of random geometric graphs $G(X,r)$, one wants to find a tight upper bound on the smallest radius $r_Q(n)$ that will guarantee that $Q$ holds with high probability. The radius $r_Q(n)$ is called the critical radius if $Q$ exhibits a sharp threshold (also known as phase transition). One of the main differences between random graphs and random geometric graphs is in that $G(X,r_1)$ is a subset of $G(X,r_2)$ with high probability, when $r_2 = r_1 = o(1)$. The corresponding result does not hold for classical random graphs. The major difficulty in the study of random geometric graphs arises out of the geometry, and therefore points are related to one another in ways that depend greatly on their relative positions.

The success of the above graph models is due to the fact that there exists some “independent” perspective behind the analysis. For instance, when dealing with a counting problem, one may assume that $v_1,\ldots,v_n$ are randomly, uniformly, and independently scattered in a cube of volume $n/l$ in d-space, where $l$ is a Poisson parameter. Define $|C|$ to be the order of the component of $G(X,\lambda)$ containing $i$. It can be shown that the component of $G(X,\lambda)$ containing 0 has $k$ vertices.

$$P \left( \sum_{i=1}^{n} I[|C_i| = k] \sim p_k(\lambda) \right) \to 1. \tag{2}$$

On the other hand, the proportionate number of vertices of $G(X,\lambda)$ lying in components of order $k$ converges to $p_k(\lambda)$. The formula for $p_k(\lambda)$ was defined in [21] as

$$p_{k+1}(\lambda) = (k+1)\lambda \int_{(R^d)^{k+1}} h(x_1,\ldots,x_k) \exp(-\lambda A(0,x_1,\ldots,x_k)) dx_1\ldots dx_k, \tag{3}$$

where $h(x_1,\ldots,x_k) = 1$ if $G(0,x_1,\ldots,x_k)$ is connected, and $0,x_1,\ldots,x_k$ are in the lexicographical order; otherwise, $h(x_1,\ldots,x_k) = 0$. Moreover, $A(0,x_1,\ldots,x_k)$ is the volume of the union of $1$-balls centered at $0,x_1,\ldots,x_k$.

Random Intersection Graphs (RIG). Furthermore, the uniform random intersection graph $G(n,K)$ is a random graph as defined in [3]. Suppose that we have a vertex set $V$ and another set $W$. Let $v \in V$. We assign to each vertex $v$ a subset $W_v \subseteq W$, chosen uniformly at random from the $k$-subsets of $W$, and connect two vertices $v$ and $v'$ by an edge if and only if their assigned sets $W_v$ and $W_{v'}$ have non-empty intersection. The study of $G(n,K)$ is motivated by applications to random key pre-distribution on WSN.

The classic random key pre-distribution to accomplish this is due to Eschenauer and Gligor [10]. Each sensor node is loaded with $k$ distinct encryption keys, randomly taken from a pool of $K$ possible keys, before deployment. Two sensors can form a secure link if they are within wireless communication range and they share at least one encryption key. The uniform random intersection graph models this situation in the case when all sensors are within communication range. Most of the subsequent WSN literature model the uniform random intersection graph as a classical Erdős-Rényi random graph $G(n,p)$, a graph with $n$ vertices whose edges are chosen randomly and independently with a fixed probability $p$. They then use the asymptotic behavior of Erdős-Rényi random graphs to find good parameters for the scheme. For distinct nodes $u,v \in G(n,k,k)$, the probability that $uv$ being an edge is $p$, where the probability of key shared between two nodes is

$$p = 1 - ((K-k)!)^2/((K-2k)!K!) \tag{4}$$

as demonstrated in [4][14]. So the WSN literature models $G(n,k,k)$ by the Erdős-Rényi random graph $G(n,p)$ where $p \sim k^2/K$. Though simulations support this threshold, modeling $G(n,k,k)$ in this way is unsatisfactory since the behavior of $G(n,p)$ and $G(n,k,k)$ is sometimes radically different. Consequently, to investigate a subgraph count problem on RKP implies that the desired graph model must exploit the fact that we expect many more triangles in $G(n,k,k)$ than in $G(n,p)$ or $G(X,\lambda)$, especially when $k$ is small.

**Example 3.** Suppose that there are only two keys in the key pool, and each node randomly chooses the keys with a probability. Assume that the probability of any two nodes having a common key is $p$. What does the probability of a triangle occur? Recall that the answer is $p^3$ in Erdős-Rényi random graph. However, the exact answer to this question must be larger than $\frac{1}{2}p^2$ in RKP graph.

From an intuitive point of view, it seems complicated to trace $p_k(\lambda)$ combined with random intersection graph $G(n,k,k)$. This is due to the fact that in both models, the edges involve different levels of dependency. Thus, to overcome this problem, it uses an approximate pseudorandom graph in order to eliminate some properties that are difficult to handle.

The perspective of “dependency reduction” has been used in many similar literature. Therefore, our goal is to study pseudorandom graphs as models for networks with an additional tolerance $\varepsilon$ to control the dependency error, and this will govern our choice of results and insights we present here.
4. Preliminaries and Notations

The theory of random graphs and (ε)-regular graphs forms the basis of our work. We use v_G and v(G) to denote the number of vertices of a graph G and e_G and e(G) to denote the number of edges. We identify a path with ℓ + 1 endpoints by P^ℓ. For any vertex set W ⊆ V, let G[W] denote the graph induced by W. We denote the number of edges in G[W] by e_G(W) = e(W).

For any pair of possibly disjoint subsets U, W ⊆ V(G), let E_G(U, W) = E(U, W) denote the set of edges in G with one endpoint in U and the other endpoint in W. We also define e_G(U, W) := |E_G(U, W)|.

The expression Γ(v) refers to the neighborhood of a vertex v in G. Since we often have to consider the neighborhood into a subset W ⊆ V, we abbreviate Γ(v) ∩ W by Γ_w(v). Accordingly, we define d_w(v) := |Γ_w(v)|, which is the degree of vertex v into the set W. The neighborhood spanned by a set of vertices Q is denoted by Γ(Q); that is,

Γ(Q) := \bigcup_{v \in Q} Γ(v).

(5)

For neighborhoods inside specific partitions, we will use the abbreviations Γ_ℓ(v) := Γ(ℓ)(v) ∩ V_ℓ and d_ℓ(v) = |Γ_ℓ(v)|. Moreover, let Γ denote the neighborhoods in V_ε.

As we will frequently encounter throughout this work, density measures determine the growth rate of threshold functions, which are described later in Theorem 1. The most well-known density measure is

d(H) = \frac{e_H}{v_H}\frac{1}{2}\frac{e}{e_G}

(6)

for any subgraph H ⊆ V. In studying the asymptotic property of subgraph count, when H is considered as a path of length ℓ, we are interested in the following generalization of d.

d := \frac{e_H}{v_H} = \frac{\ell}{\ell - 1}.

(7)

We say that the pair (U, W) is (ε, p)-regular, if for all U' ⊆ U and W' ⊆ W with |U'| ≥ ε|U| and |W'| ≥ ε|W|, we have

|d_\rho(U', W') − d_\rho(U, W)| ≤ \epsilon.

(9)

Below, we shall sometimes use the expression (ε)-regular with respect to its density p to mean that (U, W) is an (ε, p)-regular pair. If B = (U, W, E) is a bipartite graph and (U, W) is an (ε, p)-regular pair, we say that B is an (ε, p)-regular bipartite graph.

Intuitively, a bipartite graph (V_1 ∪ V_2, E) is (ε)-regular if its edges are distributed in a random-like way. The parameter ε reflects the uniformity of this distribution. The smaller the ε, the more uniform the edge distribution in G.

Definition 1. For a path P^ℓ, let G(ℓ, n, m) be the family of graphs on vertex set V = ∪_{i∈[ℓ]} V_ℓ, where the sets V_ℓ are pairwise disjoint sets of vertices of size n, and edge set E = ∪_{i∈[ℓ]} E_{\ell, i} where E_{\ell, i} ⊆ V_\ell × V_\ell and |E_{\ell, i}| = m. Let G(ℓ, n, m, ε) ⊆ G(ℓ, n, m) denote the set of graphs in G(ℓ, n, m) satisfying that each (V_\ell ∪ V_\ell, E_{\ell, i}) is an (ε)-regular graph.

Definition 2. Let B(ℓ, n, m, δ) ⊆ G(ℓ, n, m) denote the subfamily consisting of all members that contain fewer than

(1 - \delta)n^{\ell+1}\left(\frac{m}{n}\right)^{\ell}

(10)
copies of P^\ell.

Observe that apart from the factor (1 - \delta), if the edges are randomly distributed between the vertex sets of any graph in G(ℓ, n, m), we expect

n^{\ell+1}\left(\frac{m}{n}\right)^{\ell}

(11)
copies of P^\ell in any member of this family.

5. Main Results

Given an arbitrary graph G (WSN) with n vertices (sensor nodes) and m edges (links), the behavior of the RKP sensor network can be taken as a random subgraph S_{G_r} = S(G, p) of G by keeping each edge of G with probability p independently.

Our main result of this paper is the proof of Theorem 2, a counting version of the KLR-Conjecture, in the case when H is the path P^\ell of size ℓ and m \geq Cn^{(\ell+1)/\ell}.

Theorem 1. (Bad graphs) For 3 ≤ ℓ ≤ log n and every δ, β > 0, there exists \varepsilon_0 such that for all ε ≤ \varepsilon_0, 1/ log n, there exist a constants C such that

|\mathcal{B}(ℓ, n, m, δ) \cap \mathcal{G}(ℓ, n, m)| \leq \beta\left(\frac{n}{m}\right)^{\ell},

(12)

for m \geq Cn^{(\ell+1)/\ell}.
Theorem 2. (Bad extension) For $3 \leq \ell \leq \log n$ and every $\delta, \beta > 0$, there exists $\varepsilon_0$ such that for all $\varepsilon \leq \varepsilon_0$, $p \gg 1/\log n$, there exists a constant $C$ such that

$$|G(\ell, n, dn^2, \varepsilon : \mathcal{B}(\ell, n, n, \varepsilon'))| \leq \beta^m\left(\frac{n^2}{m}\right)^{\ell},$$

for $m \geq Cn^{(\ell+1)/\ell}$.\(^{(13)}\)

Theorem 2 asserts that the number of bad graphs is very small while Theorem 3 states that there are at most $\beta^m\left(\frac{n^2}{m}\right)^{\ell}$ graphs belonging to $G(\ell, n, dn^2, \varepsilon)$ that could be extended by bad $(\varepsilon', p)$-regular graphs.

The following corollary asserts that for a given graph $G \in G(\ell, n, n, \varepsilon)$, at least $(1-\varepsilon)n$ vertices can establish a secure path to its $\ell$-hop neighbors.

Corollary 3. For $3 \leq \ell \leq \log n$ and every $\delta, \beta > 0$, there exists $\varepsilon_0$ such that for all $\varepsilon \leq \varepsilon_0$, $p \gg 1/\log n$, there exists a constant $C$ such that every RKP subgraph satisfying $p \geq Cn^{(\ell-1)/\ell}$ has all but at most $\varepsilon n$ nodes that can establish a secure path to its $\ell$-hop neighbors with probability $1 - o(1)$.

5.1. Counting Lemmas

The objects to be counted will be referred to as “bad,” because later we will identify them with substructures of graphs for which the occurrence of a complete subgraph cannot be guaranteed. These structures must be shown to occur very rarely in order to prove Corollary 3. Since almost all vertices in an $(\varepsilon, d)$-regular graph have $\Theta(m/n)$ neighbors, one can deduce that the neighborhood of $\Theta(n^2/m)$ vertices typically has size $\Theta(m/n \cdot n^2/m) = \Theta(n)$. Furthermore, the neighborhoods of the vertices have disjoint parts of size $\Theta(m/n)$; that is, every vertex contributes an equal part to the combined neighborhood of all vertices. In the sequel, we give a formal definition of this structure $\mathcal{E}$.

Definition 3. (Covering vertices) Let $P \subseteq V_i$. $I(P; v)$ be defined as an (ordered) subset of $P \subseteq V$ with cardinality such that the following property is satisfied. Let $P = \{v_1, \ldots, v_\ell\}$. Then, there exist pairwise disjoint sets $W_1 \subseteq \Gamma_j(v_1), \ldots, W_\ell \subseteq \Gamma_j(v_\ell)$ with $|W_1| = \cdots = |W_\ell| = q_{\varphi} := (1-\varepsilon)n$. For $P \subseteq V_i$ the sets $(W_i)_{i=1}^\ell$ are called covering neighborhoods.

Definition 4. (v-cover) Consider a graph $G \in G(\ell, n, m, \varepsilon)$ and $\nu > 0$. A set $P \subseteq V_i$ with $|P| = p_{\varphi} := \nu(3d)$ is called a v-cover of $V_j$ if $|I(P; v)| \geq (1-\varepsilon')p_{\varphi}$, such that $|\Gamma_j(v_i)| \geq (1-\varepsilon')n$.

Definition 5. (v-multicovers) Let $\nu, \varepsilon' > 0$. We call a set $P \subseteq V_i$ a v-multicover of $V_j$ if there exist pairwise disjoint subsets $P_1, \ldots, P_r \subseteq Q$ with $r := |Q|/p_{\varphi}$ and $|P_i| = \cdots = |P_j| = p_{\varphi}$ such that $P_i$ is a v-cover of $V_j$, $i = 1, \ldots, r$.

We define two kinds of bad structures contained in each $G \in \mathcal{B}(\ell, n, m, \varepsilon)$.

Definition 6. (Bad graphs) Let $\alpha > 0$.

(a) for $1 \leq i \leq \ell$, there exist sets $W_i \subseteq V_i$ of size $|W_i| \leq \alpha n$ such that for $1 \leq i \leq k$ and $1 \leq j \leq \alpha n/|x|$, there exist less than $(1-\alpha)n/|x|$ pairwise disjoint subsets $X_1, \ldots, X_{k+1}$ that form an $(\varepsilon', p)$-regular for each pair.

(b) for $\ell \leq i \leq \ell + 1$, there exists a set $X \subseteq V_i$ of size at least $\varepsilon n$ such that for all $q \geq C(\ell-1)/\ell$, at least $\beta^q$ are not v-multicover

Now we construct all possibilities of bad events for $\ell$-tuple $(V_1, \ldots, V_\ell)$. Let $Q_1$ be a subset in $V_1$. A set $\varepsilon''n \leq Q' \subseteq Q_1$ is called bad if $|\Gamma(Q', Q_2)| \leq (1-\delta)n$. The following lemma states that if we choose a bad set $Q$ with respect to $v$ within $V_1$, then the number of such sets that are allowed to be selected is very small.

Lemma 1. (Count (a)) For $\beta, \nu, \alpha > 0$, there exists $\varepsilon_0 = \varepsilon_0(\beta, \nu, \alpha) > 0$ such that for $\varepsilon \leq \varepsilon_0$, all but at most $\beta^q$ partitions of $V_i$ into subsets $Q'_i$ with size $(1+\varepsilon')x, 1 \leq j \leq \alpha n/|x|$ contain at most

$$\beta^q\left(\frac{n}{(1+\varepsilon')x}\right)^{\ell+1}$$

bad $\ell$-tuples $(Q'_1, \ldots, Q'_\ell)$.

Lemma 2. (Count (b)) For $\beta, \nu, \alpha > 0$, there exists a constant $\varepsilon_0 = \varepsilon_0(\beta, \nu, \alpha')$ such that for $\varepsilon \leq \varepsilon_0$, any pair $(V_i, V_j)$ satisfies the number of sets with size $s$ that are not v-multicover is at most

$$\beta^s\left(\frac{n}{s}\right)^{\ell+1}$$

provided that $p_{\varphi}(\ell-1) \leq n^2/4$ and $p_{\varphi} \leq s \leq q$.

Observe that if all families of disjoint sets $Q_i$ that satisfy a certain undesired property satisfy $\sum |Q_i| \leq \varepsilon n$, then we can delete at most $\varepsilon n$ vertices such that none of them have this bad property.

6. The Proofs

Theorem 3 is an easy consequence from Lemma 3 and Lemma 4, we omit its proof here. Theorem 4 is more complicated to prove. Our proof strategy is as follows. First, we define a family of bad graphs to be excluded from the family $\mathcal{G}(\ell, n, m, \varepsilon)$. Then, we shall show that there are merely a tiny fraction of all graphs in $\mathcal{B}(\ell, n, m, \varepsilon)$ that can be extended to an $(\varepsilon, d)$-regular graph.

Proof. (of Theorem 4) Let $x \geq \frac{n^2}{m} \log n$ and $\varepsilon' x^n \leq y \leq (1-\varepsilon')x^n$, $m$. In particular, Lemma 3 shows that all but a $\beta^q$ fraction of all tuples are good. We prove Theorem 4 by constructing all graphs that satisfy the above properties. First, we select $m$ edges between the $(\varepsilon', p)$-regular pairs $(V_i, V_j), 1 \leq i < j \leq \ell$. There are at most

$$\binom{n^2}{m}^{\ell-1}$$

(16)
ways to do that. Second, we choose pairwise disjoint sets $Q_i^j$, $j \leq \ell$. There are less than
\[ n^{\frac{\ell m}{x} + \frac{x}{x}} \leq n^{\frac{\ell m}{x} + \frac{1}{x}} \tag{17} \]
ways to do so.

Next we distribute the $y$ edges between the pair $(Q_i^j, Q_j^i)$. There are at most
\[
\prod_{1 \leq i < j \leq \ell - 1} \left( \frac{\ell m}{x} \right) \geq n^{\frac{\ell m}{x} + \frac{x}{x}} \tag{18}
\]
ways to choose the graph $G_i^j$.

Then, we want to choose extensions of $G_i^j$ to satisfy $(\varepsilon, d)$-regular. As these extensions have to belong to $\mathcal{G}(\ell - 1, x, d^2, \varepsilon)$, the numbers of ways to select the neighborhoods of vertices in $V_j$ is thus at most
\[
\prod_{v \in B_j} \left( \beta^n \left( \frac{n}{d(v)} \right) \right) \leq \beta^n n^{\frac{n}{d(v)}} \tag{19}
\]
Similarly, there are at most $\beta^n \left( \frac{n^2}{m} \right)$ possibilities to choose the pair $(V_j, V_{j+1})$. Then, we show that there are at most
\[
\beta^n \left( \frac{n^2}{m} \right) + \beta^n \left( \frac{n^2}{m} \right) = \beta^n \left( \frac{n^2}{m} \right) \leq \beta^n \left( \frac{n^2}{m} \right) \tag{20}
\]
ways to extend it to $V_{j+1}$.

Put all of these together, we obtain the upper bound
\[
\left( \frac{n}{m} \right)^{\ell - 1} n^{\frac{\ell m}{x} + \frac{x}{x}} \prod_{1 < i < j \leq \ell} \left( \frac{m}{x} \beta^n \left( \frac{n^2}{m} \right) \right) \leq 2^{\frac{\ell m}{x} + \frac{x}{x}} \beta^n \left( \frac{n^2}{m} \right) \tag{21}
\]
Let $\beta'' = 2^{\frac{\ell m}{x} + \frac{x}{x}} \beta^n$. We complete the proof.

Here we give a proof of Corollary 3.

Proof. First of all, let $\alpha$ and $\varepsilon'$ be given. We set
\[ \beta = \left( \frac{1}{n} \right)^{1/2} \frac{1}{\sqrt{\varepsilon'}} \left( \frac{1}{\varepsilon} \right)^{2/|d|} \cdot \tag{22} \]
By applying $\beta$ and $\varepsilon'$ to Theorem 2, we obtain $\varepsilon_0$ and $C$. Then let $\varepsilon = \min\{d/2, \varepsilon'/4, \varepsilon_0\}$. Suppose $V_1, V_2, \ldots, V_t$ is an $\ell$-partite partition graph from $\mathcal{B}(n, T, \varepsilon', \delta)$. We shall show that an RKP subgraph of $\mathcal{B}$ is unlikely to appear in $S_{G_i^j}$. Since the bipartite subgraph $S_{G_i^j}(V_j, V_j)$ contains at least $(d - \varepsilon)n^2$ edges each. By the definition of $\mathcal{F}$, then, there is a set $W \subset V_j$ with $|W| \leq \delta n$ such that $|V_j(V_j, \ldots, V_{j-1})| = \Pi^{d-1}(W) \leq (1 - \delta)\varepsilon n$. Then, Theorem 2, infers that there are at most
\[ \beta^n \left( \frac{n^2}{m} \right) \tag{23} \]
possibilities for choosing all such subgraphs. Now the number of all possible graphs belonging to $(\varepsilon, d)$-regular are at most
\[ \sum_{|W| \leq \delta n} \left( \frac{n^2}{m} \right) \left( \frac{n^2}{m} \right)^{d - \varepsilon n^2} \tag{24} \]
Hence, the probability of the random subgraph $S_G \in \mathcal{F}$ is bounded from above by
\[
\sum_{|W| \leq \delta n} \left( \frac{n^2}{m} \right) \beta^n \left( \frac{n^2}{m} \right)^{d - \varepsilon n^2} \beta^n \left( \frac{n^2}{m} \right)^{d - \varepsilon n^2} \leq \exp(2\ell \log n + \ell n \log N - 4\ell n^{1-1/\ell}) \tag{25}
\]
The result now follows as noted by Markov’s inequality.

Now we prove Lemma 1.

Proof. (of Lemma 1) Since we are supposed to build a bad pair of $(Q_1, Q_2)$ that satisfies condition (a), there are at least $q/2$ indices $i$ such that $|\Gamma(v_i, Q)| \leq (1 - \delta)x_i(q/2)$. Let $\delta \leq 2^{-\varepsilon\gamma\gamma n}$. Since there only $\delta x$ choices for $v_i$, the number of bad sets $Q_i'$ is at most
\[
\left( \frac{q}{|q/2|} \right)^{q/2} \left( \frac{x}{q/2} \right)^{q/2} \left( \frac{e}{\varepsilon'} \right)^{q/2} \left( \frac{\varepsilon'}{\varepsilon'\gamma n} \right) \leq \delta x \left( \frac{\varepsilon'}{\varepsilon'\gamma n} \right) \tag{26}
\]
where $\varepsilon' \gamma n \leq q/4 \leq n/4$. Since that subsets $Q_1', \ldots, Q_{t'}$ of size $x$ are chosen sequentially, any choice of a bad subset $Q_i'$ may depend on the previously chosen subsets $(Q_1', \ldots, Q_{i-1}')$. Moreover, at least $\delta \varepsilon x$ sets $Q_i'$ must be bad and do not depend on the order of
choice. Hence, the probability of a subset $Q \geq \varepsilon'x$ being bad is at most
\[
2 \left( \frac{\delta^{\varepsilon'n}|n|}{\binom{n}{x}} \right)^{\frac{an}{x}} \leq 2'\left(\varepsilon e\right)^{-\alpha x/2} \leq \delta^{\alpha x/2}
\]  
(26)

In other words, there are at most $\delta^{\alpha x/2}(\frac{n}{(1+\varepsilon')s})$ bad sets $Q$. Then, the proof follows.

Finally comes the proof of Lemma 2.

Proof. (of Lemma 2) Let $Q$ denote the family of sets of size $s$ in $V_i$ that is a $\gamma$-multicover of $V_j$, $1 \leq i < j \leq \ell + 1$. First, Lemma 1 states that the probability of a randomly chosen set $X \subseteq Q$ with $|X| = p_r$ is at most $\beta^{|s|\binom{n}{p_r}}$. Moreover, at least $\delta r$ sets $X_i$ must be bad and do not depend on the order of choice. Hence, we first count the number of sets $Q$ that contain $\delta r$ bad subsets. We start by partitioning the set $Q$ into $r := \lfloor s/p_r \rfloor$ disjoint sets of size $s$ and an additional set of size $s - rp_r$. Hence, we have at most
\[
\left( \frac{r}{\delta r} \right) \left( \frac{s^{|s/\nu|}}{\binom{n}{p_r}} \right) \left( \frac{n - rp_r}{s - rp_r} \right)
\]  
\[
\leq \left( \frac{\beta^{|s|/s}}{\beta^{|s|/p_r}} \right) \frac{1}{p_r!} \left( \frac{n}{s} \right)
\]  
(27)

ways to do that. Since there are $\frac{1}{p_r!} \binom{n}{s}$ ways to partition, the lemma follows.

7. Simulation

In this section, we show the threshold phenomena of our desired property “every nodes has a secure path to $\ell$-hop neighbors.”

For clarity, with a little modification to the above property, let A EVP denote the property “at least $1-\varepsilon$ nodes has a secure path to the $\ell$-hop neighbors.” One can see that the probability that satisfies A EVP is rapidly increased to 1. In what follows, $n$ is the number of nodes and $\ell$ is the average degree. Although our results in Section 5 has probability $1 - \alpha(1)$, one can observe that the probabilities in the simulation have exponential decay behaviors. The reason is similar to the different between binomial model with uniform model. In words, we fix the number of edges and choose each bad graph uniformly at random in our proofs. However, for simulations, we taking each pair of vertices to be an edge with probability $p$ independently. Consequently, the probabilities in the simulations are equivalent to those in the main results.

8. Conclusion and Future Work

In this paper, we showed that pseudorandom subgraphs with size smaller than the density of the given $\varepsilon$ graphs still inherit good property with high probability. Our main contribution is determined a desired RKP probability $p$ such that every vertex can find a secure path to its $\ell$-hop neighbors, provided that $p \geq Cn^{-(\ell-1)/\ell}$ and some appropriate constants are selected.
The most challenging part is when the sensor network is rather sparse. In this case, the number of all possible KRP subgraphs is too large to expand the exponential bounds. Thus approaches based on large deviation inequality generally seem to fail to provide the results we described here (see Rödl’s work for more details). To see why this may be expected, the expected value of path $P^\epsilon$ has the order of $O(n)$ in a sparse case, rather than $O(n^2)$ in the dense case, while the number of sets we need to control is $\exp(t(\Omega))$, and most common large deviation inequalities need such density $d$ in Theorem 1 to be large compared with $p$. In the future, it seems valuable to take some geometric properties into account. The WSN application is based on many previous researches. In recent years, lots of key distribution schemes for heterogeneous WSN are presented, thus modifying the regularity lemma to fit the heterogeneous/multiple environment could be more suitable for WSN.

References


