ISR TECHNICAL REPORT 2010-16

On random graphs associated with a pairwise key distribution scheme for wireless sensor networks (Extended version)

Osman Yagan, Armand M. Makowski

On random graphs associated with a pairwise key distribution scheme for wireless sensor networks (Extended version)

Osman Yağan and Armand M. Makowski
Department of Electrical and Computer Engineering
and the Institute for Systems Research
University of Maryland at College Park
College Park, Maryland 20742
oyagan@umd.edu, armand@isr.umd.edu

Abstract— **The pairwise key distribution scheme of Chan et al. was proposed as an alternative to the key distribution scheme of Eschenauer and Gligor (EG) to enable network security in wireless sensor networks. In this paper we consider the random graph induced by this pairwise scheme under the assumption of full visibility. We first establish a zero-one law for graph connectivity. Then, we discuss the number of keys needed in the memory of each sensor in order to achieve secure connectivity (with high probability). For a network of n sensors the required number of keys is shown to be on the order of $\log n$, a key ring size comparable to that of the EG scheme (in realistic scenarios).**

Keywords: Wireless sensor networks, Security, Key predistribution, Random graphs, Connectivity, Zero-one laws.

I. Introduction

Wireless sensor networks (WSNs) are distributed collections of sensors with *limited* computing and communications resources. Security is expected to be a key challenge for WSNs deployed in hostile environments where communications are monitored, and nodes are subject to capture and surreptitious use by an adversary. However, traditional key exchange and distribution protocols have been found inadequate for use in large-scale WSNs; see [7], [13], [15] for detailed discussions of some of the challenges.

Recently, *random* key predistribution schemes have been proposed to address some of these difficulties. The idea of randomly assigning secure keys to the sensor nodes prior to network deployment was first introduced by Eschenauer and Gligor [7]. The EG scheme, as we refer to it hereafter, has been investigated in the context of *random key graphs* by several authors [1], [4], [14], [18], [19]. Random key graphs are random graphs induced by the EG scheme under the assumption of full visibility, i.e., when nodes are all within communication range of each other. To be sure, the full visibility assumption does away with the wireless nature of the communication infrastructure supporting WSNs. In return, this simplification makes it possible to focus on how randomizing the key selections affects the establishment of a secure network, and the connectivity results for the underlying random key graph then provide helpful (though optimistic) guidelines to dimension the EG scheme.

Following the original work of Eschenauer and Gligor, a number of other key distribution schemes have been suggested. The q-composite scheme [3] is a variation on the EG scheme where two nodes need to share at least q keys (with $q > 1$) in order to establish a secure link between them. The q-composite scheme improves resiliency against small-scale attacks as the network becomes more vulnerable to large attacks. Du et al. [5] have proposed a key predistribution scheme which also improves resiliency but at the cost of increased overheads. Although these schemes somewhat improve network resiliency, they all fail to provide *perfect* resiliency against node capture attacks. Moreover, none of them enables a node to authenticate the identity of a neighbor with which it communicates. In terms of network security this is a major drawback because *node-to-node authentication* can help detect node misbehavior, and provides resistance against node replication attacks [3].

To address this last point, Chan et al. [3] have proposed a random pairwise key predistribution scheme with the following properties: (i) Even if some nodes are captured, the secrecy of the remaining nodes is *perfectly* preserved; (ii) Unlike earlier schemes, this pairwise scheme enables both node-to-node authentication and quorum-based node revocation. The pairwise distribution scheme can be implemented through the following *offline* construction: Before deployment, each of the n sensor nodes is paired (offline) with K distinct nodes which are randomly selected from amongst all other nodes. For each such pair of sensors, a unique (pairwise) key is generated and stored in the memory modules of each of the paired sensors along with the id of the other node. A secure link can then be established between two nodes if at least one of them is assigned to the other, i.e., if they have at least one pairwise key in common. Precise definitions and implementation details are given in Section II.

Let $\mathbb{H}(n; K)$ denote the random graph on the vertex set $\{1, \ldots, n\}$ where distinct nodes i and j are adjacent if they have at least a pairwise key in common; this corresponds to modelling the random pairwise distribution scheme under full visibility. The main goal of this paper is to give conditions on n and K under which $\mathbb{H}(n;K)$ is a connected graph with high probability as n grows large. As in the case of the EG scheme, such conditions might provide helpful guidelines for dimensioning purposes. In the original paper of Chan et al. [3] (as in the reference [9]), the connectivity of $\mathbb{H}(n;K)$ is analyzed by *equating* it with the Erdős-Renyi graph $\mathbb{G}(n;p)$ where $p = \frac{2K}{n}$; this constraint ensures that the link probabilities in the two graphs are asymptotically matched. A formal transfer of well-known connectivity results from Erdős-Renyi graphs to $\mathbb{H}(n;K)$ suggests that the parameter K should behave like $c \log n$ for some $c > \frac{1}{2}$ in order for $\mathbb{H}(n;K)$ to be connected with a probability approaching 1 for n large. With this conclusion as a point of departure, the maximum supportable networks size was evaluated [3], [9], and the random pairwise key predistribution scheme was deemed *not* scalable.

Here we show that transferring connectivity results from Erdős-Renyi graphs to $\mathbb{H}(n;K)$ leads to *misleading* conclusions. Indeed by a *direct* analysis we show the following zero-one law: With $K \geq 2$ (resp. $K = 1$), the probability that $\mathbb{H}(n;K)$ is a connected graph approaches 1 (resp. 0) as n grows large, and the desired connectivity is therefore achievable under very small values of K (much smaller than prescribed by the transfer from Erdős-Renyi graphs). Furthermore, at the connectivity threshold obtained here, i.e., when $K = 2$, we show that the expected degree of a node in $\mathbb{H}(n;K)$ is less than 4; this suggests a major difference from many classical random graph structures where the connectivity threshold appears when the expected node degree equals to $\log n$, e.g., see Erdős-Rényi graphs [2], random key graphs [1], [4], [14], [18], random intersection graphs [16] and random geometric graphs [12].

We then discuss the required number of keys to be kept in the memory module of each sensor in order to achieve secure connectivity. Since sensor nodes are expected to have very limited memory, it is crucial for a key distribution scheme to have *low* memory requirements [5]. In contrast with the EG scheme (and its variants), the key rings produced by the pairwise scheme of Chan et al. have variable size between K and $K + (n-1)$. Still, with the average size of a key ring being $2K$, we identify minimal conditions on how to scale the parameter K with the number n of nodes so that the size of any key ring hovers around $2K_n$ (in some probabilistic sense). Next, we show that the *maximum* key ring size is on the order $\log n$ with very high probability provided $K = O(\log n)$. Such a concentration result, together with the fact that very small K values suffice for the connectivity of $\mathbb{H}(n;K)$, points to the possibility of turning the pairwise scheme into a scalable one.

As with available results regarding the EG scheme based on random key graphs, the results given here under full visibility may lead to a dimensioning of the pairwise scheme which is too optimistic. This is due to the fact that the unreliable nature of wireless links has not been incorporated in the model. We do take a first step towards addressing this issue in the companion paper [22]; there the connectivity properties of the pairwise scheme are analyzed under a simplified communication model where unreliable wireless links are represented as on/off channels. Despite these limitations, the study of the random graph $\mathbb{H}(n;K)$ is nevertheless of independent interest as it models a very basic random pairing mechanism with potential applications in areas beyond wireless sensor networks, e.g., social; networks, where full visibility is not an issue.

We close by noting that this paper considers only the case when the sensor nodes $1, \ldots, n$ are all deployed at the same time. However, in practice the initially deployed network may have fewer than n nodes. In that case only a subset of $\{1, \ldots n\}$ will be deployed initially and the remaining sensor labels will be used at a later time if additional nodes are needed to be deployed. The implementation details and the connectivity results regarding the case where the network is deployed *gradually* are discussed in [21].

The rest of the paper is organized as follows: In Section II we give a formal model for the random pairwise distribution scheme of Chan et al. The random graph $\mathbb{H}(n;K)$ is contrasted against Erdős-Rényi graphs and regular random graphs in Section III. Results concerning connectivity are presented in Section IV, and properties of the key rings are discussed in Section V. Proofs can be found in Sections VI, VII and VIII.

A word on notation: All statements involving limits, including asymptotic equivalences, are understood with n going to infinity. The cardinality of any discrete set S is denoted by $|S|$. Also, we use the notation $=_{st}$ to indicate distributional equality.

II. The random pairwise scheme

The random pairwise key predistribution scheme of Chan et al. is parametrized by two positive integers n and K such that $K < n$. There are n nodes which are labelled $i = 1, \ldots, n$. with unique ids $\mathrm{Id}_1, \ldots, \mathrm{Id}_n$. Write $\mathcal{N} := \{1, \ldots n\}$ and set $\mathcal{N}_{-i} := \mathcal{N} - \{i\}$ for each $i = 1, \ldots, n$. With node i we associate a subset $\Gamma_{n,i}$ of nodes selected at *random* from $\mathcal{N}_{-i}$ – We say that each of the nodes in $\Gamma_{n,i}$ is paired to node i. Thus, for any subset $A \subseteq \mathcal{N}_{-i}$, we require

$$\mathbb{P}\left[\Gamma_{n,i} = A\right] = \begin{cases} \binom{n-1}{K}^{-1} & \text{if } |A| = K \\ 0 & \text{otherwise} \end{cases}$$

ensuring that the selection of $\Gamma_{n,i}$ is done *uniformly* amongst all subsets of $\mathcal{N}_{-i}$ which are of size exactly K. The rvs $\Gamma_{n,1}, \ldots, \Gamma_{n,n}$ are assumed to be mutually independent so that

$$\mathbb{P}\left[\Gamma_{n,i} = A_i, \ i = 1, \ldots, n\right] = \prod_{i=1}^{n} \mathbb{P}\left[\Gamma_{n,i} = A_i\right]$$

for arbitrary $A_1, \ldots, A_n$ subsets of $\mathcal{N}_{-1}, \ldots, \mathcal{N}_{-n}$, respectively.

Once this *offline* random pairing has been created, we construct the key rings $\Sigma_{n,1}, \ldots, \Sigma_{n,n}$, one for each node, as follows: Assumed available is a collection of nK distinct cryptographic keys $\{\omega_{i|\ell}, \ i = 1, \ldots, n; \ \ell = 1, \ldots, K\}$ – These keys are drawn from a very large pool of keys; in practice the pool size is assumed to be much larger than nK, and can be safely taken to be infinite for the purpose of our discussion.

Now, fix $i = 1, \ldots, n$ and let $\ell_{n,i} : \Gamma_{n,i} \rightarrow \{1, \ldots, K\}$ denote a labeling of $\Gamma_{n,i}$. For each node j in $\Gamma_{n,i}$ paired to i, the cryptographic key $\omega_{i|\ell_{n,i}(j)}$ is associated with j. For instance, if the random set $\Gamma_{n,i}$ is realized as $\{j_1, \ldots, j_K\}$ with $1 \leq j_1 < \ldots < j_K \leq n$, then an obvious labeling consists in $\ell_{n,i}(j_k) = k$ for each $k = 1, \ldots, K$ with key $\omega_{i|k}$ associated with node j_k. Of course other labeling are possible. e.g., according to decreasing labels or according to a random permutation. Finally, the pairwise key

$$\omega^\star_{n,ij} = [\mathrm{Id}_i | \mathrm{Id}_j | \omega_{i|\ell_{n,i}(j)}]$$

is constructed and inserted in the memory modules of both nodes i and j. Inherent to this construction is the fact that the key $\omega^\star_{n,ij}$ is assigned *exclusively* to the pair of nodes i and j, hence the terminology pairwise distribution scheme. The key ring $\Sigma_{n,i}$ of node i is the set

$$\Sigma_{n,i} := \{\omega^\star_{n,ij}, \ j \in \Gamma_{n,i}\} \cup \{\omega^\star_{n,ji}, \ i \in \Gamma_{n,j}\}. \qquad (1)$$

As mentioned earlier, under full visibility, two nodes, say i and j, can establish a secure link if at least one of the events $i \in \Gamma_{n,j}$ or $j \in \Gamma_{n,j}$ is taking place. Note that both events can take place, in which case the memory modules of node i and j both contain the distinct keys $\omega^\star_{n,ij}$ and $\omega^\star_{n,ji}$. It is also plain that by construction this scheme supports node-to-node authentication.

This pairwise distribution scheme naturally gives rise to the following class of random graphs: With $n = 2, 3, \ldots$ and positive integer $K < n$, we say that the distinct nodes i and j are adjacent, written $i \sim j$, if and only if they have at least one key in common in their key rings, namely

$$i \sim j \quad \text{iff} \quad \Sigma_{n,i} \cap \Sigma_{n,j} \neq \emptyset. \qquad (2)$$

Let $\mathbb{H}(n; K)$ denote the undirected random graph on the vertex set $\{1, \ldots, n\}$ induced by the adjacency notion (2). To keep the notation simple we have omitted the dependence on K for most of the quantities introduced so far. In what follows we largely abide by this practice, although we shall make the dependence on K explicit in a few places when scaling K with the number n of users.

III. Comparing with other random graphs

First some notation: Fix positive integers $n = 2, 3, \ldots$ and K with $K < n$. The edge assignments in the random graph $\mathbb{H}(n; K)$ are characterized by the $\{0, 1\}$-valued rvs $\{\xi_{n,ij}, \ j \in \mathcal{N}_{-i}, \ i = 1, \ldots, n\}$ defined by

$$\xi_{n,ij} := \mathbf{1}\left[i \in \Gamma_{n,j} \ \vee \ j \in \Gamma_{n,i}\right], \qquad \begin{array}{c} i \neq j \\ i, j = 1, \ldots, n \end{array}$$

with $\vee$ standing for logical disjunction. Thus, $\xi_{n,ij} = 1$ (resp. $\xi_{n,ij} = 0$) if i and j are adjacent (resp. not adjacent) in $\mathbb{H}(n; K)$, with $\xi_{n,ij} = \xi_{n,ji}$ by the undirected nature of the graph. In the calculations that follow we shall find it helpful to exploit the relation

$$1 - \xi_{n,ij} = \mathbf{1}\left[i \notin \Gamma_{n,j}, \ j \notin \Gamma_{n,i}\right]. \qquad (3)$$

Comparing with Erdős-Rényi graphs: Pick distinct $i, j = 1, \ldots, n$. It is plain that

$$\mathbb{P}\left[i \in \Gamma_{n,j}\right] = \frac{\binom{n-2}{K-1}}{\binom{n-1}{K}} = \frac{K}{n-1},$$

so that

$$\begin{aligned} \mathbb{P}\left[i \notin \Gamma_{n,j}, \ j \notin \Gamma_{n,i}\right] &= \mathbb{P}\left[i \notin \Gamma_{n,j}\right] \mathbb{P}\left[j \notin \Gamma_{n,i}\right] \\ &= \left(1 - \frac{K}{n-1}\right)^2 \qquad (4) \end{aligned}$$

by independence. As a result,

$$\mathbb{E}\left[\xi_{n,ij}\right] = 1 - \left(1 - \frac{K}{n-1}\right)^2. \qquad (5)$$

Put differently,

$$\mathbb{P}\left[i \sim j\right]_{n,K} = \frac{K}{n-1}\left(2 - \frac{K}{n-1}\right). \qquad (6)$$

Next, as we turn to the evaluation of correlations between edge assignment rvs, pick the vertices $i, j, k, \ell = 1, \ldots, n$ with $i \neq j$ and $k \neq \ell$. If the indices i, j, k and ℓ are all distinct, then by virtue of (3) the rvs $\xi_{n,ij}$ and $\xi_{n,k\ell}$ are independent, whence $\mathrm{Cov}[\xi_{n,ij}, \xi_{n,k\ell}] = 0$. It remains to consider the cases when the indices i, j, k and ℓ are *not* all distinct, e.g., without loss of generality, take the case $i = k$ with i, j and ℓ distinct. Then from (3) we get

$$\begin{aligned} &\mathrm{Cov}[\xi_{n,ij}, \xi_{n,i\ell}] \\ &= \mathrm{Cov}[1 - \xi_{n,ij}, 1 - \xi_{n,i\ell}] \\ &= \mathrm{Cov}[\mathbf{1}\left[i \notin \Gamma_{n,j}, \ j \notin \Gamma_{n,i}\right], \mathbf{1}\left[i \notin \Gamma_{n,\ell}, \ \ell \notin \Gamma_{n,i}\right]] \\ &= \mathbb{P}\left[i \notin \Gamma_{n,j}, \ j \notin \Gamma_{n,i}, \ i \notin \Gamma_{n,\ell}, \ \ell \notin \Gamma_{n,i}\right] \\ &\quad - \mathbb{P}\left[i \notin \Gamma_{n,j}, \ j \notin \Gamma_{n,i}\right] \mathbb{P}\left[i \notin \Gamma_{n,\ell}, \ \ell \notin \Gamma_{n,i}\right] \\ &= \mathbb{P}\left[i \notin \Gamma_{n,j}\right] \mathbb{P}\left[i \notin \Gamma_{n,\ell}\right] \mathbb{P}\left[j \notin \Gamma_{n,i}, \ \ell \notin \Gamma_{n,i}\right] \\ &\quad - \mathbb{P}\left[i \notin \Gamma_{n,j}, \ j \notin \Gamma_{n,i}\right] \mathbb{P}\left[i \notin \Gamma_{n,\ell}, \ \ell \notin \Gamma_{n,i}\right] \\ &= \mathbb{P}\left[i \notin \Gamma_{n,j}\right] \mathbb{P}\left[i \notin \Gamma_{n,\ell}\right] \mathbb{P}\left[j \notin \Gamma_{n,i}, \ \ell \notin \Gamma_{n,i}\right] \\ &\quad - \mathbb{P}\left[i \notin \Gamma_{n,j}\right] \mathbb{P}\left[j \notin \Gamma_{n,i}\right] \mathbb{P}\left[i \notin \Gamma_{n,\ell}\right] \mathbb{P}\left[\ell \notin \Gamma_{n,i}\right] \end{aligned}$$

by the independence of the rvs $\Gamma_{n,i}$, $\Gamma_{n,j}$ and $\Gamma_{n,\ell}$. Noting that

$$\mathbb{P}\left[j \notin \Gamma_{n,i}, \ \ell \notin \Gamma_{n,i}\right] = \frac{\binom{n-3}{K}}{\binom{n-1}{K}},$$

we easily conclude that

$$\begin{aligned} &\mathrm{Cov}[\xi_{n,ij}, \xi_{n,i\ell}] \qquad (7) \\ &= \left(\frac{\binom{n-2}{K}}{\binom{n-1}{K}}\right)^2 \left(\frac{\binom{n-3}{K}}{\binom{n-1}{K}} - \left(\frac{\binom{n-2}{K}}{\binom{n-1}{K}}\right)^2\right) < 0 \end{aligned}$$

by elementary calculations. It is now plain that the random graph $\mathbb{H}(n; K)$ is not an Erdős-Rényi graph [2] – Edge assignments are (negatively) correlated in $\mathbb{H}(n; K)$ while independent in Erdős-Rényi graphs.

In fact, the rvs $\{\xi_{n,ij}, \ j \in \mathcal{N}_{-i}, \ i = 1, \ldots, n\}$ turn out to exhibit a strong form of negative correlation in that they are *negatively associated* in the sense of Joag-Dev and Proschan [11]. To see this, consider the rvs

$$\eta_{n,ij} := \mathbf{1}\left[j \in \Gamma_{n,i}\right], \qquad \begin{matrix} i \neq j \\ i,j = 1,\dots,n. \end{matrix}$$

Under the enforced assumptions, it is clear that $\{\eta_{n,1j},\ j \in \mathcal{N}_{-1}\}, \{\eta_{n,2j},\ j \in \mathcal{N}_{-2}\}, \dots, \{\eta_{n,nj},\ j \in \mathcal{N}_{-n}\}$ are *independent* families of rvs, each of which is negatively associated [11, Example 3.2(c)]. More precisely, for each $i = 1,\dots,n$, the rvs $\{\eta_{n,ij},\ j \in \mathcal{N}_{-i}\}$ are negatively associated since $\Gamma_{n,i}$ represents a random sample (without replacement) of $\mathcal{N}_{-i}$. Thus, the entire collection of rvs $\{\eta_{n,ij},\ j \in \mathcal{N}_{-i},\ i = 1,\dots,n\}$ is negatively associated by the "closure under products" property of negative association [6, p. 35]. Now, for distinct $i,j = 1,\dots,n$ we note from (3) that

$$\begin{aligned} \xi_{n,ij} &= 1 - (1-\xi_{n,ij})(1-\xi_{n,ji}) \\ &= f(\eta_{n,ij}, \eta_{n,ji}) \end{aligned} \qquad (8)$$

with non-decreasing function $f : \mathbb{R}^2 \to \mathbb{R} : (x,y) \to 1-(1-x)(1-y)$. Hence, by the disjoint monotone aggregation property [6, p. 35] of negative association, the family of edge indicator rvs $\{\xi_{n,ij},\ 1 \leq i < j \leq n\}$ is also negatively associated. As a result, with $\mathcal{A} = \{\{i,j\} :\ 1 \leq i < j \leq n\}$, it is plain that

$$\mathbb{P}\left[i \sim j, \{i,j\} \in A\right]_{n,K} \leq \prod_{\{i,j\} \in A} \mathbb{P}\left[i \sim j\right]_{n,K}, \quad A \subseteq \mathcal{A}.$$

Comparing with random regular graphs: For each $i = 1,2,\dots,n$, let $D_{n,i}$ denote the degree of node i in the *undirected* graph $\mathbb{H}(n;K)$. We have

$$\begin{aligned} D_{n,i} &= \sum_{j=1, j\neq i}^{n} \mathbf{1}\left[i \in \Gamma_{n,j}\ \vee\ j \in \Gamma_{n,i}\right] \\ &= K + \sum_{j=1, j \notin \Gamma_{n,i}\cup\{i\}}^{n} \mathbf{1}\left[i \in \Gamma_{n,j}\right] \end{aligned} \qquad (9)$$

where we note that

$$|\{j = 1,\dots,n :\ j \notin \Gamma_{n,i} \cup \{i\}\}| = n - K - 1.$$

Therefore, by independence, the sum appearing in (9) is a binomial rv with $n-K-1$ trials and success probability $\frac{K}{n-1}$, whence

$$D_{n,i} =_{st} K + \text{Bin}\left(n-K-1, \frac{K}{n-1}\right). \qquad (10)$$

It is now plain that the nodes in $\mathbb{H}(n;K)$ have different (random) degree, and therefore $\mathbb{H}(n;K)$ is not a random regular graph [2, p. 50] [10, Chap. 9, p. 233].

IV. Connectivity

Fix positive integers $n = 2,3,\dots$ and $K < n$. Throughout we set

$$P(n;K) := \mathbb{P}\left[\mathbb{H}(n;K) \text{ is connected}\right].$$

The first technical result of this paper, given next, is established in Section VI; the proof adapts classical arguments used for proving the one law for connectivity in Erdős-Rényi graphs.

Theorem 4.1: With any positive integer $K \geq 2$, the bound

$$P(n;K) \geq 1 - \frac{(K+1)^{K^2-1}}{2} \cdot n^{-(K^2-2)} \qquad (11)$$

holds for all $n = 2,3,\dots$ sufficiently large, say $n \geq n(K)$ for some finite integer $n(K) > e(K+1)$ which depends on K.

The bound (11) gives some indication as to how fast the convergence $\lim_{n\to\infty} P(n;K) = 1$ occurs when $K \geq 2$, with the convergence becoming faster with larger K as would be expected; see also (13) below. Although the right handside of (11) may be negative for small values of n (in which case the bound is trivial), it is already active (i.e., positive) when $n = 2(K+1)$ (and beyond past $n(K)$).

For $K = 2$, the bound (11) takes the simpler form

$$P(n;2) \geq 1 - \frac{27}{2n^2}, \quad n \geq n(2). \qquad (12)$$

For each $n = 1,2,\dots$, a simple coupling argument yields the comparison

$$P(n;2) \leq P(n,K), \quad 2 \leq K < n. \qquad (13)$$

Making use of (12) we then conclude that

$$P(n;K) \geq 1 - \frac{27}{2n^2}, \quad n \geq \max(K, n(2)) \qquad (14)$$

for any $K \geq 2$.

A zero-one law for connectivity is presented next.

Theorem 4.2: With any positive integer K, it holds that

$$\lim_{n\to\infty} P(n;K) = \begin{cases} 0 & \text{if } K = 1 \\ & \\ 1 & \text{if } K \geq 2. \end{cases} \qquad (15)$$

The one-law in Theorem 4.2 is an easy consequence of the bound (11) (or (14)), while the zero-law of Theorem 4.2 is proved separately in Section VII.

Theorem 4.2 easily yields the behavior of graph connectivity as the parameter K is scaled with n. First some terminology: We refer to any mapping $K : \mathbb{N}_0 \to \mathbb{N}_0$ as a *scaling* provided it satisfies the natural conditions

$$K_n < n, \quad n = 1,2,\dots. \qquad (16)$$

Corollary 4.3: For any scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$, we have

$$\lim_{n\to\infty} P(n;K_n) = 1 \qquad (17)$$

provided $K_n \geq 2$ for all n sufficiently large.

Proof. For each $n = 1,2,\dots$, a simple coupling argument yields

$$P(n;K) \leq P(n,K'), \quad K < K' < n.$$

Under the scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$, it follows that $P(n;2) \leq P(n,K_n)$ for all n sufficiently large as soon as $K_n \geq 2$. Letting n go to infinity in this last inequality, we get (17) by invoking Theorem 4.2. ■

Because $\mathbb{H}(n;K)$ cannot be equated with an Erdős-Renyi graph, neither Theorem 4.1 nor Corollary 4.3 are consequences of classical results for Erdős-Renyi graphs [2]. Indeed, consider the following well-known zero-one law for Erdős-Rényi graphs: For any scaling $p : \mathbb{N}_0 \to [0,1]$ satisfying

$$p_n \sim c \cdot \frac{\log n}{n} \tag{18}$$

for some $c > 0$, it holds that

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{G}(n;p_n) \text{ is connected}\right] = \begin{cases} 0 & \text{if } \ 0 < c < 1 \\ \\ 1 & \text{if } \ 1 < c. \end{cases}$$

As seen from (6), $\frac{2K_n}{n-1} - \frac{K_n^2}{(n-1)^2}$ stands for the probability of link assignment in $\mathbb{H}(n;K_n)$ and therefore plays a role analogous to that of p_n in Erdős-Rényi graphs. Thus, a *transfer* of the connectivity results from $\mathbb{G}(n;p_n)$ to $\mathbb{H}(n;K_n)$ suggests scaling K such that

$$\frac{2K_n}{n-1} - \frac{K_n^2}{(n-1)^2} \sim c\frac{\log n}{n},$$

or equivalently

$$2K_n \sim c \log n \tag{19}$$

in the practically relevant case when $K_n = 0(n)$. This would then lead formally to the zero-one law

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{H}(n;K_n) \text{ is connected}\right] = \begin{cases} 0 & \text{if } \ 0 < c < 1 \\ \\ 1 & \text{if } \ 1 < c. \end{cases}$$

to hold under (19). Clearly, this yields the misleading conclusion that K_n has to behave like $c\log n$ for some $c > \frac{1}{2}$ for $\mathbb{P}\left[\mathbb{H}(n;K_n)\right]$ to be asymptotically almost surely connected– In fact, by Theorem 4.2 it is only needed to have $K_n \geq 2$.

Also, observe from (10) that

$$\begin{aligned} \mathbb{E}\left[D_{n,i}\right] &= K + (n-K-1)\frac{K}{n-1} \\ &= K\left(2 - \frac{K}{n-1}\right). \end{aligned} \tag{20}$$

Thus, when $K = 2$ the expected degree of a node in $\mathbb{H}(n;2)$ is less than 4. However, as can be seen from Theorem 4.2, the random graph $\mathbb{H}(n;K)$ is asymptotically almost surely connected. This already points out to a significant difference with many other random graph structures discussed in the literature where the threshold for connectivity appears when the expected node degree equals to $\log n$, e.g., Erdős-Rényi graphs [2], random key graphs [1], [4], [14], [18], random intersection graphs [16] and random geometric graphs [12].

To further drive this point, note the following: In many known classes of random graphs, the absence of isolated nodes and graph connectivity are asymptotically equivalent properties, e.g., Erdős-Rényi graphs [2], random geometric random graphs [12] and random key graphs [14], [17]. This equivalence, when it holds, is used to advantage by first establishing the zero-one law for the absence of isolated nodes, a step which is usually much simpler to complete with the help of the method of first and second moments [10, p. 55]. However, there are no isolated nodes in $\mathbb{H}(n;K)$ since each node has degree at least K. Thus, the class of random graphs studied here provides an example where graph connectivity and the absence of isolated nodes are not asymptotically equivalent properties; in fact this is what makes the proof of the zero-law more intricate.

V. Key ring sizes

Fix $n = 2,3,\ldots$ and positive integer K with $K < n$. For each $i = 1,2,\ldots,n$, node i is assigned a key ring $\Sigma_{n,i}$ whose size is given by

$$|\Sigma_{n,i}| = |\Gamma_{n,i}| + \sum_{j=1,\ j\neq i}^{n} \mathbf{1}\left[i \in \Gamma_{n,j}\right]. \tag{21}$$

This is a simple consequence of the definition (1), and should be contrasted with the definition (9) for the degree $D_{n,i}$ of node i. In the latter case, if both events $j \in \Gamma_{n,i}$ and $i \in \Gamma_{n,j}$ are realized, this produces *only* a unit contribution to both $D_{n,i}$ and $D_{n,j}$, although two distinct pairwise keys are generated for the nodes i and j (and both are included in the key rings). We also define the maximal key ring size as

$$M_n := \max_{i=1,\ldots,n} |\Sigma_{n,i}|.$$

It is easy to see that

$$|\Sigma_{n,i}| = K + B_{n,i} \tag{22}$$

where $B_{n,i}$ is the rv determined through

$$B_{n,i} := \sum_{j=1,\ j\neq i}^{n} \mathbf{1}\left[i \in \Gamma_{n,j}\right].$$

Under the enforced independence assumptions, the rv $B_{n,i}$ is a binomial rv $\text{Bin}(n-1, \frac{K}{n-1})$, with

$$\mathbb{E}\left[B_{n,i}\right] = (n-1)\cdot\frac{K}{n-1} = K$$

and

$$\text{Var}[B_{n,i}] = (n-1)\cdot\frac{K}{n-1}\cdot\frac{n-1-K}{n-1}.$$

As a result, $\mathbb{E}\left[|\Sigma_{n,i}|\right] = 2K$ and

$$\text{Var}[|\Sigma_{n,i}|] = K\left(1 - \frac{K}{n-1}\right).$$

It is now plain that

$$\begin{aligned} \mathbb{E}\left[\left|\frac{|\Sigma_{n,i}|}{\mathbb{E}\left[|\Sigma_{n,i}|\right]} - 1\right|^2\right] &= \frac{\text{Var}[|\Sigma_{n,i}|]}{\mathbb{E}\left[|\Sigma_{n,i}|\right]^2} \\ &= \frac{1}{4}\left(\frac{1}{K} - \frac{1}{n-1}\right) \end{aligned} \tag{23}$$

so that

$$\mathbb{E}\left[\left|\frac{|\Sigma_{n,i}|}{2K} - 1\right|^2\right] = \frac{1}{4}\left(\frac{1}{K} - \frac{1}{n-1}\right). \tag{24}$$

In general the key ring sizes satisfy the bounds

$$K \leq |\Sigma_{n,i}| \leq K + (n-1), \quad i = 1, \ldots, n. \tag{25}$$

We give minimal conditions on a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ to ensure that the key ring of a node has size roughly of the order (of its mean) $2K_n$ when n is large.

Lemma 5.1: For any scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$*, we have*

$$\frac{|\Sigma_{n,1}(K_n)|}{2K_n} \xrightarrow{P}_n 1$$

as soon as $\lim_{n\to\infty} K_n = \infty$.

Proof. Under the enforced assumptions, we have

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\frac{|\Sigma_{n,1}(K_n)|}{2K_n} - 1\right|^2\right] = 0$$

by the earlier calculations (24), and the result follows. ■

Thus, $|\Sigma_{n,1}(K_n)|$ fluctuates from K_n to $K_n + (n-1)$ with a propensity to hover about $2K_n$ when n is large under the conditions of Lemma 5.1. Next we provide a concentration result that quantifies how the maximal key ring size deviates from $2K_n$.

Theorem 5.2: Consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ *of the form*

$$K_n \sim \gamma \log n, \quad n = 2, 3, \ldots \tag{26}$$

with $\gamma > 0$*. If* $\gamma > \gamma^\star := (2\log 2 - 1)^{-1} \simeq 2.6$*, then there exists* $c(\gamma)$ *in the interval* $(0, \gamma)$ *such that*

$$\lim_{n\to\infty} \mathbb{P}\left[|M_n(K_n) - 2K_n| \geq c \log n\right] = 0 \tag{27}$$

whenever $c(\gamma) < c < \gamma$.

In the course of proving Theorem 5.2 in Section VIII, we also show that

$$\mathbb{P}\left[|M_n(K_n) - 2K_n| \geq c \log n\right] \leq 2n^{-h(\gamma;c)} \tag{28}$$

for all $n = 1, 2, \ldots$ whenever $c(\gamma) < c < \gamma$ with $h(\gamma; c) > 0$ specified at (57).

We present experimental results that validate Lemma 5.1 and Theorem 5.2: For fixed values of n and K we have constructed key rings according to the mechanism presented in Section II. For each pair of parameters n and K, the experiments have been repeated $1,000$ times yielding $1,000 \times n$ key rings for each parameter pair. The results are depicted in Figures 1-4 which show the key ring sizes according to their frequency of occurrence. The histograms in blue consider all of the produced $1,000 \times n$ key rings, while the histograms in white consider only the $1,000$ maximal key ring sizes, i.e., only the largest key ring among n nodes in an experiment.

It is immediate from Figures 1-4 that the key ring sizes tend to concentrate around $2K$, validating the claim of Lemma 5.1. As would be expected, this concentration becomes more evident as n gets large. It is also clear that, in almost all cases the maximum size of a key ring (out of n nodes) is less than $3K$ validating the claim of Theorem 5.2.

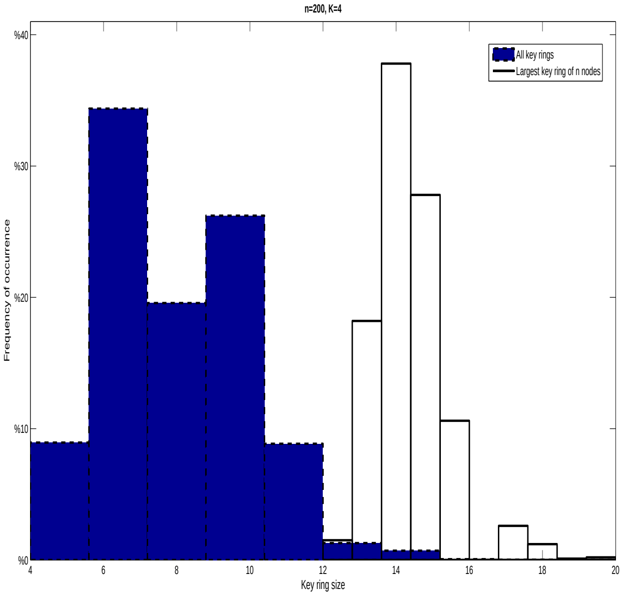

Fig. 1. Key ring sizes observed in $1,000$ experiments for $n = 200$ and $K = 4$ – Only 2% of the key rings are larger than $3K$ and the largest key ring has size 20.

VI. A proof of Theorem 4.1

Fix $n = 2, 3, \ldots$ and consider a positive integer K. The conditions

$$2 \leq K \quad \text{and} \quad e(K+1) < n \tag{29}$$

are assumed enforced throughout; the second condition is made to avoid degenerate situations which have no bearing on the final result. There is no loss of generality in doing so as we eventually let n go to infinity. In particular we

For any non-empty subset S of nodes, i.e., $S \subseteq \{1, \ldots, n\}$, we define the graph $\mathbb{H}(n;K)(S)$ (with vertex set S) as the subgraph of $\mathbb{H}(n;K)$ restricted to the nodes in S. We say that S is *isolated* in $\mathbb{H}(n;K)$ if there are no edges (in $\mathbb{H}(n;K)$) between the nodes in S and the nodes in the complement $S^c = \{1, \ldots, n\} - S$. This is characterized by the event $B_n(K;S)$ given by

$$B_n(K;S) := \cap_{i \in S} \cap_{j \in S^c} \left[i \notin \Gamma_{n,j},\ j \notin \Gamma_{n,i}\right].$$

Since each node in $\mathbb{H}(n;K)$ is connected to at least K other nodes, a set S can be isolated in $\mathbb{H}(n;K)$ only if $|S| \geq K+1$.

Also, we let $C_n(K;S)$ denote the event that the induced subgraph $\mathbb{H}(n;K)(S)$ is itself connected. Finally, we set

$$A_n(K;S) := C_n(K;S) \cap B_n(K;S).$$

The discussion starts with the following basic observation: If $\mathbb{H}(n;K)$ is *not* connected, then there must exist a subset S of nodes with $|S| \geq K+1$ such that $\mathbb{H}(n;K)(S)$ is connected while S is isolated in $\mathbb{H}(n;K)$. Thus, if $C_n(K)$ denotes the event that $\mathbb{H}(n;K)$ is connected, we have the inclusion

$$C_n(K)^c \subseteq \cup_{S \in \mathcal{P}_n:\ |S| \geq K+1}\ A_n(K;S) \tag{30}$$

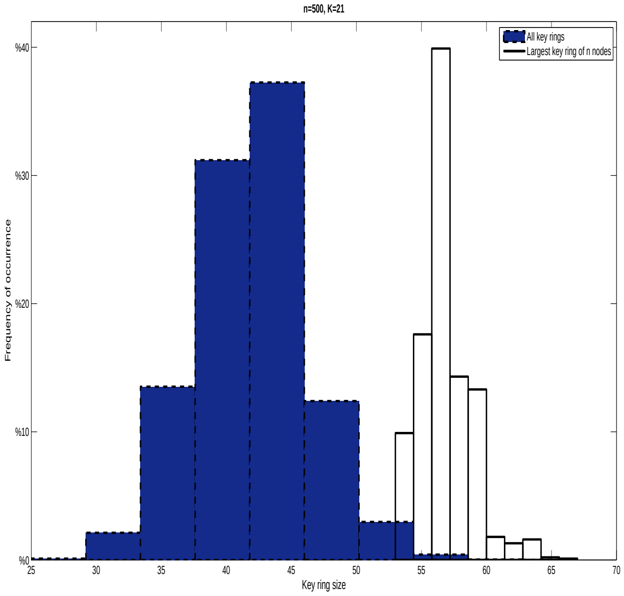

Fig. 2. Key ring sizes observed in $1,000$ experiments for $n = 500$ and $K = 21$ – Out of the $500,000$ key rings produced only 9 happened to be larger than $3K$ while the largest size observed is 67.

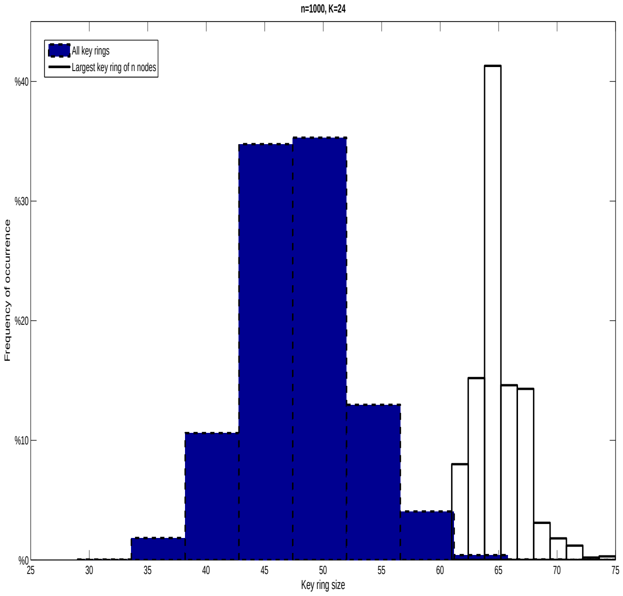

Fig. 3. Key ring sizes observed in $1,000$ experiments for $n = 1,000$ and $K = 24$ – $1,000,000$ key rings are produced. Only 5 of them happened to be larger than $3K$ and the largest observed key ring size is 75.

where $\mathcal{P}_n$ stands for the collection of all non-empty subsets of $\{1,\ldots,n\}$. A moment of reflection should convince the reader that this union need only be taken over all subsets S of $\{1,\ldots,n\}$ with $K+1 \le |S| \le \lfloor \frac{n}{2} \rfloor$. A standard union bound argument immediately gives

$$\begin{aligned}\mathbb{P}\left[C_n(K)^c\right] &\le \sum_{S\in\mathcal{P}_n : K+1\le|S|\le\lfloor\frac{n}{2}\rfloor} \mathbb{P}\left[A_n(K;S)\right] \\ &= \sum_{r=K+1}^{\lfloor\frac{n}{2}\rfloor}\left(\sum_{S\in\mathcal{P}_{n,r}} \mathbb{P}\left[A_n(K;S)\right]\right) \qquad (31)\end{aligned}$$

where $\mathcal{P}_{n,r}$ denotes the collection of all subsets of $\{1,\ldots,n\}$ with exactly r elements.

For each $r = 1,\ldots,n$, we simplify the notation by writing $A_{n,r}(K) := A_n(K;\{1,\ldots,r\})$, $B_{n,r}(K) := B_n(K;\{1,\ldots,r\})$ and $C_{n,r}(K) := C_n(K;\{1,\ldots,r\})$. For $r = n$, the notation $C_{n,n}(K)$ coincides with $C_n(K)$ as defined earlier. Under the enforced assumptions, it is a simple matter to check by exchangeability that

$$\mathbb{P}\left[A_n(K;S)\right] = \mathbb{P}\left[A_{n,r}(K)\right], \quad S\in\mathcal{P}_{n,r}$$

and the expression

$$\sum_{S\in\mathcal{P}_{n,r}} \mathbb{P}\left[A_n(K;S)\right] = \binom{n}{r}\,\mathbb{P}\left[A_{n,r}(K)\right] \tag{32}$$

follows since $|\mathcal{P}_{n,r}| = \binom{n}{r}$. Substituting into (31) we obtain the bounds

$$\mathbb{P}\left[C_n(K)^c\right] \le \sum_{r=K+1}^{\lfloor\frac{n}{2}\rfloor}\binom{n}{r}\,\mathbb{P}\left[B_{n,r}(K)\right] \tag{33}$$

as we note the inclusion $A_{n,r}(K)\subseteq B_{n,r}(K)$.

For each $r = K+1,\ldots,n$, it is easy to check that

$$\mathbb{P}\left[B_{n,r}(K)\right] = \left(\frac{\binom{r-1}{K}}{\binom{n-1}{K}}\right)^r \cdot \left(\frac{\binom{n-r-1}{K}}{\binom{n-1}{K}}\right)^{n-r}. \tag{34}$$

Reporting (34) into (33) we get

$$\mathbb{P}\left[C_n(K)^c\right] \le \sum_{r=K+1}^{\lfloor\frac{n}{2}\rfloor}\binom{n}{r}\left(\frac{\binom{r-1}{K}}{\binom{n-1}{K}}\right)^r\left(\frac{\binom{n-r-1}{K}}{\binom{n-1}{K}}\right)^{n-r}. \tag{35}$$

For $0 \le K \le x \le y$, we have

$$\frac{\binom{x}{K}}{\binom{y}{K}} = \prod_{\ell=0}^{K-1}\left(\frac{x-\ell}{y-\ell}\right) \le \left(\frac{x}{y}\right)^K$$

since $\frac{x-\ell}{y-\ell}$ decreases as ℓ increases from $\ell = 0$ to $\ell = K-1$. Using this fact into (35) together with the standard bound

$$\binom{n}{r} \le \left(\frac{ne}{r}\right)^r, \quad r = 1,\ldots,n$$

we conclude that

$$\begin{aligned}&\mathbb{P}\left[C_n(K)^c\right] \\ &\le \sum_{r=K+1}^{\lfloor\frac{n}{2}\rfloor}\left(\frac{ne}{r}\right)^r\left(\frac{r-1}{n-1}\right)^{rK}\left(1-\frac{r}{n-1}\right)^{K(n-r)} \\ &\le \sum_{r=K+1}^{\lfloor\frac{n}{2}\rfloor}\left(\frac{ne}{r}\right)^r\left(\frac{r}{n}\right)^{rK}\left(1-\frac{r}{n}\right)^{K(n-r)} \\ &\le \sum_{r=K+1}^{\lfloor\frac{n}{2}\rfloor}\left(\frac{ne}{r}\right)^r\left(\frac{r}{n}\right)^{rK} e^{-rK\frac{(n-r)}{n}}\end{aligned}$$

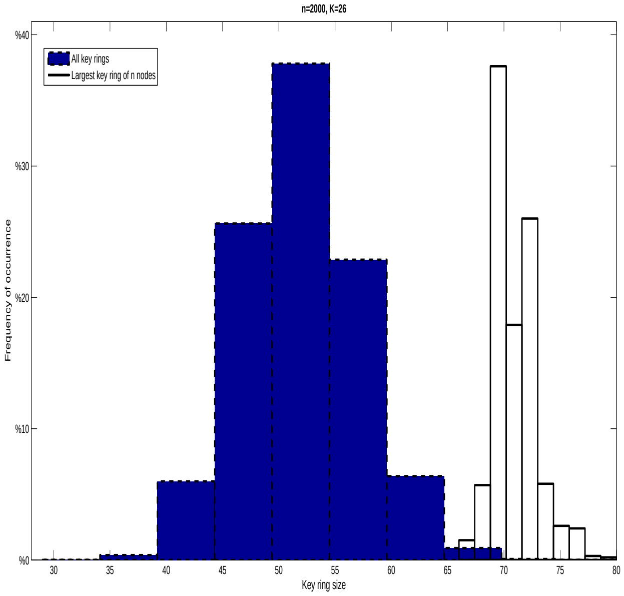

Fig. 4. Key ring sizes observed in $1,000$ experiments for $n = 2,000$ and $K = 26$ – Out of the 2000000 key rings produced only 2 happened to be larger than $3K$ the largest of them having 80 keys.

$$= \sum_{r=K+1}^{\lfloor \frac{n}{2} \rfloor} \left(\left(\frac{r}{n}\right)^{K-1} e^{1-K\frac{(n-r)}{n}} \right)^r . \qquad (36)$$

On the range $r = K+1, \ldots, \lfloor \frac{n}{2} \rfloor$ with $K \geq 2$, we have

$$K \ \frac{n-r}{n} \geq K \ \frac{n - \lfloor \frac{n}{2} \rfloor}{n} \geq \frac{K}{2} \geq 1,$$

whence

$$e^{1-K\frac{(n-r)}{n}} \leq 1.$$

Reporting this fact into (36) we find

$$\mathbb{P}\left[C_n(K)^c\right] \leq \sum_{r=K+1}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{r}{n}\right)^{r(K-1)} . \qquad (37)$$

For each $n = 1, 2, \ldots$, write

$$\left(\frac{x}{n}\right)^{x(K-1)} = e^{(K-1)f_n(x)}, \quad x \geq 1 \qquad (38)$$

with

$$f_n(x) = x\left(\log x - \log n\right).$$

It is plain that

$$f_n'(x) = 1 + \log x - \log n.$$

Therefore, $f_n(r)$ is monotone decreasing on the range $r = K+1, \ldots, \lfloor \frac{n}{e} \rfloor$ and monotone increasing on the range $r = \lfloor \frac{n}{e} \rfloor + 1, \ldots, \lfloor \frac{n}{2} \rfloor$, whence

$$f_n(r) \leq \max\left(f_n(K+1), f_n\left(\left\lfloor \frac{n}{2} \right\rfloor \right) \right)$$

for $r = K+1, \ldots, \lfloor \frac{n}{2} \rfloor$. It is also a simple matter to check by direct inspection that $f_n(K+1)$ is larger than $f_n\left(\lfloor \frac{n}{2} \rfloor\right)$ for n large enough, say $n \geq n(K)$ for some finite integer $n(K)$ which depends on K (and which can be taken to satisfy (29)). Using (38) together with the fact that

$$f_n(K+1) = (K+1)\log\left(\frac{K+1}{n}\right),$$

we obtain the equality

$$\max\left(\left(\frac{r}{n}\right)^{r(K-1)} : \ r = K+1, \ldots, \left\lfloor \frac{n}{2} \right\rfloor \right)$$
$$= \left(\frac{K+1}{n}\right)^{K^2-1} \qquad (39)$$

for all $n \geq n(K)$. Reporting (39) into (37), we conclude that

$$\mathbb{P}\left[C_n(K)^c\right] \leq \sum_{r=K+1}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{K+1}{n}\right)^{K^2-1} \leq \frac{n}{2} \cdot \left(\frac{K+1}{n}\right)^{K^2-1}$$

for all $n \geq n(K)$, and (11) is established. ■

VII. A proof of the zero-law in Theorem 4.2

First some terminology: When $K = 1$, the random sets $\Gamma_{n,1}, \ldots, \Gamma_{n,n}$ are now singletons, and can be interpreted as $\{1, \ldots, n\}$-valued rvs (as we do from now on) such that $\Gamma_{n,i} \neq i$ for each $i = 1, \ldots, n$. Thus, $\Gamma_{n,i}$ is the node selected at random which becomes associated (paired) with node i.

With this in mind, a *formation* is any sequence $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)$ such that for each $i = 1, \ldots, n$, the component γ_i is an element of $\{1, \ldots, n\}$ such that $\gamma_i \neq i$. In other words, $\boldsymbol{\gamma}$ is one of the $(n-1)^n$ possible realizations of the rvs $(\Gamma_{n,1}, \ldots, \Gamma_{n,n})$.

With any formation $\boldsymbol{\gamma}$ we associate a *directed* graph on the vertex set $\{1, \ldots, n\}$ in an obvious manner: There is a directed edge from node i to node j if $\gamma_i = j$. This directed graph is denoted by $H_{\boldsymbol{\gamma}}(n)$. As there are $(n-1)^n$ possible formations, there are $(n-1)^n$ distinct directed graphs so defined. Under the pairwise distribution scheme considered here, each of these graphs is equally likely, so that we have

$$P(n;1) = \frac{\sum_{\boldsymbol{\gamma}} \mathbf{1}\left[H_{\boldsymbol{\gamma}}(n) \text{ is connected } \right]}{(n-1)^n} \qquad (40)$$

where the summation $\sum_{\boldsymbol{\gamma}}$ is taken over all possible formations. Here, we have used the conventional notion of connectivity for directed graphs: A directed graph is connected if and only if the underlying *undirected* graph is connected – This is to be distinguished from the notion of *strong* connectivity defined for directed graphs. The desired zero-law will be established if we can show that

$$\lim_{n\to\infty} \frac{\sum_{\boldsymbol{\gamma}} \mathbf{1}\left[H_{\boldsymbol{\gamma}}(n) \text{ is connected } \right]}{(n-1)^n} = 0. \qquad (41)$$

From now on, let $H^\star_{\boldsymbol{\gamma}}(n)$ denote the underlying undirected graph of $H_{\boldsymbol{\gamma}}(n)$. We note that $H^\star_{\boldsymbol{\gamma}}(n)$ is a realization of the random graph $\mathbb{H}(n;1)$ when $(\Gamma_{n,1}, \ldots, \Gamma_{n,n}) = \boldsymbol{\gamma}$. For each formation $\boldsymbol{\gamma}$, we can easily validate the following observations:

1) By definition, $H^{\star}_{\boldsymbol{\gamma}}(n)$ is connected if and only if $H_{\boldsymbol{\gamma}}(n)$ is connected.
2) The undirected graph $H^{\star}_{\boldsymbol{\gamma}}(n)$ can have *at most* n edges since $H_{\boldsymbol{\gamma}}(n)$ has *exactly* n directed edges (as each of the n nodes has out-degree 1).
3) If $H_{\boldsymbol{\gamma}}(n)$ is connected (and hence $H^{\star}_{\boldsymbol{\gamma}}(n)$ is connected), then $H^{\star}_{\boldsymbol{\gamma}}(n)$ should have *at least* $n-1$ edges. In this case
   I. If $H^{\star}_{\boldsymbol{\gamma}}(n)$ has $n-1$, edges then $H^{\star}_{\boldsymbol{\gamma}}(n)$ is necessarily a *tree* and $H_{\boldsymbol{\gamma}}(n)$ has exactly one bi-directional edge.
   II. If $H^{\star}_{\boldsymbol{\gamma}}(n)$ has n edges, then $H_{\boldsymbol{\gamma}}(n)$ has exactly one *cycle*.

Case I – $\mathbb{H}(n;1)$ is connected and has $n-1$ edges: Thus, $\mathbb{H}(n;1)$ is a tree. With $\mathcal{T}_n$ denoting the collection of labelled trees on the set of vertices $\{1,\ldots,n\}$, we have $|\mathcal{T}_n| = n^{n-2}$ by Cayley's formula. Noting also that a given tree is the underlying undirected graph for $n-1$ different formations (corresponding to $n-1$ possible places for the single bi-directional edge), we get

$$\begin{aligned}
&\mathbb{P}\left[\mathbb{H}(n;1) \text{ is connected and has } n-1 \text{ edges}\right] \\
&\quad= \frac{1}{(n-1)^n}\cdot\sum\nolimits_{\boldsymbol{\gamma}} \mathbf{1}\left[\begin{array}{c} H_{\boldsymbol{\gamma}}(n) \text{ is connected and} \\ \text{has one bi-directional edge}\end{array}\right] \\
&\quad= \frac{1}{(n-1)^n}\cdot\sum\nolimits_{\boldsymbol{\gamma}} \sum_{T\in\mathcal{T}_n} \mathbf{1}\left[H^{\star}_{\boldsymbol{\gamma}}(n) = T\right] \\
&\quad= \frac{1}{(n-1)^n}\cdot(n-1)\cdot n^{n-2} \\
&\quad= \frac{1}{n}\cdot\left(\frac{n}{n-1}\right)^{n-1}. \qquad (42)
\end{aligned}$$

It is now clear that

$$\lim_{n\to\infty} \mathbb{P}\left[\begin{array}{c}\mathbb{H}(n;1) \text{ is connected} \\ \text{and has } n-1 \text{ edges}\end{array}\right] = 0. \qquad (43)$$

Case II – $\mathbb{H}(n;1)$ is connected and has n edges: This corresponds to all formations $\boldsymbol{\gamma}$ such that $H^{\star}_{\boldsymbol{\gamma}}(n)$ is connected and has exactly one cycle. It is not difficult to see that a connected graph with only one cycle can be the underlying undirected graph for two different formations (corresponding to the two possible orientations of the cycle). For instance, consider a connected graph on n nodes with exactly one cycle. This graph necessarily has n edges and therefore the original directed graph $H_{\boldsymbol{\gamma}}(n)$ cannot have a bi-directional edge. Without loss of generality, assume that the cycle consists of nodes $1, 2, 3, 4$ with edges $1\sim 2, 2\sim 3, 3\sim 4, 4\sim 1$. Then the two possible formations are $\{2,3,4,1,\gamma_5,\gamma_6,\ldots\gamma_n\}$ and $\{4,1,2,3,\gamma_5,\gamma_6,\ldots\gamma_n\}$. Similar arguments can be made for all possible cycles. Since there can be no other cycles or bi-directional edges in the rest of the graph, these two formations will be the only ones that give rise to that particular undirected structure.

Now let $\mathcal{T}^{+}_n$ denote the set of undirected graphs on n nodes which are connected and have exactly n edges. We find

$$\begin{aligned}
&\mathbb{P}\left[\mathbb{H}(n;1) \text{ is connected and has } n \text{ edges}\right] \\
&\quad= \frac{1}{(n-1)^n}\cdot\sum\nolimits_{\boldsymbol{\gamma}} \mathbf{1}\left[\begin{array}{c} H_{\boldsymbol{\gamma}}(n) \text{ is connected and} \\ \text{has exactly one cycle}\end{array}\right] \\
&\quad= \frac{1}{(n-1)^n}\cdot\sum\nolimits_{\boldsymbol{\gamma}} \sum_{G\in\mathcal{T}^{+}_n} \mathbf{1}\left[H^{\star}_{\boldsymbol{\gamma}}(n) = G\right] \\
&\quad= \frac{1}{(n-1)^n}\cdot 2\cdot|\mathcal{T}^{+}_n|. \qquad (44)
\end{aligned}$$

However, it is known [8, p. 133-134] that

$$|\mathcal{T}^{+}_n| \sim \frac{1}{4}\sqrt{2\pi}n^{n-\frac{1}{2}},$$

and reporting this fact into (44) gives

$$\begin{aligned}
&\mathbb{P}\left[\mathbb{H}(n;1) \text{ is connected and has } n \text{ edges}\right] \\
&\quad\sim \frac{\sqrt{2\pi}}{2}\left(\frac{n}{n-1}\right)^n n^{-\frac{1}{2}} \\
&\quad\sim \frac{\sqrt{2\pi}e}{2}n^{-\frac{1}{2}}. \qquad (45)
\end{aligned}$$

It is now immediate that

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{H}(n;1) \text{ is connected and has } n \text{ edges}\right] = 0.$$

Together with (43) and Facts 2-3, we now conclude that (41) holds. ■

VIII. A proof of Theorem 5.2

Fix the positive integers $n = 2, 3, \ldots$ and K with $K < n$. Using (22) we readily get

$$\left(\max_{i=1,\ldots,n} |\Sigma_{n,i}|\right) - 2K = \max_{i=1,\ldots,n} \left(B_{n,i} - K\right).$$

Therefore, with any given $t > 0$, we find

$$\begin{aligned}
&\mathbb{P}\left[\left|\left(\max_{i=1,\ldots,n} |\Sigma_{n,i}|\right) - 2K\right| > t\right] \\
&\quad= \mathbb{P}\left[\left|\max_{i=1,\ldots,n} \left(B_{n,i} - K\right)\right| > t\right] \\
&\quad= \mathbb{P}\left[\max_{i=1,\ldots,n} B_{n,i} > K + t\right] \\
&\qquad + \mathbb{P}\left[\max_{i=1,\ldots,n} B_{n,i} < K - t\right]. \qquad (46)
\end{aligned}$$

We take each term in turn. First a simple union argument shows that

$$\begin{aligned}
&\mathbb{P}\left[\max\nolimits_{i=1,\ldots,n} B_{n,i} > K + t\right] \\
&\quad= \mathbb{P}\left[\cup_{i=1}^{n} [B_{n,i} > K + t]\right] \\
&\quad\leq \sum_{i=1}^{n} \mathbb{P}\left[B_{n,i} > K + t\right] \\
&\quad= n\mathbb{P}\left[B_{n,1} > K + t\right] \qquad (47)
\end{aligned}$$

since the rvs $B_{n,1},\ldots,B_{n,n}$ are identically distributed (but not independent). Next we note that

$$\begin{aligned} \mathbb{P}\left[\max_{i=1,\ldots,n} B_{n,i} < K-t\right] &= \mathbb{P}\left[B_{n,i} < K-t,\ i=1,\ldots n\right] \\ &\leq \min_{i=1,\ldots,n} \mathbb{P}\left[B_{n,i} < K-t\right] \\ &= \mathbb{P}\left[B_{n,1} < K_n - t\right]. \end{aligned} \tag{48}$$

To proceed we recall standard bounds for the tails of binomial rvs [12, lemma 1.1, p. 16]: With

$$H(t) := 1 - t + t\log t,$$

we have the concentration inequalities

$$\mathbb{P}\left[B_{n.1} > K+t\right] \leq e^{-K\cdot H(\frac{K+t}{K})}$$

and

$$\mathbb{P}\left[B_{n,1} < K-t\right] \leq e^{-K\cdot H(\frac{K-t}{K})}$$

where the additional condition $0 < t < K$ is required for the second inequality to hold. Simple calculations on the appropriate ranges show that

$$-K\cdot H\left(\frac{K\pm t}{K}\right) = \pm t - (K\pm t)\cdot\log\left(1\pm\frac{t}{K}\right).$$

Thus, by the first concentration inequality, we conclude from (47) that

$$\mathbb{P}\left[\max_{i=1,\ldots,n} B_{n,i} > K+t\right] \leq e^{A_n(K;t)} \tag{49}$$

with

$$A_n(K;t) := \log n + t - (K+t)\cdot\log\left(1+\frac{t}{K}\right).$$

The second concentration inequality and (48) together yield

$$\mathbb{P}\left[\max_{i=1,\ldots,n} B_{n,i} < K-t\right] \leq e^{B_n(K;t)} \tag{50}$$

with

$$B_n(K;t) := -t - (K-t)\cdot\log\left(1-\frac{t}{K}\right)$$

under the additional constraint $0 < t < K$.

Now consider a scaling $K : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ of the form (26) for some $\gamma > 0$, and select the sequence $t : \mathbb{N}_0 \rightarrow \mathbb{R}_+$ given by

$$t_n = clogn, \quad n = 1,2,\ldots$$

with c in the interval $(0,\gamma)$ (so that $0 < t_n < K_n$ for all n sufficiently large). Under appropriate

Under appropriate conditions on γ and c, we shall show that

$$\lim_{n\rightarrow\infty} A_n(K_n;t_n) = -\infty \tag{51}$$

and

$$\lim_{n\rightarrow\infty} B_n(K_n;t_n) = -\infty. \tag{52}$$

The convergence statements

$$\lim_{n\rightarrow\infty} \mathbb{P}\left[\max_{i=1,\ldots,n} B_{n,i}(K_n) > K_n + t_n\right] = 0$$

and

$$\lim_{n\rightarrow\infty} \mathbb{P}\left[\max_{i=1,\ldots,n} B_{n,i}(K_n) < K_n - t_n\right] = 0$$

then follow from (49) and (50), respectively, and the desired conclusion (27) flows from (46).

With the selections made above, we get $A_n(K_n;t_n) \sim a(\gamma;c)\log n$ and $B_n(K_n;t_n) \sim b(\gamma;c)\log n$ with coefficients $a(\gamma;c)$ and $b(\gamma;c)$ given by

$$a(\gamma;c) := 1 + c - (\gamma + c)\cdot\log\left(1+\frac{c}{\gamma}\right), \quad c > 0$$

and

$$b(\gamma;c) := -c - (\gamma - c)\cdot\log\left(1-\frac{c}{\gamma}\right), \quad 0 < c < \gamma.$$

Thus, in order to ensure (51) and (52), we need to find c in the interval $(0,\gamma)$ such that $a(\gamma;c) < 0$ and $b(\gamma;c) < 0$, respectively. To that end, we first note that

$$\frac{\partial a}{\partial c}(\gamma;c) = -\log\left(1+\frac{c}{\gamma}\right) < 0, \quad c > 0$$

and

$$\frac{\partial b}{\partial c}(\gamma;c) = \log\left(1-\frac{c}{\gamma}\right) < 0, \quad 0 < c < \gamma.$$

Therefore, both mappings $c \rightarrow a(\gamma;c)$ and $c \rightarrow b(\gamma;c)$ are *strictly* decreasing on the intervals $(0,\infty)$ and $(0,\gamma)$, respectively. Since $\lim_{c\downarrow 0} b(\gamma;c) = 0$, it is plain that $b(\gamma;c) < 0$ on the entire interval $(0,\gamma)$. On the other hand, it is easy to check that $\lim_{c\downarrow 0} a(\gamma;c) = 1$ and

$$\lim_{c\uparrow\gamma} a(\gamma;c) = 1 - \gamma\left(2\log 2 - 1\right) = 1 - \frac{\gamma}{\gamma^\star}.$$

Hence, if we select $\gamma > \gamma^\star$, then $a(\gamma;c) < 0$ for all $c > c(\gamma)$ where $c(\gamma)$ is the unique solution to the equation

$$a(\gamma;c) = 0, \quad c > 0. \tag{53}$$

Uniqueness is a consequence of the strict monotonicity mentioned earlier.

The proof will be completed by showing that the constraint

$$c(\gamma) < \gamma, \quad \gamma > \gamma^\star \tag{54}$$

indeed holds. For each $\gamma > 0$, define the quantity $x(\gamma) := \frac{c(\gamma)}{\gamma}$. In view of (53) it is the unique solution to the equation

$$\frac{1}{\gamma} + x - (1+x)\log\left(1+x\right) = 0, \quad x > 0. \tag{55}$$

This equation is equivalent to

$$\frac{1}{\gamma} = \varphi(x), \quad x > 0 \tag{56}$$

where the mapping $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is given by

$$\varphi(x) = (1+x)\log\left(1+x\right) - x, \quad x \geq 0.$$

This mapping $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is strictly monotone increasing with $\lim_{x\downarrow 0}\varphi(x) = 0$ and $\lim_{x\uparrow\infty}\varphi(x) = \infty$, so that φ is a bijection from $\mathbb{R}_+$ onto itself. It then follows from (56) that $x(\gamma)$ is strictly decreasing as γ increases. Since $\varphi(1) = \left(\gamma^\star\right)^{-1}$, we get $x(\gamma^\star) = 1$ by uniqueness, whence $x(\gamma) < x(\gamma^\star) = 1$ for $\gamma > \gamma^\star$, a statement equivalent to (54).

Careful inspection of the proof shows that (28) holds with

$$h(\gamma;c) := -\max\left(a(\gamma;c), b(\gamma;c)\right) \tag{57}$$

on the range $c(\gamma) < c < \gamma$, and it is clear from the discussion above that $h(\gamma;c) > 0$ when $\gamma > \gamma^\star$. ■

ACKNOWLEDGMENT

This work was supported by NSF Grant CCF-07290. We thank Dr. H. Chan of CyLab at Carnegie Mellon University for pointing out reference [3] and for some insightful comments concerning this work. We also thank Dr. A. Barg of the University of Maryland for reference [8].

REFERENCES

[1] S.R. Blackburn and S. Gerke, “Connectivity of the uniform random intersection graph,” *Discrete Mathematics* **309** (2009), pp. 5130-5140.

[2] B. Bollobás, *Random Graphs*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge (UK), 2001.

[3] H. Chan, A. Perrig, D. Song, “Random Key Predistribution Schemes for Sensor Networks,” in Proceedings of the 2003 IEEE Symposium on Research in Security and Privacy (SP 2003), Oakland (CA), May 2003, pp. 197-213.

[4] R. Di Pietro, L.V. Mancini, A. Mei, A. Panconesi and J. Radhakrishnan, “Redoubtable sensor networks,” *ACM Transactions on Information Systems Security* **TISSEC 11** (2008), pp. 1-22.

[5] W. Du, J. Deng, Y.S. Han and P.K. Varshney, “A pairwise key predistribution scheme for wireless sensor networks,” in Proceedings of the 10th ACM Conference on Computer and Communications Security (CCS 2003), Washington (DC), October 2003, pp. 42-51.

[6] D.P. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms*, Cambridge University Press, Cambridge (UK), 2009.

[7] L. Eschenauer and V.D. Gligor, “A key-management scheme for distributed sensor networks,” in Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS 2002), Washington (DC), November 2002, pp. 41-47.

[8] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge (UK), January 2009.

[9] J. Hwang and Y. Kim, “Revisiting random key pre-distribution schemes for wireless sensor networks,” in Proceedings of the Second ACM Workshop on Security of Ad Hoc And Sensor Networks (SASN 2004), Washington (DC), October 2004.

[10] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, 2000.

[11] K. Joag-Dev and F. Proschan, “Negative association of random variables, with applications,” *The Annals of Statistics* **11** (1983), pp. 266-295

[12] M.D. Penrose, *Random Geometric Graphs*, Oxford Studies in Probability **5**, Oxford University Press, New York (NY), 2003.

[13] A. Perrig, J. Stankovic and D. Wagner, “Security in wireless sensor networks,” *Communications of the ACM* **47** (2004), pp. 53–57.

[14] K. Rybarczyk “Diameter, connectivity and phase transition of the uniform random intersection graph,” Submitted to *Discrete Mathematics*, July 2009.

[15] D.-M. Sun and B. He, “Review of key management mechanisms in wireless sensor networks,” *Acta Automatica Sinica* **12** (2006), pp. 900-906.

[16] K.B. Singer, *Random Intersection Graphs*, Ph.D. Thesis, Department of Mathematical Sciences, The Johns Hopkins University, Baltimore (MD), 1995.

[17] O. Yağan and A.M. Makowski, “On the random graph induced by a random key predistribution scheme under full visibility,” in Proceedings of the IEEE International Symposium on Information Theory (ISIT 2008), Toronto (ON), June 2008.

[18] O. Yağan and A.M. Makowski, “Connectivity results for random key graphs,” in Proceedings of the IEEE International Symposium on Information Theory (ISIT 2009), Seoul (S. Korea), June 2009.

[19] O. Yağan and A.M. Makowski, “Zero-one laws for connectivity in random key graphs,” Available online at arXiv:0908.3644v1 [math.CO], August 2009. Earlier draft available online at http://www.lib.umd.edu/drum/handle/1903/8716, January 2009.

[20] O. Yağan and A. M. Makowski, “On random graphs associated with a pairwise key distribution scheme for wireless sensor networks (Extended version).” Available online at http://www.lib.umd.edu/drum/

[21] O. Yağan and A. M. Makowski, “On the gradual deployment of random pairwise key distribution schemes,” submitted for inclusion in the program of Infocom 2010, Shanghai (PRC), June 2010. Available online at http://www.lib.umd.edu/drum/

[22] O. Yağan and A. M. Makowski, “Modeling the pairwise key distribution scheme in the presence of unreliable links,” submitted for inclusion in the program of Infocom 2010, Shanghai (PRC), June 2010. Extended version available online at http://www.lib.umd.edu/drum/