ABSTRACT

Title of dissertation: RANDOM GRAPH MODELING OF KEY DISTRIBUTION SCHEMES IN WIRELESS SENSOR NETWORKS

Osman Yağan, Doctor of Philosophy, 2011

Dissertation directed by: Professor Armand M. Makowski
Department of Electrical and Computer Engineering

Wireless sensor networks (WSNs) are distributed collections of sensors with limited capabilities for computations and wireless communications. It is envisioned that such networks will be deployed in hostile environments where communications are monitored, and nodes are subject to capture and surreptitious use by an adversary. Thus, cryptographic protection will be needed to ensure secure communications, as well as to support sensor-capture detection, key revocation and sensor disabling. Recently, random key predistribution schemes have been introduced to address these issues, and they are by now a widely accepted solution for establishing security in WSNs.

The main goal of the dissertation is to investigate and compare two *popular* random key predistribution schemes, namely the Eschenauer-Gligor (EG) scheme and the pairwise key distribution scheme of Chan, Perrig and Song. We investigate both schemes through their induced random graph models and develop scaling laws that corresponds to desirable network properties, e.g., absence of secure nodes that are

isolated, secure connectivity, resiliency against attacks, scalability, and low memory load – We obtain conditions on the scheme parameters so that these properties occur with high probability as the number of nodes becomes large. We then compare these two schemes explaining their relative advantages and disadvantages, as well as their feasibility for several WSN applications.

In the process, we first focus on the "full visibility" case, where sensors are all within communication range of each other. This assumption naturally leads to studying the random graph models induced by the aforementioned key distribution schemes, namely the *random key graph* and the *random pairwise graph*, respectively. In a second step, we remove the assumption of full visibility by integrating a wireless communication model with the random graph models induced under full visibility. We study the connectivity of WSNs under this new model and obtain conditions (for both schemes) that lead to the secure connectivity of the network.

RANDOM GRAPH MODELING OF KEY DISTRIBUTION SCHEMES IN WIRELESS SENSOR NETWORKS

by

Osman Yağan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor Armand M. Makowski, Chair/Advisor
Professor Alexander Barg
Professor Prakash Narayan
Professor Richard J. La
Professor Aravind Srinivasan, Dean's Representative

DEDICATION

To my family

ACKNOWLEDGMENTS

First and foremost I would like to thank my advisor, Professor Armand M. Makowski for his greatest guidance, support and friendship throughout my PhD studies. Without his broad knowledge, perfectionism, and dedication for excellence this thesis could not have been written. Throughout the four years I have spent at the University of Maryand, his mentorship was most helpful to me not only in my academic career but in every aspect of my life. I am deeply grateful for everything he has done for me and I will always feel fortunate that I have been his student.

I would like to thank Professors Prakash Narayan, Alexander Barg, Richard J. La, and Aravind Srinivasan for reading my thesis, taking part in my PhD dissertation defense committee, and providing me with valuable comments and suggestions.

I have been very fortunate to meet many friends at the University of Maryland. They all have brought lots of joy to my life and made it much easier for me to achieve this task. In particular, I am most grateful to Ersen Ekrem with whom I have started this journey together. Throughout, he was not only my room mate but also my best friend whose support I could always count on. I am also deeply grateful to Bora Çetin who has always been there to help and cheer me up; without him, life at UMD would me miserable. Special thanks should go to Dr. Yalın Sağduyu who has been a great friend and mentor to me; I will always miss our trio Ersen-Yalın-Osman which have had lots of fun in DC nights. I thank Himanshu Tyagi and Shalabh Jain for their

great friendship; I feel deeply fortunate to meet them. I also owe a word of thanks to my friends Dr. Alkan Soysal, Kaustubh Jain, Kapil Anand, Nitesh Shroff, Dr. Ravi Tandon, Dr. Arya Mazumdar, Dr. Anna Panteliodou, Ömur Özel, Filiz Yeşilköy, Marcos Vasconcelos, Dr. Ion Matei, Dr. Kiran Somasundaram, Raef Bassily and Dr. Prasanth Anthapadmanabhan for several helpful and cheerful discussions. I apologize to all of the friends that I have inadvertently left out who made my graduate study memorable.

Last but not least, I would like to thank my parents, my father Akif and my mom Sevil, and my brother Bahadır, for their constant unconditional love, support and encouragement. Finally, I would like to thank my wife, Nejla for being the love of my life, for supporting me to an extent that makes me feel like I can achieve everything I desire, and for making me the happiest man in the world. Without her, nothing would be meaningful if possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Wireless sensor networks (WSNs) are distributed collections of sensors with limited capabilities for computations and wireless communications. It is envisioned that [1] WSNs will be used in a wide range of applications areas such as health (e.g. patient monitoring), military (e.g., battlefield surveillance) and home (e.g., home automation). In many applications, both civilian and military, WSNs are likely to be deployed in hostile environments where communications can be monitored, and nodes are subject to capture and surreptitious use by an adversary. Under such circumstances, cryptographic protection will be needed to ensure secure communications and to enable sensor-capture detection, key revocation and sensor disabling. However, many security schemes developed for general network environments are inapplicable to WSNs owing to their unique features. For instance, public key cryptography is not feasible due to limitations on the physical memory and power consumption of the sensors. Also, traditional key exchange and distribution protocols are based on thrusting third parties, and turn out to be inadequate for large-scale WSNs due to

unknown network topology prior to deployment. See discussions in [14, 28, 33] on some of the challenges.

1.1 Random key predistribution schemes

Recently *random* key predistribution schemes have been proposed to address some of the difficulties mentioned above. In essence, a random key predistribution scheme can be described as follows – We start with a large set of (cryptographic) keys, the so-called *key pool*. For each sensor node i, a subset Σ_i of the key pool, called the *key ring* of node i, is generated by some random mechanism and inserted in the memory module of node i prior to deployment. Distinct nodes i and j can then establish a secure communication link if they are within wireless communication range of each other and the *security* condition

$$\Sigma_i \cap \Sigma_j \neq \emptyset \tag{1.1}$$

is satisfied. Most schemes in the literature and all schemes considered here have this form; they differ only in the way the random sets $\Sigma_1, \ldots, \Sigma_n$ are generated.

As we seek to understand various properties of these schemes, we study two cases. We begin with the "full visibility" case, namely that nodes are all within communication range of each other and secure communication between two nodes requires only that their key rings share at least one key. Understanding this case leads naturally to

studying the random graph structure whose adjacency is defined as follows: Distinct nodes i and j are adjacent, written $i \sim j$, if the condition (1.1) holds. Later on, we shall remove the full visibility assumption by integrating a wireless communication model with the random graph models induced under full visibility. More specifically, a new model is constructed where distinct nodes i and j are adjacent if and only if (1.1) holds *and* the wireless communication link between i and j is available; this amounts to an underlying random graph model formed by *intersecting* the full-visibility graph and the communication graph. Many of the questions raised in the dissertation deal with properties of these underlying random graphs, namely, absence of isolated nodes, connectivity, security, etc.

The idea of randomly assigning secure keys to sensor nodes prior to network deployment was first proposed by Eschenauer and Gligor [14]. Their scheme, hereafter called the EG scheme, operates according to the following key selection mechanism: Before network deployment, each sensor is independently assigned K distinct keys which are selected at random from a very large pool of P keys; see [14] for implementation details. Following the original work of Eschenauer and Gligor, a number of other key distribution schemes have been suggested. The q-composite scheme [7] is a slight 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. In other words, in the q-composite scheme, nodes i and j are connected only if they satisfy (1.1) with $|\Sigma_i \cap \Sigma_j| \geq q$. Another modification to the EG scheme was suggested by Du et al. [11], who combine the EG scheme and Blom's key management scheme. Chan et al. [7]

also proposed a random *pairwise* key predistribution whose implementation is quite different than the EG scheme: Each of the n sensor nodes is paired (offline) with K distinct nodes which are randomly selected from amongst all other nodes. Then, 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. The reader is referred to [5] for a detailed survey of various key distribution schemes for WSNs.

With a number of key predistribution schemes available, a basic question arises as to how these proposed ways of securing a WSN compare to each other. A natural approach would be to first investigate each particular scheme through its random graph *model* and then compare these schemes in terms of various metrics. This approach comes with the additional benefit of providing helpful guidelines to *dimension* the key distribution schemes. In other words, investigating the random graph model induced by a key distribution scheme might inform the selection of its parameters in order to ensure certain desirable properties in the network.

In view of this, the main goal of this dissertation is to investigate the EG scheme and the pairwise scheme of Chan et al and to make a thorough comparison. The evaluation metrics to be used in this comparison are given in Section 1.3 together with a motivation as to why each metric is being considered. For each metric, we summarize the relevant results we have obtained in Chapter 2 (regarding the EG scheme) and Chapter 3 (regarding the pairwise scheme). Then, in Chapter 4 we compare the two schemes in some details, explaining their relative advantages and disadvantages, as

well as their feasibility in various applications of WSNs. The mathematical tools to be used are collected in Chapter 5 while the main results of the dissertation are established in Chapter 6 through Chapter 10.

We now give a very brief introduction on random graphs in order to introduce the notation and some of the concepts to be used.

1.2 Random graphs

A graph G is an ordered pair $G = (V(G), E(G))$ where $V(G)$ denotes a (non-empty) set of vertices and $E(G)$ denotes a set of edges, i.e., set of ordered pairs of vertices. For the purpose of our discussion, all graphs under consideration are assumed to be undirected, simple and finite; see e.g., [34] for an extended introduction to graph theory. Given a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, a random graph $\mathbb{G}$ on the vertex set $V_n = \{1, \ldots, n\}$ is a graph-valued random variable (rv) defined by

$$\mathbb{G}(n) : \Omega \to \mathcal{G}(V_n)$$

where $\mathcal{G}(V_n)$ is the set of all undirected simple graphs on V_n.

Many results in this dissertation deal with establishing *zero-one laws* for certain properties of a random graph $\mathbb{G}$. Fix $n = 2, 3, \ldots$, and assume that $\mathbb{G}(n)$ has the vertex set $V_n = \{1, \ldots, n\}$. Often, the pmf of $\mathbb{G}(n)$ depends on a parameter, say υ, in some subset $\Upsilon \subseteq \mathbb{R}^d$, a fact indicated through the notation $\mathbb{G}(n; \upsilon)$. The parameter

υ is sometimes scaled with n so that $\{\mathbb{G}(n;\upsilon_n), n = 2, 3, \ldots\}$ now defines a family of random graphs. The main goal of the dissertation is to obtain conditions on the scaling $\upsilon : \mathbb{N}_0 \to \Upsilon$ such that either

$$\lim_{n\to\infty} \mathbb{P}[\ G(n;\upsilon_n) \text{ has property } \mathcal{A}\ ] = 0, \quad (\textbf{Zero-law})$$

or

$$\lim_{n\to\infty} \mathbb{P}[\ G(n;\upsilon_n) \text{ has property } \mathcal{A}\ ] = 1 \quad (\textbf{One-law})$$

for given a graph property $\mathcal{A}$.

To better discriminate between specific classes of random graphs we sometimes use letters other than $\mathbb{G}$ in the notation. This includes the random key graph $\mathbb{K}(n;K,P)$ and the random pairwise graph $\mathbb{H}(n;K)$, introduced in Chapter 2 and Chapter 3, respectively.

1.3 Evaluation metrics

1.3.1 Connectivity

A basic question concerning a *random* key predistribution scheme is its ability to achieve *secure connectivity* among participating nodes. Indeed, due to the randomness involved in the key distribution mechanism, there is a positive probability that secure connectivity will *not* be achieved – This will be so even in the best of cases

when the communication graph is itself connected, namely the full visibility case.[1] Therefore, it is of interest to obtain conditions on the scheme parameters so that the induced random graph model has certain connectivity properties. In particular, we study the absence of isolated nodes and graph connectivity. Interest in these properties stems from the fact that the desired level of connectivity may differ amongst various applications of WSNs. Since these properties are closely related to each other, studying one of them often proves useful in studying the other.

As mentioned before, we study the connectivity properties of a key distribution scheme in two phases. First, we consider the full visibility case according to which sensor 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 medium supporting WSNs. In return, this simplification makes it possible to focus on how randomization of the key distribution mechanism alone affects the establishment of a secure network in the best of circumstances, i.e., when there are no link failures.

In the second part we study the connectivity properties under more realistic assumptions that account for the possibility that communication links between nodes may not be available – This could occur due to the presence of physical barriers between nodes or because of harsh environmental conditions severely impairing transmission. To study such situations, we introduce a simple communication model where channels are mutually independent, and are either on or off. This amounts to modeling the communication graph by an Erdős-Rényi (ER) model [4, 13], and the overall

[1]The communication graph refers to the graph induced by the communication process whereby two nodes are adjacent if they are wireless neighbors, e.g., the disk model or the SINR graph [20].

system model is constructed by *intersecting* the random graph model of a key distribution scheme (under full visibility) and an ER graph.

We now give precise definitions of the connectivity properties mentioned above: Consider an undirected graph $G = (V, E)$ with vertex set $V = \{1, \ldots, n\}$ and edge set $E \subseteq V \times V$.

Definition 1.3.1 *A node i in V is said to be* **isolated** *if there exists no edges (in E) between i and any j in V distinct from i. The graph G has the property of "absence of isolated nodes" if no node in V is isolated.*

Definition 1.3.2 *Two distinct nodes i, j in V are said to be connected if E contains a path from i to j. The graph G is said to be* **connected** *if every pair of distinct nodes i, j in V are connected.*

1.3.2 Security

In modeling and comparing key distribution schemes for WSNs it is also fundamental to study the security properties of the resulting networks. In particular, for each of the schemes considered here, we wish to understand the resiliency of the network against external attacks. As in [26], we consider extremely severe forms of massive attacks. The adversary captures a certain number of nodes and owns the key rings of these selected nodes. We assume that a link between two nodes, say i and j, is compromised if the adversary owns a key which is stored in *both* the key rings

of i and j. We also assume that the adversary has unlimited computing power, and can minimize the number of nodes that it needs to capture in order to compromise a given number of links in the network.

In many applications of WSNs, the network as a whole can still operate in a useful manner even though a *small* number of sensors fail, i.e., taken under the control of an adversary [26]. In these cases, it can be much more important to protect the global functionality of the network than a few individual communication links. However, if the adversary is capable of capturing a large fraction of the nodes, then there is not much that can be done to salvage the network functionality. Therefore, in evaluating the security provided by a key predistribution scheme, it is essential to ask whether a *severe* attack can be achieved cheaply, namely by capturing just a small number of the nodes.

With these considerations in mind we now introduce the notions of a network being *unassailable* and *unsplittable*; see [26]. Let C_r denote the *maximum* number of links that an adversary can compromise by capturing r nodes. Similarly, let I_r denote the size of the *largest* subset of sensors whose communications with the rest of the network can be compromised by capturing r nodes. Both unassailability and unsplittability are defined in the asymptotic regime where the total number n of nodes grows unboundedly large, and the total number r_n of captured nodes grows sub-linearly with n, i.e., $r_n = o(n)$.

Definition 1.3.3 *A network is said to be* **unassailable** *if*

$$C_{r_n} = o(E_n) \quad \textit{whenever} \quad r_n = o(n), \tag{1.2}$$

where E_n denotes the total number of links in the network.

Definition 1.3.4 *A network is said to be* **unsplittable** *if*

$$I_{r_n} = o(n) \quad \textit{whenever} \quad r_n = o(n). \tag{1.3}$$

These conditions are highly desirable as they imply that an adversary cannot impair a considerable part of the network without capturing a considerable number of nodes.

Unassailability and unsplittability of a given scheme are both related to the structure of the random graph model induced by the particular scheme. However, in our evaluation of security we also consider a number of *protocol based* criteria that represent desirable characteristics in a key distribution scheme for sensor networks [7]. **Resistance against node replication** is related to whether the adversary can insert additional hostile nodes into the network. **Revocation** deals with whether a misbehaving node, once detected, can be dynamically removed from the system. Finally, a protocol provides **node-to-node authentication** if any node can ascertain the identity of the nodes with which it communicates.

1.3.3 Memory usage: Key-ring sizes

As stated earlier, sensors in a WSN have limited memory so that the number of secure keys that can be stored in each sensor's memory module is constrained. Therefore, in evaluating a key predistribution scheme, it is also essential to consider the key ring sizes that are required to achieve certain connectivity and security properties in the network.

1.3.4 Scalability and implementation issues

As explained by Chan et al. [7], the security characteristics of a network may be weakened as the number of nodes in the network grows. Also, certain network properties may only be achieved by key ring sizes that grow with the increasing network size. Recalling that WSNs are likely to have considerable size, it is therefore desirable that a key distribution scheme is able to support a large number of sensors, and this leads to studying the *scalability* of key distribution schemes.

We also take into account possible implementation difficulties in practical scenarios. For instance, in some applications it might be required to increase the size of the network after initial deployment. When the network is deployed gradually, it is essential to ask whether a key distribution scheme is capable of accommodating sensors added at a later time such that secure connections can be established with the already deployed nodes.

1.4 Notations and conventions

All limiting statements, including asymptotic equivalences, are understood with the number of sensor nodes n going to infinity. The rvs under consideration are all defined on the same probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. Probabilistic statements are made with respect to this probability measure $\mathbb{P}$, and we denote the corresponding expectation operator by $\mathbb{E}$. Also, we use the notation $=_{st}$ to indicate distributional equality. The indicator function of an event E is denoted by $\mathbf{1}[E]$. For any discrete set S we write $|S|$ for its cardinality. For sequences $a, b : \mathbb{N}_0 \to \mathbb{R}_+$, we write $a_n \sim b_n$ as a shorthand for the asymptotic equivalence $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. Similarly, we write $a_n = o(b_n)$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$, whereas $a_n = O(b_n)$ means that there exists $c > 0$ such that $a_n \leq c \cdot b_n$ for all n sufficiently large. Finally, we have $a_n = \Omega(b_n)$ if $b_n = O(a_n)$, or equivalently, if there exists $c > 0$ such that $a_n \geq c \cdot b_n$ for all n sufficiently large.

Chapter 2

Modeling the EG scheme: Random key graphs

The first scheme considered in this dissertation is the random key distribution scheme of Eschenauer and Gligor [14], hereafter referred to as the EG scheme. We start by explaining the details of its implementation and then introduce the induced random graph model under the assumption of full visibility. Finally, we give a summary of the results that we have obtained for several properties of the EG scheme and of its induced graph. These results will later be established in Chapters 6 and 7.

2.1 The Eschenauer-Gligor Scheme

The EG scheme operates in three phases: Consider a collection of n sensor nodes equipped with wireless transmitters, and assume available a large set of P cryptographic keys, also known as the *key pool*.

(i) **Initialization phase:** Before network deployment, each node randomly selects a set of K *distinct* keys from the pool. These K keys form the *key ring* of the node, and are

inserted into its memory module. Key rings are selected independently across nodes.

(ii) **Key setup phase:** After deployment, each node discovers its *wireless neighbors*, i.e., nodes which are within its wireless communication range. When a node finds a wireless neighbor with whom it shares a key, they mutually authenticate the key to verify that the other party actually owns it. At the end of this phase, wireless neighbors which have keys in common can now communicate securely with each other in one hop.

(iii) **Path-key identification phase:** The key rings being randomly selected, there is a possibility that some pairs of wireless neighbors may not share a key. If a path made up of nodes sharing keys pairwise exists between such a pair of wireless neighbors, this (secure) path can be used to exchange a *path-key* to establish a direct (and secure) link between them.

A basic question concerning the EG scheme is its ability to achieve *secure connectivity*. This will be studied through induced random graph models by developing scaling laws which correspond to desirable network properties, e.g., absence of secure nodes which are isolated and secure connectivity. In the process we shall obtain conditions on the scheme parameters so that these properties occur with high probability as the number of nodes becomes large. We start by introducing the random graph induced by the EG scheme under the assumption of full visibility, i.e., *the random key graph*.

2.2 Random key graphs

Random key graphs are parametrized by three positive integers, namely the number n of nodes, the size P of the key pool and the size K of each key ring with $K \leq P$. To lighten the notation we group the integers P and K into the ordered pair $\theta \equiv (K, P)$. With $\mathcal{P}$ the set of cryptographic keys, let $\mathcal{P}_K$ to denote the collection of all subsets of $\mathcal{P}$ which contain exactly K elements – Obviously, with $|\mathcal{P}| = P$, we have $|\mathcal{P}_K| = \binom{P}{K}$. The key rings $\Sigma_1(\theta), \ldots, \Sigma_n(\theta)$ are assumed to be *i.i.d.* random variables (rvs) with

$$\mathbb{P}\left[\Sigma_i(\theta) = S\right] = \binom{P}{K}^{-1}, \quad S \in \mathcal{P}_K \tag{2.1}$$

for any $i = 1, \ldots, n$. This corresponds to selecting keys randomly and *without* replacement from the key pool.

Distinct nodes $i, j = 1, \ldots, n$ are said to be adjacent if they share at least one key in their key rings in which case an undirected link is assigned between nodes i and j. The resulting random graph defines the *random key graph* (RKG) on the vertex set $\{1, \ldots, n\}$, hereafter denoted by $\mathbb{K}(n; \theta)$. For distinct $i, j = 1, \ldots, n$, it is a simple matter to check that

$$\mathbb{P}\left[\Sigma_i(\theta) \cap \Sigma_j(\theta) = \emptyset\right] = q(\theta) \tag{2.2}$$

with

$$q(\theta) = \begin{cases} 0 & \text{if } \; P < 2K \\ \\ \dfrac{\binom{P-K}{K}}{\binom{P}{K}} & \text{if } \; 2K \leq P, \end{cases} \tag{2.3}$$

whence the probability of edge occurrence between any two nodes is equal to $1-q(\theta)$. The expression (2.3) is a simple consequence of the fact that

$$\mathbb{P}\left[S \cap \Sigma_i(\theta) = \emptyset\right] = \frac{\binom{P-|S|}{K}}{\binom{P}{K}}, \quad i = 1, \ldots, n \tag{2.4}$$

for every subset S of $\{1, \ldots, P\}$ with $|S| \leq P - K$. The case $P < 2K$ is clearly not interesting: It corresponds to an edge existing between every pair of nodes, so that $\mathbb{K}(n;\theta)$ coincides with the complete graph $\mathcal{K}_n$.

Random key graphs form a subclass in the family of *random intersection* graphs. However, the model adopted here differs from the random intersection graphs discussed by Singer-Cohen et al. in [15, 24, 31] where each node is assigned a key ring, one key at a time according to a Bernoulli-like mechanism (so that each key ring has a random size and has a positive probability of being empty).

Despite strong similarities, we stress that the random graph $\mathbb{K}(n;\theta)$ is *not* an Erdős-Rényi graph $\mathbb{G}(n;p)$ [22] even if we use

$$p = 1 - q(\theta). \tag{2.5}$$

This is so because edge assignments are correlated in $\mathbb{K}(n;\theta)$ but independent in $\mathbb{G}(n;p)$: Indeed, let $i \sim j$ denote the event that there exists an edge between the nodes i and j. With $K = 2$ and $P = 4 \times 10^4$, it follows from (2.3) that

$$\mathbb{P}\left[i \sim j\right] \simeq 10^{-4}$$

whereas

$$\mathbb{P}\left[i \sim j \mid i \sim k, j \sim k\right] \simeq \frac{1}{2}.$$

Throughout, with $n = 2, 3, \ldots$ and positive integers K and P such that $K \leq P$, let $P(n;\theta)$ denote the probability that the random key graph $\mathbb{K}(n;\theta)$ is connected, namely

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

Similarly, let $P^{\star}(n;\theta)$ denote the probability that the random key graph has no isolated nodes, i.e.,

$$P^{\star}(n;\theta) := \mathbb{P}\left[\mathbb{K}(n;\theta) \text{ contains no isolated nodes}\right].$$

In the full visibility case assumed here, $P(n;\theta)$ coincides with the probability of secure connectivity mentioned earlier.

2.3 Main results

We now summarize the main results obtained for the EG scheme. As a first step, for the full visibility case, we obtain results regarding the connectivity properties of the random key graph. Next, we remove the full visibility assumption and give analogous results in Section 2.3.4 for the case when a wireless communication model is also integrated to the random key graph.

2.3.1 A roadmap

We seek conditions on P and K so that $P(n;\theta)$ (and $P^{\star}(n;\theta)$) is as large (i.e., as close to one) as possible. This issue naturally leads to zero-one laws for graph connectivity in random key graphs when P and K are appropriately scaled with n. Such zero-one laws are available for Erdős-Rényi graphs [13, 22] and can provide helpful guidelines in establishing their analogs for random key graphs. Below, we outline a possible approach which is inspired by the discussion of this issue given by Eschenauer and Gligor in their original work [14]; see also the discussion in [9, 10].

(i) Let $\mathbb{G}(n;p)$ denote the Erdős-Rényi (ER) graph on n vertices with edge probability p $(0 < p \leq 1)$ [4, 22]. As mentioned before, the random key graph $\mathbb{K}(n;\theta)$ is *not* stochastically equivalent to an Erdős-Rényi graph $\mathbb{G}(n;p)$. Yet, setting aside this fact, we note that $\mathbb{K}(n;\theta)$ can be matched naturally to an Erdős-Rényi graph $\mathbb{G}(n;p)$ with p and θ related through

$$p = 1 - q(\theta). \tag{2.6}$$

This constraint ensures that link assignment probabilities in $\mathbb{K}(n;\theta)$ and $\mathbb{G}(n;p)$ coincide. Moreover, under (2.6) it is easy to check that the degree of a node in either random graph is a Binomial rv with the same parameters, namely $n-1$ and $p = 1 - q(\theta)$![1] Given that the degree distributions in a random graph are often taken as a good indicator of its connectivity properties, it has been conjectured that the zero-one law for graph connectivity in random key graphs can be inferred from the

[1]For Erdős-Rényi graphs this result is well known, while for random key graphs this characterization is a straightforward consequence of (2.4).

analog result for Erdős-Rényi graphs when matched through the condition (2.6).

(ii) To perform such a "transfer" we first recall that in Erdős-Rényi graphs the property of absence of isolated nodes is known to exhibit the following zero-one law [4]: If we scale the edge assignment probability p according to

$$p_n = \frac{\log n + \alpha_n}{n}, \quad n = 1, 2, \dots \tag{2.7}$$

for some deviation sequence $\alpha : \mathbb{N}_0 \to \mathbb{R}$, then

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{G}(n; p_n) \text{ contains no isolated nodes}\right] = \begin{cases} 0 & \text{if } \lim_{n\to\infty} \alpha_n = -\infty \\ \\ 1 & \text{if } \lim_{n\to\infty} \alpha_n = +\infty. \end{cases} \tag{2.8}$$

Furthermore, it is known [4] that the absence of isolated nodes and graph connectivity are asymptotically equivalent properties in ER graphs: Under (2.7) it is the case that

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{G}(n; p_n) \text{ is connected}\right] = \begin{cases} 0 & \text{if } \lim_{n\to\infty} \alpha_n = -\infty \\ \\ 1 & \text{if } \lim_{n\to\infty} \alpha_n = +\infty. \end{cases} \tag{2.9}$$

(iii) Under the matching condition (2.6), these classical results suggest scaling the parameters K and P with n according to

$$1 - \frac{\binom{P_n - K_n}{K_n}}{\binom{P_n}{K_n}} = \frac{\log n + \alpha_n}{n}, \quad n = 1, 2, \dots \tag{2.10}$$

for some sequence $\alpha : \mathbb{N}_0 \to \mathbb{R}$. In view of (2.9) it is then not too unreasonable to expect that the zero-one laws

$$\lim_{n\to\infty} P^\star(n;\theta_n) = \begin{cases} 0 & \text{if } \lim_{n\to\infty}\alpha_n = -\infty \\ \\ 1 & \text{if } \lim_{n\to\infty}\alpha_n = +\infty \end{cases} \tag{2.11}$$

and

$$\lim_{n\to\infty} P(n;\theta_n) = \begin{cases} 0 & \text{if } \lim_{n\to\infty}\alpha_n = -\infty \\ \\ 1 & \text{if } \lim_{n\to\infty}\alpha_n = +\infty \end{cases} \tag{2.12}$$

should hold (possibly under some additional assumptions).

Of course, for this approach to be operationally useful, a good approximation to the left handside of (2.10) is needed. It will become apparent (Lemma 7.4.4) that

$$1 - \frac{\binom{P-K}{K}}{\binom{P}{K}} \simeq \frac{K^2}{P} \tag{2.13}$$

under natural assumptions. Thus, if we scale the parameters according to

$$\frac{K_n^2}{P_n} = \frac{\log n + \alpha_n}{n}, \quad n = 1, 2, \ldots \tag{2.14}$$

it is natural to conjecture that the zero-one laws (2.11)-(2.12) should still hold; we will show that this is indeed the case.

2.3.2 A zero one law for the absence of isolated nodes

As with Erdős-Rényi graphs, we address the connectivity problem by first showing the existence of a zero-one law for the absence of isolated nodes in RKGs, thereby establishing the validity of the conjecture (2.11) under (2.14).

Any pair of functions $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ defines a *scaling* provided the natural conditions

$$K_n \leq P_n, \quad n = 1, 2, \ldots \tag{2.15}$$

are satisfied. For any such scaling we can associate a sequence $\alpha : \mathbb{N}_0 \to \mathbb{R}$ through the relation (2.14). In other words, we set

$$\alpha_n := n\frac{K_n^2}{P_n} - \log n, \quad n = 1, 2, \ldots$$

We refer to this sequence $\alpha : \mathbb{N}_0 \to \mathbb{R}$ as the *deviation function* associated with the scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$. As the terminology suggests, the deviation function measures by how much the scaling deviates from the critical scaling $\frac{\log n}{n}$.

The next result, established in Chapter 6, verifies the conjectured zero-one law (2.11) for the absence of isolated nodes.

Theorem 2.3.1 *For any scaling $K, P : \mathbb{N}_0 \to \mathbb{N}_0$, it holds that*

$$\lim_{n\to\infty} P^\star(n;\theta_n) = \begin{cases} 0 & \text{if } \lim_{n\to\infty} \alpha_n = -\infty \\ \\ 1 & \text{if } \lim_{n\to\infty} \alpha_n = +\infty \end{cases} \tag{2.16}$$

where the function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ is determined through (2.14).

A proof of Theorem 2.3.1 is given in Chapter 6. Theorem 2.3.1 was also obtained independently by Blackburn and Gerke [3].

If the random key graph $\mathbb{K}(n;\theta)$ is connected, then it does not contain any isolated nodes, whence

$$P(n;\theta) \leq P^\star(n;\theta). \tag{2.17}$$

As a result, Theorem 2.3.1 already implies

$$\lim_{n\to\infty} P(n;\theta_n) = 0 \quad \text{if } \lim_{n\to\infty} \alpha_n = -\infty,$$

thereby establishing the zero-law of the conjecture (2.12) under (2.14).

2.3.3 A zero-one law for connectivity

As mentioned earlier, Theorem 2.3.1 already implies a zero law for the property of graph connectivity and establishes a part of the conjecture (2.12). The remaining

part (namely the one-law) of the conjecture (2.12) (under (2.14)) is discussed next:

A scaling $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ is said to be *admissible* if

$$2 \leq K_n \tag{2.18}$$

for *all* $n = 1, 2, \ldots$ *sufficiently* large. We can now present the zero-one law for connectivity in RKG's:

Theorem 2.3.2 *Consider an admissible scaling* $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ *with deviation function* $\alpha : \mathbb{N}_0 \to \mathbb{R}$ *determined through (2.14). We have*

$$\lim_{n\to\infty} P(n; \theta_n) = 0 \quad \textit{if } \lim_{n\to\infty} \alpha_n = -\infty. \tag{2.19}$$

On the other hand, if there exists some $\sigma > 0$ *such that*

$$\sigma n \leq P_n \tag{2.20}$$

for all $n = 1, 2, \ldots$ *sufficiently large, then we have*

$$\lim_{n\to\infty} P(n; \theta_n) = 1 \quad \textit{if } \lim_{n\to\infty} \alpha_n = \infty. \tag{2.21}$$

A detailed discussion and a proof of Theorem 2.3.2 are provided in Chapter 7. The-

orem 2.3.2 readily implies the following zero-one law.

Corollary 2.3.3 *Consider an admissible pair of functions $P, K : \mathbb{N}_0 \to \mathbb{N}$ such that*

$$\frac{K_n^2}{P_n} \sim c\frac{\log n}{n} \tag{2.22}$$

holds for some $c > 0$. If there exists some $\sigma > 0$ such that (2.20) holds for all n sufficiently large, then we have

$$\lim_{n\to\infty} P(n;\theta_n) = \begin{cases} 0 & \text{if } 0 < c < 1 \\ \\ 1 & \text{if } 1 < c. \end{cases} \tag{2.23}$$

Indeed, it suffices to use Theorem 2.3.2 with any admissible pair of functions $P, K : \mathbb{N}_0 \to \mathbb{N}$ whose function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies

$$\alpha_n = (c-1)\,(1+o(1)) \cdot \log n, \quad n = 1, 2, \ldots$$

2.3.4 Connectivity results under an ON-OFF channel

The results presented in the previous sections are obtained under the assumption of full visibility. Now, we remove the full visibility assumption and seek to account for the possibility that communication links between nodes may not be available. Namely,

we assume a wireless communication model that consists of independent channels each of which can be either on or off. Thus, with p in $(0,1)$, let $\{B_{ij}(p), 1 \leq i < j \leq n\}$ denote i.i.d. $\{0,1\}$-valued rvs with success probability p. The channel between nodes i and j is available (resp. up) with probability p and unavailable (resp. down) with the complementary probability p.

Distinct nodes i and j are said to be B-adjacent, written $i \sim_B j$, if $B_{ij}(p) = 1$. The notion of B-adjacency defines an Erdős-Rényi graph $\mathbb{G}(n;p)$ on the vertex set $\{1,\ldots,n\}$. Obviously,

$$\mathbb{P}\left[i \sim_B j\right] = p.$$

To study the connectivity of the EG scheme under this ON-OFF channel, we consider a random graph model that is obtained by *intersecting* the random key graph $\mathbb{K}(n;\theta)$ with the ER graph $\mathbb{G}(n;p)$. More precisely, the distinct nodes i and j are said to be adjacent, written $i \sim j$, if and only if they are both adjacent in the random key graph (i.e., they share a key) and B-adjacent, namely

$$i \sim j \quad \text{iff} \qquad \Sigma_i(\theta) \cap \Sigma_j(\theta) \neq \emptyset \quad \text{and} \quad B_{ij}(p) = 1. \tag{2.24}$$

The resulting *undirected* random graph defined on the vertex set $\{1,\ldots,n\}$ through this notion of adjacency is denoted $\mathbb{K} \cap \mathbb{G}(n;\theta,p)$.

Throughout the collections of rvs $\{K_1(\theta),\ldots,K_n(\theta)\}$ and $\{B_{ij}(p), 1 \leq i < j \leq n\}$ are assumed to be independent, in which case the edge occurrence probability in

$\mathbb{K} \cap \mathbb{G}(n;\theta,p)$ is given by

$$\mathbb{P}\left[i \sim j\right] = (1 - q(\theta)) \cdot \mathbb{P}\left[i \sim_B j\right] = p(1 - q(\theta)). \tag{2.25}$$

We now present connectivity results for the random graph $\mathbb{K} \cap \mathbb{G}(n;\theta,p)$. To fix the terminology, we refer to any mapping $p : \mathbb{N}_0 \to (0,1)$ as a scaling for ER graphs. Very recently we have established a zero-one law for the absence of isolated nodes [47].

Theorem 2.3.4 *Consider an admissible scaling $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ and a scaling $p : \mathbb{N}_0 \to (0,1)$ such that*

$$p_n(1 - q(\theta_n)) \sim c\frac{\log n}{n}, \quad n = 1, 2, \ldots \tag{2.26}$$

for some $c > 0$. If $\lim_{n\to\infty} p_n \log n = p^\star$ exists, then we have

$$\lim_{n\to\infty} \mathbb{P}\left[\begin{array}{c} \mathbb{K} \cap \mathbb{G}(n;\theta_n,p_n) \text{ contains} \\ \text{no isolated nodes} \end{array}\right] = \begin{cases} 0 & \text{if } c < 1 \\ \\ 1 & \text{if } c > 1. \end{cases} \tag{2.27}$$

An analog of Theorem 2.3.4 also holds for the property of graph connectivity.

Theorem 2.3.5 *Consider an admissible scaling $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ and a scaling $p : \mathbb{N}_0 \to (0,1)$ such that (2.26) holds for some $c > 0$. If $\lim_{n\to\infty} p_n \log n = p^\star$ exists,*

then we have

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{K} \cap \mathbb{G}(n;\theta_n,p_n) \;\; \textit{is connected}\right] = 0 \quad \textit{if} \;\; c < 1. \tag{2.28}$$

On the other hand, if there exists some $\sigma > 0$ *such that*

$$\sigma n \leq P_n \tag{2.29}$$

for all $n = 1, 2, \ldots$ *sufficiently large, then we have*

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{K} \cap \mathbb{G}(n;\theta_n,p_n) \;\; \textit{is connected}\right] = 1 \quad \textit{if} \;\; c > 1. \tag{2.30}$$

We do not give the proofs of Theorem 2.3.4 and Theorem 2.3.5 in this dissertation, but they can be found in [47].

It is now natural to wonder as to whether the zero-one laws of Theorem 2.3.1-Theorem 2.3.2 for the full visibility case can be deduced from the more general results given in Theorem 2.3.4-Theorem 2.3.5, say by setting $p_n = 1$ for each $n = 2, 3, \ldots$ sufficiently large. Indeed, in view of (2.13), it is a simple matter to check that Theorem 2.3.5 implies a weaker version of Theorem 2.3.2, namely Corollary 2.3.3. Nevertheless, one can easily see that Theorem 2.3.2 does not follow from Theorem 2.3.5: Consider an admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ such that (2.20) holds and the deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ determined through (2.14) satisfies $\alpha_n = o(\log n)$ and $\lim_{n\to\infty} \alpha_n = \infty$.

Then, if $p_n = 1$ for all n sufficiently large, we have (see Lemma 7.4.4)

$$p_n(1 - q(\theta_n)) \sim 1 - q(\theta_n) \sim \frac{K_n^2}{P_n} = \frac{\log n + \alpha_n}{n} \sim \frac{\log n}{n},$$

whence (2.26) holds with $c = 1$. Thus, under the enforced assumptions, Theorem 2.3.5 does not tell anything about the asymptotic behavior of the probability that $\mathbb{K} \cap \mathbb{G}(n; \theta_n, 1)$ (or, equivalently $\mathbb{K}(n; \theta_n)$) is connected. Yet, we see from Theorem 2.3.2 that $\lim_{n\to\infty} \mathbb{P}\left[\mathbb{K}(n; \theta_n) \text{ is connected}\right] = 1$ as soon as $\lim_{n\to\infty} \alpha_n = \infty$.

Chapter 3

Modeling the pairwise key distribution scheme: Random pairwise graphs

In this chapter, we consider the random pairwise key distribution scheme of Chan, Perrig and Song [7] which was proposed as an alternative to the EG scheme. As was done in the previous chapter for the EG scheme, we start with a brief discussion of the implementation details. We then introduce the random graph model induced under the assumption of full visibility; in this case the induced model will be referred to as the random pairwise graph. Finally, we summarize the results which we have obtained for several properties of the scheme and its induced graph. These results will be established in Chapters 8, 9, and 10.

3.1 Pairwise key distribution scheme of Chan, Perrig and Song

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 K 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 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 $\ell_{n,i} : \Gamma_{n,i} \to \{1, \ldots, K\}$ 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

$$\Sigma_{n,i} := \{\omega^{\star}_{n,ij},\ j \in \Gamma_{n,i}\} \cup \{\omega^{\star}_{n,ji},\ i \in \Gamma_{n,j}\}. \tag{3.1}$$

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. It is not excluded 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 $\omega^{\star}_{n,ji}$. By construction this scheme supports node-to-node authentication.

This construction of the key rings constitutes the *initialization phase* of the pairwise scheme where sensors are loaded with secure keys before deployment. Once the network is deployed, the *key setup phase* and the *path-key identification phase* will take place exactly as in the case of EG scheme; see Chapter 2.1 for details.

One of the main questions regarding the feasibility of the pairwise scheme is its ability to achieve secure connectivity amongst *deployed* sensors. This is so because, given the randomness involved in the pairing mechanism, there is a positive probability that connectivity will not be achieved even in the best of circumstances when the communication graph is fully connected. Therefore, it is desirable to understand how should the parameter K be selected (for a given number n of sensors) so that the resulting network is connected with high probability. We proceed as in the case of the EG scheme by first investigating the random graph model induced (under full visibility) by the pairwise scheme, i.e., the random pairwise graph.

3.2 Random pairwise graphs

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

$$i \sim j \quad \text{iff} \quad \Sigma_{n,i} \cap \Sigma_{n,j} \neq \emptyset. \tag{3.2}$$

Let $\mathbb{H}(n; K)$ denote the undirected random graph on the vertex set $\{1, \ldots, n\}$ induced by the adjacency notion (3.2). From now on, we shall refer to this random graph as the random *pairwise* graph.

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

$$\xi_{n,ij} := \mathbf{1}\left[i \in \Gamma_{n,j}\ \vee\ j \in \Gamma_{n,i}\right], \quad i \neq j,\ \ i,j = 1,\ldots,n$$

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

$$1 - \xi_{n,ij} = \mathbf{1}\left[i \notin \Gamma_{n,j},\ j \notin \Gamma_{n,i}\right]. \tag{3.3}$$

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

$$\mathbb{P}\left[i \notin \Gamma_{n,j},\ j \notin \Gamma_{n,i}\right] = \mathbb{P}\left[i \notin \Gamma_{n,j}\right]\mathbb{P}\left[j \notin \Gamma_{n,i}\right] = \left(1 - \frac{K}{n-1}\right)^2 \tag{3.4}$$

by independence. As a result,

$$\mathbb{E}\left[\xi_{n,ij}\right] = 1 - \left(1 - \frac{K}{n-1}\right)^2. \tag{3.5}$$

Put differently, the link assignment probability $\lambda_n(K)$ in random pairwise graphs is given by

$$\lambda_n(K) \;=\; 1-\left(1-\frac{K}{n-1}\right)^2 = \frac{2K}{n-1}-\left(\frac{K}{n-1}\right)^2. \tag{3.6}$$

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 (3.3) 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 (3.3) 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

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

by elementary calculations. It is now plain that the random graph $\mathbb{H}(n; K)$ is not an Erdős-Rényi graph [4] – 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 [23]; see Chapter 10.6 for details.

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.

3.3 Main results

3.3.1 Zero-one laws for connectivity

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

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

We wish to determine conditions on K and n so that $P(n; K)$ is as large (i.e., as close to one) as possible. The first technical result of this chapter, given next, provides a lower bound on $P(n; K)$.

Theorem 3.3.1 *Consider any positive integer $K \geq 2$. With $n(K) = \lceil e(K+1) \rceil$, we have*

$$P(n; K) \geq 1 - \frac{(K+1)^{K^2-1}}{2} \cdot n^{-(K^2-2)}, \qquad n \geq n(K) \tag{3.7}$$

Theorem 3.3.1 is established in Chapter 8.2. The bound (3.7) gives some indication as to how fast the convergence $\lim_{n\to\infty} P(n; K) = 1$ occurs when $K \geq 2$. As would be expected, the convergence becomes faster with larger K; see also (3.9) below. For $K = 2$, $n(K) = 9$, the bound (3.7) takes the simpler form

$$P(n; 2) \geq 1 - \frac{27}{2n^2}, \quad n \geq 9. \tag{3.8}$$

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

$$P(n; 2) \leq P(n, K), \quad 2 \leq K < n. \tag{3.9}$$

Making use of (3.8) we then conclude that

$$P(n; K) \geq 1 - \frac{27}{2n^2}, \qquad \begin{array}{l} 2 \leq K < n \\ \\ n \geq n(K) \end{array} . \tag{3.10}$$

A zero-one law for connectivity is presented next.

Theorem 3.3.2 *With any positive integer K, it holds that*

$$\lim_{n\to\infty} P(n; K) = \begin{cases} 0 & \text{if } K = 1 \\ \\ 1 & \text{if } K \geq 2. \end{cases} \tag{3.11}$$

The one-law in Theorem 3.3.2 is an easy consequence of the bound (3.7) (or (3.10)), while the zero-law of Theorem 3.3.2 is proved separately in Chapter 8.3.

Theorem 3.3.2 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

$$K_n < n, \quad n = 1, 2, \ldots. \tag{3.12}$$

Corollary 3.3.3 *For any scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$, we have*

$$\lim_{n\to\infty} P(n; K_n) = 1 \tag{3.13}$$

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

Because $\mathbb{H}(n; K)$ cannot be equated with an Erdős-Renyi graph, neither Theorem 3.3.1 nor Corollary 3.3.3 are consequences of classical results for Erdős-Renyi graphs [4]. Indeed, consider the following well-known zero-one law for Erdős-Rényi graphs (easily induced from (2.7)-(2.8)): For any scaling $p : \mathbb{N}_0 \to [0, 1]$ satisfying

$$p_n \sim c \cdot \frac{\log n}{n} \tag{3.14}$$

for some $c > 0$, it holds that

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{G}(n; p_n) \text{ is connected}\right] = \begin{cases} 0 & \text{if } \ 0 < c < 1 \\ \\ 1 & \text{if } \ 1 < c. \end{cases} \tag{3.15}$$

In view of (3.6), $\frac{2K_n}{n-1} - \frac{K_n^2}{(n-1)^2}$ 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 $\mathbb{H}(n;K_n)$ (akin to the one discussed in Chapter 2.3.1 for random key graphs) would suggest scaling K such that

$$\frac{2K_n}{n-1}-\frac{K_n^2}{(n-1)^2}\sim c\frac{\log n}{n}, \tag{3.16}$$

for some $c>0$. Any scaling $K:\mathbb{N}_0\to\mathbb{N}_0$ which behaves like (3.16) must necessarily satisfy $K_n=o(n)$, and it is easy to see that requiring (3.16) is equivalent to

$$2K_n\sim c\log n. \tag{3.17}$$

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 (3.17). Clearly, this yields the misleading conclusion that K_n has to behave like $c\log n$ for some $c>\frac{1}{2}$ in order for $\mathbb{H}(n;K_n)$ to be asymptotically almost surely (a.a.s.) connected– In fact, by Theorem 3.3.2 it is only needed to have $K_n\geq 2$ for all n sufficiently large.

3.3.2 Key ring sizes associated with the pairwise scheme

The mere fact that $\mathbb{H}(n;K)$ becomes connected even with very small K values does not imply that the *number* of keys (i.e., the size $|\Sigma_{n,i}|$) to achieve connectivity is necessarily small. Indeed, in contrast with the EG scheme and its variants, the pairwise scheme produces key rings of variable size between K and $K+n-1$. In this dissertation, we explore this issue further and obtain conditions on a scaling $K:\mathbb{N}_0\to\mathbb{N}_0$ which ensure that all key rings have size of order $\log n$.

First observe via (3.1) that

$$|\Sigma_{n,i}| =_{st} K + \text{Bin}\left(n-1, \frac{K}{n-1}\right), \quad i=1,\ldots,n$$

where $\text{Bin}(n,p)$ stands for a *binomial* rv with n trials and success probability p. Thus, we have

$$\mathbb{E}\left[|\Sigma_{n,i}|\right] = 2K, \quad i=1,\ldots,n. \tag{3.18}$$

It is also a simple matter to check that

$$|\Sigma|_{avg} := \frac{|\Sigma_{n,1}|+\cdots+|\Sigma_{n,n}|}{n} = 2K \tag{3.19}$$

with $|\Sigma|_{avg}$ denoting the *average* key ring size in the network.

We begin by noting that

$$\frac{|\Sigma_{n,1}(K_n)|}{2K_n} \xrightarrow[n]{P} 1 \tag{3.20}$$

for any scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ as soon as $\lim_{n\to\infty} K_n = \infty$; this is an easy consequence of Lemma 3.3.4 below. Thus, when K_n becomes large with n, although the key ring size $|\Sigma_{n,1}(K_n)|$ fluctuates from K_n to $K_n+(n-1)$, it does so with a propensity to hover about its mean $2K_n$. This can be given a precise meaning as the next concentration result shows. First we set

$$a(\tau) := (1+\tau)\cdot\log(1+\tau) - \tau, \quad \tau > -1, \tag{3.21}$$

and

$$b(\tau) := \begin{cases} 2 & \text{if } 0 < \tau < 1 \\ 1 & \text{if } 1 \le \tau. \end{cases} \tag{3.22}$$

Lemma 3.3.4 *Consider positive integers K and n such that $K < n$. For any $c > 0$, we have*

$$\mathbb{P}\left[||\Sigma_{n,1}(K)| - 2K| > cK\right] \le b(c)\cdot e^{-a(c)K} \tag{3.23}$$

for all $n = 2, 3, \ldots$ with $a(c) > 0$ given by (3.21) and $b(c)$ given by (3.22).

Lemma 3.3.4, which is established in Chapter 8.4, has several consequences: If the parameter K is scaled according to some scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that $\lim_{n\to\infty} K_n = \infty$, then with high probability, the number of keys $|\Sigma_{n,1}(K_n)|$ stored at a node will be between $(2-c)K_n$ and $(2+c)K_n$ for any $c > 0$. For sake of concreteness consider the case $c = 1$: Since it is always the case that $|\Sigma_{n,1}(K_n)| \ge K_n$ (e.g., see (8.19)), the

inequality (3.23) reduces to

$$\mathbb{P}\left[|\Sigma_{n,1}(K_n)| > 3K_n\right] \leq e^{-a(1)K_n} \tag{3.24}$$

for all $n = 2, 3, \ldots$. It is easily seen from (3.24) that $\mathbb{P}\left[|\Sigma_{n,1}(K_n)| > 3K_n\right]$ is already negligible (i.e., "vanishes") for $K_n \geq 15$.

A related result also holds for the maximal key ring size when restricting attention to the subclass of logarithmic scalings. First define the maximal key ring size as

$$M_n(K) := \max_{i=1,\ldots,n} |\Sigma_{n,i}|. \tag{3.25}$$

Theorem 3.3.5 *Consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that*

$$K_n \sim \lambda \log n \tag{3.26}$$

with $\lambda > 0$. Then, there exists $c(\lambda) > 0$ such that for any $c > c(\lambda)$ we have

$$\mathbb{P}\left[|M_n(K_n) - 2K_n| > cK_n\right] \leq n^{-(h(\lambda;c)+o(1))} \tag{3.27}$$

for all $n = 1, 2, \ldots$ with

$$h(\lambda; c) := -1 + \lambda a(c) > 0, \quad c(\lambda) < c. \tag{3.28}$$

Theorem 3.3.5 is established in Chapter 8.5. As with Lemma 3.3.4, Theorem 3.3.5 shows that with high probability, the maximum key ring size $M_n(K_n)$ will be less than $(c+2)K_n$ for all $c > c(\lambda)$. Arguments given in Chapter 8.5 show that if (3.26) is satisfied with $\lambda > \lambda^\star := (2\log 2 - 1)^{-1} \simeq 2.6$, then $c(\lambda) < 1$. The maximal key ring size is now seen to be less than $3K_n$ with high probability as we use (3.27) with $c = 1$. Finally, by an easy monotonicity argument, Theorem 3.3.5 implies that if $K_n = O(\log n)$ then all key rings have size $O(\log n)$ with high probability.

3.3.3 Connectivity results for the gradual deployment scenario

We continue our study of the connectivity properties of the pairwise scheme from a different perspective: In many applications, the sensor nodes are expected to be deployed gradually over time. Yet, the pairwise key distribution is an *offline* pairing mechanism which simultaneously involves all n nodes. Thus, once the network size n is set, there is no way to add more nodes to the network and still *recursively* expand the pairwise distribution scheme (as is possible for the EG scheme). However, as we explain below, the gradual deployment of a large number of sensor nodes is nevertheless feasible from a practical viewpoint. In that context we are now interested in understanding how the parameter K needs to scale with n large in order to ensure that connectivity is *maintained* a.a.s. throughout gradual deployment. We start by describing the implementation that allows the deployment of the pairwise scheme in multiple stages.

The implementation model: Initially, generate n node identities and construct key rings $\Sigma_{n,1}, \ldots, \Sigma_{n,n}$ as described in Chapter 3.1 – Here n stands for the maximum possible network size and should be selected large enough. This key selection procedure does not require the physical presence of the sensor entities and can be implemented completely on the software level. We now describe how this offline pairwise key distribution scheme can support gradual network deployment in consecutive stages. In the initial phase of deployment, with $0 < \gamma_1 < 1$, let $\lfloor \gamma_1 n \rfloor$ sensors be produced and given the labels $1, \ldots, \lfloor \gamma_1 n \rfloor$. The key rings $\Sigma_{n,1}, \ldots, \Sigma_{n,\lfloor \gamma_1 n \rfloor}$ are then inserted into the memory modules of the sensors $1, \ldots, \lfloor \gamma_1 n \rfloor$, respectively. Imagine now that more sensors are needed, say $\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor$ sensors with $0 < \gamma_1 < \gamma_2 \leq 1$. Then, $\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor$ additional sensors would be produced, this second batch of sensors would be assigned labels $\lfloor \gamma_1 n \rfloor + 1, \ldots, \lfloor \gamma_2 n \rfloor$, and the key rings $\Sigma_{n,\lfloor \gamma_1 n \rfloor + 1}, \ldots, \Sigma_{n,\lfloor \gamma_2 n \rfloor}$ would be inserted into their memory modules. Once this is done, these $\lfloor \gamma_2 n \rfloor - \lfloor \gamma_1 n \rfloor$ new sensors are added to the network (which now comprises $\lfloor \gamma_2 n \rfloor$ deployed sensors). This step may be repeated a number times: In fact, for some finite integer ℓ, consider positive scalars $0 < \gamma_1 < \ldots < \gamma_\ell \leq 1$ (with $\gamma_0 = 0$ by convention). We can then deploy the sensor network in ℓ consecutive phases, with the k^{th} phase adding $\lfloor \gamma_k n \rfloor - \lfloor \gamma_{k-1} n \rfloor$ new nodes to the network for each $k = 1, \ldots, \ell$.

The results: With the network deployed gradually over time as described above, we are interested in understanding how the parameter K needs to be scaled with large n to ensure that connectivity is *maintained* a.a.s. throughout gradual deployment. Consider positive integers $n = 2, 3, \ldots$ and K with $K < n$. With γ in the

interval $(0,1)$, let $\mathbb{H}_\gamma(n;K)$ denote the subgraph of $\mathbb{H}(n;K)$ restricted to the nodes $\{1,\ldots,\lfloor\gamma n\rfloor\}$. Given scalars $0<\gamma_1<\ldots<\gamma_\ell\leq 1$, we seek conditions on the parameters K and n such that $\mathbb{H}_{\gamma_i}(n;K)$ is a.a.s. connected for each $i=1,2,\ldots,\ell$.

First we write

$$P_\gamma(n;K):=\mathbb{P}\left[\mathbb{H}_\gamma(n;K)\text{ is connected}\right]=\mathbb{P}\left[C_{n,\gamma}(K)\right]$$

where $C_{n,\gamma}(K)$ denote the event that $\mathbb{H}_\gamma(n;K)$ is connected. The fact that $\mathbb{H}(n;K)$ is connected does *not* imply that $\mathbb{H}_\gamma(n;K)$ is necessarily connected. Indeed, with distinct nodes $i,j=1,\ldots,\lfloor\gamma n\rfloor$, the path that exists in $\mathbb{H}(n;K)$ between these nodes (as a result of the assumed connectivity of $\mathbb{H}(n;K)$) may comprise edges that are not in $\mathbb{H}_\gamma(n;K)$. The next result, established in Chapter 9.2.1, provides an analog of Corollary 3.3.3 in this new setting.

Theorem 3.3.6 *With γ in the unit interval $(0,1)$ and $c>0$, consider a scaling $K:\mathbb{N}_0\to\mathbb{N}_0$ such that*

$$K_n\sim c\,\frac{\log n}{\gamma}. \tag{3.29}$$

Then, we have

$$\lim_{n\to\infty}P_\gamma(n;K_n)=\begin{cases}0 & \text{if } c<r(\gamma)\\ \\ 1 & \text{if } c>1\end{cases} \tag{3.30}$$

where the threshold $r(\gamma)$ is given by

$$r(\gamma) := \left(1 - \frac{\log(1-\gamma)}{\gamma}\right)^{-1}. \tag{3.31}$$

Theorem 3.3.6 does not provide a full zero-one law for the connectivity of $\mathbb{H}_\gamma(n; K_n)$ as there is a gap between the threshold $r(\gamma)$ of the zero-law and the unit threshold of the one-law. Yet, it can easily be seen that the gap between the thresholds of the zero-law and the one-law is quite small with $\frac{1}{2} < 1 - r(\gamma) < 1$. More importantly, Theorem 3.3.6 already implies (via a monotonicity argument) that it is necessary *and* sufficient to keep the parameter K_n on the order of $\log n$ to ensure that the graph $\mathbb{H}_\gamma(n; K_n)$ is a.a.s. connected. It is worth pointing out that the simulation results in Section 9.6 indeed suggest the existence of a full zero-one law for $P_\gamma(n; K_n)$ with a threshold resembling $r(\gamma)$. This would not be surprising since in many known classes of random graphs, the absence of isolated nodes and graph connectivity are asymptotically equivalent properties, e.g., Erdős-Rényi graphs [4] and random key graphs [29, 30, 35], among others.

Finally we turn to gradual network deployment as discussed earlier.

Theorem 3.3.7 *With $0 < \gamma_1 < \gamma_2 < \ldots < \gamma_\ell \leq 1$, consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that*

$$K_n \sim c\, \frac{\log n}{\gamma_1} \tag{3.32}$$

for some $c > 1$. Then we have

$$\lim_{n\to\infty} \mathbb{P}\left[C_{n,\gamma_1}(K_n) \cap \ldots \cap C_{n,\gamma_\ell}(K_n)\right] = 1. \tag{3.33}$$

Theorem 3.3.7 will be established in Chapter 9.5. The event $[C_{\gamma_1,n}(K_n) \cap \ldots \cap C_{\gamma_\ell,n}(K_n)]$ corresponds to the network in *each* of its ℓ phases being connected as more nodes get added – In other words, on that event the sensors do form a connected network at each phase of deployment. As a result, we infer via Theorem 3.3.7 that the condition (3.32) (with $c > 1$) is enough to ensure that the network remains a.a.s. connected as more sensors are deployed over time. From a practical point of view, it is important to notice that the condition (3.32) requires only the knowledge of the maximal possible network size n and the fraction γ_1 of the nodes that will be deployed initially – If the network is connected at the initial stage, it will remain so throughout all stages of the deployment irrespective of the fractions $\gamma_2, \ldots, \gamma_\ell$ that are added later.

We can now combine Theorem 3.3.5 and Theorem 3.3.7, and arrive at the following conclusion:

Corollary 3.3.8 *With $0 < \gamma_1 < \gamma_2 < \ldots < \gamma_\ell \leq 1$, consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$*

such that $K_n = O(\log n)$ *with*

$$K_n \geq \max\left\{(\gamma_1)^{-1}, \lambda^\star\right\} \cdot \log n, \quad n = 2, 3, \ldots \tag{3.34}$$

Then, the following holds:

1) *The maximum number of keys kept in the memory module of each sensor will be a.a.s. less than* $3K_n$*;*

2) *The network deployed gradually in* ℓ *steps (as in Section 3.3.3) will be a.a.s. connected in each of the* ℓ *phases of deployment.*

We close by comparing the gradual deployment and the single-phase deployment in terms of the required values of K to achieve secure connectivity. It is already known via Theorem 3.3.2 that $K \geq 2$ is enough to ensure connectivity when all nodes in the network are deployed simultaneously. On the other hand, Figure 3.1 depicts the required value of K for achieving connectivity in the gradual deployment case as a function of the size γ_1 of the initial deployment. It can easily be seen that if γ_1 is small, the required K can be much larger than 2.

3.3.4 Connectivity under an ON-OFF channel

The results presented in the previous sections are obtained under the assumption of full visibility. Now, we complement them by accounting for the possibility that

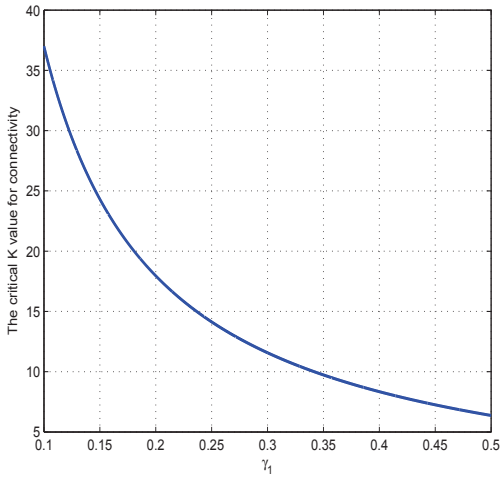


Figure 3.1: *The critical K value (according to Theorem 3.3.6) required to achieve secure connectivity at every step of the gradual deployment. The maximum possible network size is set to $n = 2000$ and only $\lceil \gamma_1 \cdot n \rceil$ sensors are deployed in the initial stage of the deployment. It is seen that if γ_1 is small, the required K can be much larger than that of the single-phase deployment case, namely $K = 2$.*

communication links between nodes may not be available. As for the EG scheme, we study such situations under a communication model that consists of independent channels each of which can be either on or off. Thus, the random graph model studied here is obtained by *intersecting* the random pairwise graph $\mathbb{H}(n; K)$ with the ER graph $\mathbb{G}(n; p)$. More precisely, in this new model, the distinct nodes i and j are said to be adjacent, written $i \sim j$, if and only they are both adjacent in the random pairwise graph (i.e., they share a key) and they are B-adjacent, namely

$$i \sim j \quad \text{iff} \quad \Sigma_{n,i} \cap \Sigma_{n,j} \neq \emptyset \ \text{ and } \ B_{ij}(p) = 1, \tag{3.35}$$

where $\{B_{ij}(p), 1 \leq i < j \leq n\}$ denote a collection of i.i.d. $\{0, 1\}$-valued rvs with success probability p. The resulting *undirected* random graph defined on the vertex

set $\{1,\ldots,n\}$ through this notion of adjacency is denoted $\mathbb{H}\cap\mathbb{G}(n;K,p)$.

Throughout, the collections of rvs $\{\Gamma_{n,1},\ldots,\Gamma_{n,n}\}$ and $\{B_{ij}(p),1\leq i<j\leq n\}$ are assumed to be independent, in which case the edge occurrence probability in $\mathbb{H}\cap\mathbb{G}(n;K,p)$ is given by

$$\mathbb{P}\left[i\sim j\right]=p\cdot\mathbb{P}\left[\Sigma_{n,i}\cap\Sigma_{n,j}\neq\emptyset\right]=p\lambda_n(K) \tag{3.36}$$

with $\lambda_n(K)$ defined in (3.6).

We now present connectivity results for the random graph $\mathbb{H}\cap\mathbb{G}(n;K,p)$. To fix the terminology any mapping $p:\mathbb{N}_0\to(0,1)$ defines a scaling for ER graphs.

The results will be expressed in terms of the threshold function $\tau:[0,1]\to[0,1]$ defined by

$$\tau(p)=\begin{cases}1 & \text{if } \ p=0\\ \frac{2}{1-\frac{\log(1-p)}{p}} & \text{if } 0<p<1\\ 0 & \text{if } p=1.\end{cases} \tag{3.37}$$

With this notation, the main results can be summarized as follows:

Theorem 3.3.9 *Consider scalings $K:\mathbb{N}_0\to\mathbb{N}_0$ and $p:\mathbb{N}_0\to(0,1)$ such that*

$$p_n\left(2K_n-\frac{K_n^2}{n-1}\right)\sim c\ \log n,\quad n=1,2,\ldots \tag{3.38}$$

for some $c > 0$. Assume also that $\lim_{n\to\infty} p_n = p^\star$ exists. Then, we have

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left[\mathbb{H} \cap \mathbb{G}(n; K_n, p_n) \textit{ contains no isolated nodes}\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left[\mathbb{H} \cap \mathbb{G}(n; K_n, p_n) \;\; \textit{is connected}\right] \\
&= \begin{cases} 0 & \textit{if} \;\; c < \tau(p^\star) \\ \\ 1 & \textit{if} \;\; c > \tau(p^\star) \end{cases}
\end{aligned} \tag{3.39}$$

with the threshold $\tau(p^\star)$ specified in (3.37).

Theorem 3.3.9 will be established in Chapter 10. We see that the class of random graphs studied here provides one more instance where the zero-one laws for absence of isolated nodes and connectivity coincide, viz. ER graphs [4], random geometric graphs [27] or the random key graphs [3, 29, 39] discussed in the previous chapter.

Theorem 3.3.2 and its Corollary 3.3.3 cannot be recovered from Theorem 3.3.9 whose zero-one laws are derived under the assumption $p_n < 1$ for all $n = 1, 2, \ldots$. Furthermore, even if the scaling $p : \mathbb{N}_0 \to (0, 1)$ were to satisfy $\lim_{n\to\infty} p_n = 1$, only the one-laws in Theorem 3.3.10 remain since $\tau(p^\star) = 0$ (and $\widehat{\tau}(p^\star) = 0$) at $p^\star = 1$. Although this might perhaps be expected given the aforementioned absence of isolated nodes in $\mathbb{H}(n; K)$, the one-laws for both the absence of isolated nodes and graph connectivity in $\mathbb{H} \cap \mathbb{G}(n; K, p)$ still require conditions on the behavior of the scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$, namely (3.41) (whereas Corollary 3.3.3 does not).

A particularly interesting case arises when $p^\star > 0$ since requiring (3.38) now

amounts to

$$\left(2K_n - \frac{K_n^2}{n-1}\right) \sim \frac{c}{p^\star} \log n \tag{3.40}$$

for some $c > 0$. Any scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ which behaves like (3.40) must necessarily satisfy $K_n = o(n)$, and it is easy to see that requiring (3.38) is equivalent to

$$K_n \sim t \log n \tag{3.41}$$

for some $t > 0$ with c and t related by $t = \frac{c}{2p^\star}$. With this reparametrization, Theorem 3.3.9 takes a simpler form:

Theorem 3.3.10 *Consider scalings $K : \mathbb{N}_0 \to \mathbb{N}_0$ and $p : \mathbb{N}_0 \to (0,1)$ such that $\lim_{n\to\infty} p_n = p^\star > 0$. Under the condition (3.41) for some $t > 0$, we have*

$$\begin{aligned} &\lim_{n\to\infty} \mathbb{P}\left[\mathbb{H} \cap \mathbb{G}(n; K_n, p_n) \textit{ contains no isolated nodes}\right] \\ &= \lim_{n\to\infty} \mathbb{P}\left[\mathbb{H} \cap \mathbb{G}(n; K_n, p_n) \;\; \textit{is connected}\right] \\ &= \begin{cases} 0 & \textit{if } \; t < \widehat{\tau}(p^\star) \\ \\ 1 & \textit{if } \; t > \widehat{\tau}(p^\star) \end{cases} \end{aligned} \tag{3.42}$$

where we have set

$$\widehat{\tau}(p) := \frac{\tau(p)}{2p} = \frac{1}{p - \log(1-p)}, \quad 0 < p \leq 1. \tag{3.43}$$

The alternate formulation given in Theorem 3.3.10 is particularly relevant for the case $p_n = p^\star$ (in $(0,1)$) for all $n = 1, 2, \ldots$, which captures situations when channel conditions are not affected by the number of users. This simplification does not occur in the more realistic case $p^\star = 0$ which corresponds to the situation where channel conditions are indeed influenced by the number of users in the system – The more users in the network, the more likely they will experience interferences from other users.

3.4 Discussion

3.4.1 Full-visibility vs On-Off model

Although the communication model considered here may be deemed simplistic, it does permit a complete analysis of the issues of interest, with the results already yielding a number of interesting observations: The obtained zero-one laws differ significantly from the corresponding results in the full visibility case. Thus, the communication model may have a significant impact on the dimensioning of the pairwise distribution algorithm, and this points to the need of possibly reevaluating guidelines developed under the full visibility assumption. Furthermore, simulations suggest that the zero-one laws obtained here for the on-off channel model may still be useful in dimensioning the pairwise scheme under the popular, and more realistic, disk model [19]; see Section 3.4.3.

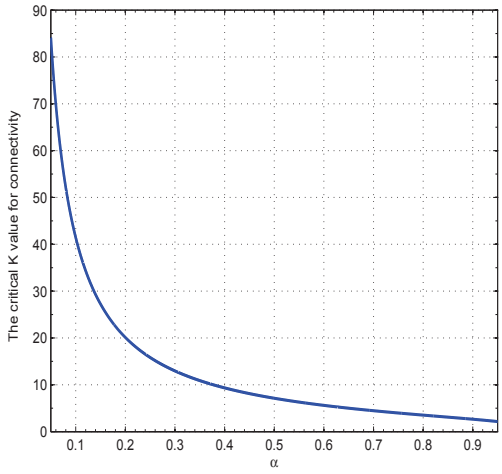


Figure 3.2: *The critical K value (according to Theorem 3.3.10) required to achieve secure connectivity in the on-off channel model. The network size is set to $n = 5000$ and the probability p of wireless link availability is varied from* 0.05 *to* 0.95. *It is seen that if the channel is poor, i.e., if p is small, the required K can be much larger than that of the full-visibility case, namely $K = 2$.*

To better asses the impact of the communication model on the connectivity of the pairwise scheme, we show in Figure 3.2 the critical K value (obtained via Theorem 3.42) required to achieve connectivity as a function of the channel parameter p. We see that if the channel conditions are poor, i.e., if the probability p of a wireless link being available is small, the required K value can be much larger than that of the full visibility case, namely $K = 2$.

3.4.2 Comparing $\mathbb{H} \cap \mathbb{G}(n; K, p)$ with ER graphs

In the original paper of Chan et al. [7] (as in the reference [21]), the connectivity analysis of the pairwise scheme was based on ER graphs [4] – It was assumed that the random graph induced by the pairwise scheme under a communication model (taken

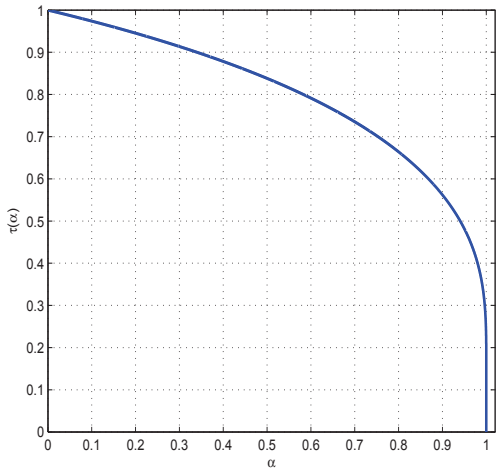


Figure 3.3: $\tau(p)$ *vs* p. *Clearly* $\tau(p^\star) = 1$ *only if* $\lim_{n\to\infty} p_n = p^\star = 0$.

mostly to be the disk model [19]) behaves *like* an ER graph; similar assumptions have been made in [14, 21] when discussing the connectivity of the EG scheme. However, this assumption was made without any formal justification. Here we have shown that the full visibility model $\mathbb{H}(n; K)$ has major differences with an ER graph. It is easy to verify that the edge assignments in $\mathbb{H} \cap \mathbb{G}(n; K, p)$ are negatively correlated (see Chapter 10.6) while independent in ER graphs. Therefore, the models $\mathbb{H}(n; K)$ and $\mathbb{H} \cap \mathbb{G}(n; K, p)$ cannot be equated with ER graphs, and the results obtained in this dissertation are *not* mere consequences of classical results for ER graphs.

However, *formal* similarities do exist between $\mathbb{H} \cap \mathbb{G}(n; K, p)$ and ER graphs. Recall the zero-one law (3.14)-(3.15) for ER graphs and observe that the condition (3.38) can be rephrased more compactly as

$$p_n \lambda_n(K_n) \sim c \, \frac{\log n}{n}, \qquad c > 0$$

with the result (3.39) remaining unchanged. Hence, in both ER graphs and $\mathbb{H} \cap \mathbb{G}(n; K, p)$, the zero-one laws can be expressed as a comparison of the probability of link assignment against the critical scaling $\frac{\log n}{n}$; this is also the case for random geometric graphs [27], and random key graphs [3, 29, 39]. Note however that the condition $c > \tau(p^\star)$ that ensures a.a.s. connectivity in $\mathbb{H} \cap \mathbb{G}(n; K, p)$ is not the same as the condition $c > 1$ for a.a.s. connectivity in ER graphs; see Figure 3.3. Thus, the connectivity behavior of the model $\mathbb{H} \cap \mathbb{G}(n; K, p)$ is in general different from that in an ER graph, and a "transfer" of the connectivity results from ER graphs cannot be taken for granted. Yet, the comparison becomes intricate when the channel is poor: The connectivity behaviors of the two models do match in the practically relevant case (for WSNs) $\lim_{n\to\infty} p_n = 0$ since $\tau(0) = 1$.

3.4.3 A more realistic communication model

One possible extension of the work presented here would be to consider a more realistic communication model; e.g., the popular disk model [19] which takes into account the geographical locations of the sensor nodes. For instance, assume that the nodes are distributed over a bounded region $\mathcal{D}$ of the plane. According to the *disk model*, nodes i and j located at $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$, respectively, in $\mathcal{D}$ are able to communicate if

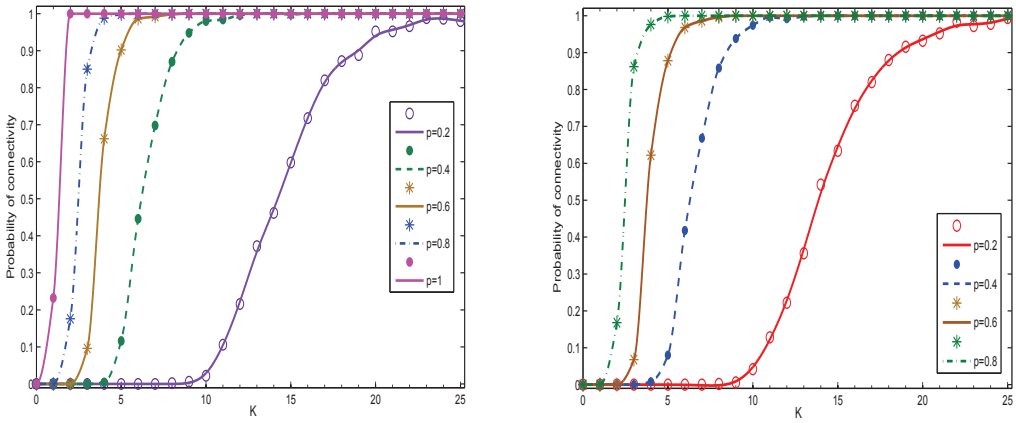
$$\| \boldsymbol{x_i} - \boldsymbol{x_j} \| < \rho \tag{3.44}$$

where $\rho > 0$ is called the transmission range. When the node locations are independently and randomly distributed over the region $\mathcal{D}$, the graph induced under the condition (3.44) is known as a random geometric graph [27], thereafter denoted $\mathbb{G}(n;\rho)$.

Under the disk model, studying the pairwise scheme of Chan et al. amounts to analyzing the intersection of $\mathbb{H}(n;K)$ and $\mathbb{G}(n;\rho)$, say $\mathbb{H}\cap\mathbb{G}(n;K,\rho)$. A direct analysis of this model seems to be very challenging; see below for more on this. However, limited simulations already suggest that the zero-one laws obtained here for $\mathbb{H}\cap\mathbb{G}(n;K,p)$ have an analog for the model $\mathbb{H}\cap\mathbb{G}(n;K,\rho)$. To verify this, consider n nodes distributed uniformly and independently over a folded unit square $[0,1]^2$ with toroidal (continuous) boundary conditions. Since there are no border effects, it is easy to check that

$$\mathbb{P}\left[\,\|\,\boldsymbol{x_i}-\boldsymbol{x_j}\,\|<\rho\,\right]=\pi\rho^2,\quad i\neq j,\ \ i,j=1,2,\ldots,n.$$

whenever $\rho<0.5$. We match the two communication models $\mathbb{G}(n;p)$ and $\mathbb{G}(n;\rho)$ by requiring $\pi\rho^2=p$. Then, we fix the number of nodes at $n=200$ and consider the channel parameters $p=0.2$, $p=0.4$, $p=0.6$, $p=0.8$, and $p=1$ (the full visibility case), while varying the parameter K from 1 to 25. For each parameter pair (K,p), we generate 500 independent samples of the graphs $\mathbb{H}\cap\mathbb{G}(n;K,p)$ and $\mathbb{H}\cap\mathbb{G}(n;K,\rho)$ while we count the number of times (out of a possible 500) that the obtained graphs are connected. Dividing the counts by 500, we obtain the (empirical) probabilities



(a) (b)

Figure 3.4: *a) Probability that* $\mathbb{H} \cap \mathbb{G}(n; K, p)$ *is connected as a function of* K *for* $p = 0.2$*,* $p = 0.4$*,* $p = 0.6$*,* $p = 0.8$ *and* $p = 1$ *with* $n = 200$*. b) Probability that* $\mathbb{H} \cap \mathbb{G}(n; K, \rho)$ *is connected as a function of* K*. The number of nodes is set to* $n = 200$ *and* ρ *is given by* $\pi\rho^2 = p$*. This figure clearly resembles Figure 3.4(a) for all* $p \neq 1$*.*

for $\mathbb{H} \cap \mathbb{G}(n; K, p)$ and $\mathbb{H} \cap \mathbb{G}(n; K, \rho)$ being connected for various values of K and p. The results are depicted in Figure 3.4(a) (on-off channel model) and Figure 3.4(b) (disk model). Clearly, these two figures are almost indistinguishable suggesting that the connectivity behaviors of the models $\mathbb{H} \cap \mathbb{G}(n; K, p)$ and $\mathbb{H} \cap \mathbb{G}(n; K, \rho)$ are quite similar. This raises the possibility that the results obtained here for the on/off communication model can also be used for dimensioning the pairwise scheme under the disk model.

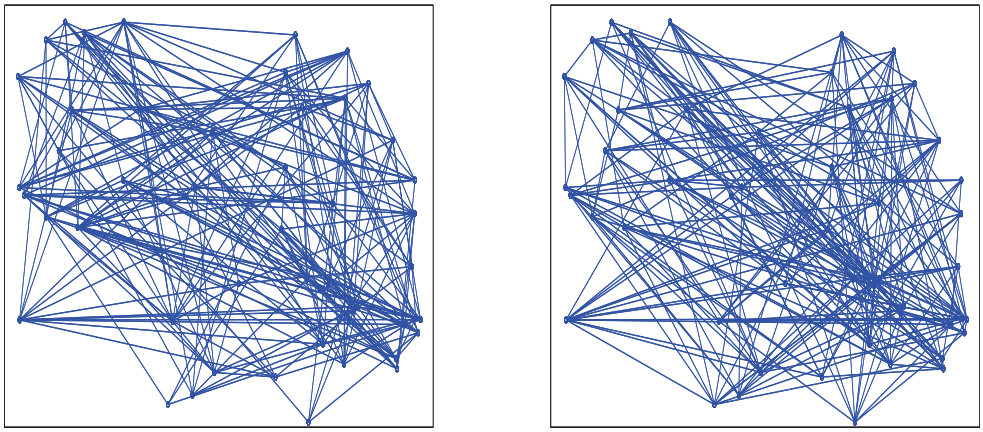
A complete analysis of $\mathbb{H} \cap \mathbb{G}(n; K, \rho)$ is likely to be very challenging given the difficulties already encountered in the analysis of similar problems. For example, the intersection of random geometric graphs with ER graphs was considered in [2, 48]. Although zero-one laws for graph connectivity are available for each component random graph, the results for the intersection model in [2,48] were limited only to the

absence of isolated nodes; the connectivity problem is still open for that model. Yi et al. [48] also consider the intersection of random key graphs with random geometric graphs, but these results are again limited to the property of node isolation. To the best of our knowledge, Theorem 3.3.2, together with Theorem 2.3.5, reported here constitute the only zero-one laws for graph connectivity in a model formed by intersecting multiple random graphs! (Except of course the trivial case where an ER graph intersects another ER graph.)

3.4.4 Intersection of random graphs

When using random graph models to study networks, situations arise where the notion of adjacency between nodes reflects multiple constraints. This can be so even when dealing with networks other than WSNs. As was the case here, such circumstances call for studying models which are constructed by taking the intersection of multiple random graphs. However, as pointed out earlier, the availability of results for each component model does not necessarily imply the availability of results for the intersection of these models; see the examples provided in the previous section.

Figures 3.5-3.6 can help better understand the relevant issues as to why this is so: Figure 3.5(a) provides a sample of an ER graph $\mathbb{G}(n, p)$ with $n = 200$ and $p = 0.2$. As would be expected from classical results, the obtained graph is very densely connected. Similarly, Figure 3.5(b) provides a sample of the pairwise random graph $\mathbb{H}(n; K)$ with $n = 200$ and $K = 5$. In line with Theorem 3.3.2, the obtained graph is connected.



(a) (b)

Figure 3.5: *a) An instantiation of ER graph $\mathbb{G}(n;p)$ with $n = 50$ and $p = 0.2$.– The graph is connected. b) An instantiation of $\mathbb{H}(n;K)$ with $n = 50$ and $K = 5$.– The graph is connected.*

On the other hand, the graph formed by intersecting these graphs turn out to be *disconnected* as shown in Figure 3.6.

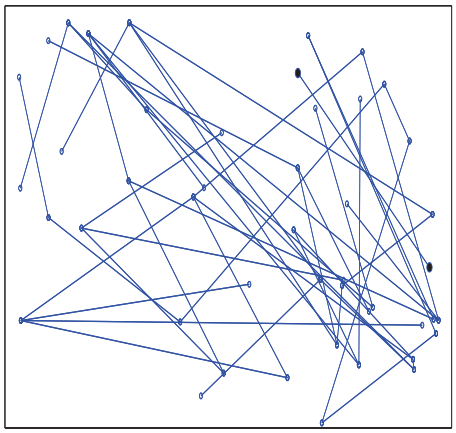


Figure 3.6: *The intersection $\mathbb{H} \cap \mathbb{G}(n;K,p)$ of the graphs in Figure 3.5(a) and Figure 3.5(b) – The graph is disconnected as the marked nodes form a component!*

To drive this point further, consider the constant parameter case for the models

$\mathbb{H}(n;K)$ and $\mathbb{G}(n;p)$, a case which cannot be recovered from Theorem 3.3.9. Nevertheless, Theorem 3.3.2 yields

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{H}(n;K) \text{ is connected}\right] = 1, \quad K \geq 2$$

while it is well known [4] that

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{G}(n;p) \text{ is connected}\right] = 1, \quad 0 < p < 1.$$

However, it can be shown that

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{H} \cap \mathbb{G}(n;K,p) \text{ contains no isolated nodes}\right] = 0 \tag{3.45}$$

whence

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{H} \cap \mathbb{G}(n;K,p) \text{ is connected}\right] = 0 \tag{3.46}$$

for the same ranges of values for p and K; for details see the discussion at the end of Section 10.7. This clearly provides a *non-trivial* example (one that is not for an ER intersecting an ER graph) where the intersection of two random graphs is indeed a.a.s. *not* connected although each of the components is a.a.s. connected.

Chapter 4

A comparison of the EG scheme and the pairwise scheme

4.1 Introduction

The main goal of this dissertation is to compare the EG scheme and the pairwise scheme based on the metrics introduced in Chapter 1.3. As most of the relevant results have already been presented in the previous two chapters, we are now in a position to make this comparison. In particular, we will compare the number of keys required in each sensor's memory to ensure certain desired properties in the network. As explained in Chapter 1.3, we will look for connectivity, scalability and security.

In what follows we use $|\Sigma|_{\text{EG}}$ for the key ring size of each sensor participating in the EG scheme. In other words, according to the notation used in Chapter 2, we have $|\Sigma|_{\text{EG}} = K$. Similarly, we let $|\Sigma|_{\text{Avg}}$ be the *average* key ring size in the pairwise scheme given in (3.19). Finally, for convenience, we let $|\Sigma|_{\text{Max}}$ denote the *maximum* key ring size in the pairwise scheme, i.e., we set $|\Sigma|_{\text{Max}} = M_n(K)$ with $M_n(K)$ defined in (3.25).

4.2 Connectivity

We start with the connectivity properties.

Full visibility: First of all, it is important to notice that in principle the EG scheme can yield connectivity even when each sensor has only one key – Just let the pool size P equals to one. However, such cases are not interesting in realistic WSN scenarios as the vulnerability of the network against node capture attacks will be extremely high [8, 9]. Indeed, for the EG scheme to be operationally useful, it was stated in [10] that the parameters $|\Sigma|_{\mathrm{EG,n}}$ and P_n should satisfy

$$\frac{|\Sigma|_{\mathrm{EG,n}}}{P_n} \sim \frac{1}{n}. \tag{4.1}$$

Thus, throughout this chapter, we assume that the condition (4.1) is satisfied.

Now, recall from Corollary 2.3.3 that the random key graph will be a.a.s. connected if we have

$$\frac{|\Sigma|^2_{\mathrm{EG,n}}}{P_n} \sim c\frac{\log n}{n}.$$

for some $c > 1$. Under (4.1), this amounts to having key rings with

$$|\Sigma|_{\mathrm{EG,n}} \sim c \log n, \quad c > 1,$$

whence we conclude that the required key ring size to achieve connectivity (under full visibility) is $O(\log n)$ under the EG scheme.

Under the pairwise scheme, we know via Theorem 3.3.2 that the random pairwise graph is a.a.s. connected whenever the scheme parameter K satisfies $K \geq 2$. In view of (3.19), $K = 2$ yields an average key ring size $|\Sigma|_{\text{Avg}} = 4$, while Theorem 3.3.5 ensures that the maximum size of a key ring will be $O(\log n)$. Thus, under full visibility, the pairwise scheme can yield connectivity with key rings having size on the order of what is required for the EG scheme.

On-Off communication model: Now consider the partial visibility case and assume as in Chapter 2.3.4 and Chapter 3.3.4 that wireless links are independent from each other and each is either *on* with probability p or *off* with probability $1 - p$. In the case of the EG scheme, we see from Theorem 2.3.5 that the resulting random graph (i.e., the intersection of the random key graph and Erdős-Rényi graph) will be a.a.s. connected if we have

$$p_n(1 - q(\theta_n)) \sim c\frac{\log n}{n} \tag{4.2}$$

for some $c > 0$. Furthermore, since key ring sizes are expected to be much smaller than network size [14], i.e., $|\Sigma|_{\text{EG},n} = o(n)$, it is not unreasonable to assume that (4.1) implies

$$\lim_{n\to\infty} \frac{|\Sigma|^2_{\text{EG},n}}{P_n} = 0, \tag{4.3}$$

whence

$$1 - q(\theta_n) \sim \frac{|\Sigma|^2_{\text{EG},n}}{P_n}$$

in view of Lemma 7.4.4. As a result, under (4.1)-(4.3) the condition (4.2) is equivalent to

$$|\Sigma|_{\mathrm{EG},n} \sim c\frac{\log n}{p_n} \tag{4.4}$$

and the EG scheme will yield secure connectivity whenever the condition (4.4) is satisfied for some $c > 1$.

In the case of the pairwise scheme, we have from Theorem 3.3.9 that the induced random graph (namely, the random pairwise graph intersecting an Erdős-Rényi graph) will be a.a.s. connected whenever the condition

$$p_n \left(2K_n - \frac{K_n^2}{n-1} \right) \sim c \ \log n \tag{4.5}$$

is satisfied for some $c > \tau(p^\star)$ where $\tau(p^\star) < 1$. As in the case of the EG scheme, if we assume that the average key ring size $2K_n$ is much smaller than the network size n, i.e., $K_n = o(n)$, the condition (4.5) amounts to having

$$p_n \cdot 2K_n \sim c \log n$$

so that the induced random graph will be a.a.s. connected if

$$|\Sigma|_{\mathrm{Avg},n} \sim c\frac{\log n}{p_n} \tag{4.6}$$

for some $c > 1$. It follows via Theorem 3.3.5 that the maximal key ring size $|\Sigma|_{\mathrm{Max},n}$ will also be on the order $O\left(\frac{\log n}{p_n}\right)$ under the condition (4.6). As a result, we see from

(4.4) that the pairwise scheme can yield secure connectivity under the on-off channel model with all key ring sizes being on the order of what is required for the EG scheme.

4.3 Scalability

The EG scheme inherently supports the gradual deployment of sensors regardless of the parameter choice. Therefore the required key ring size for connectivity in the gradual deployment case is the same with the simultaneous deployment case discussed in the previous section. However, as seen from Chapter 3.3.3, for the pairwise scheme we need to have

$$K_n \sim c \log n \tag{4.7}$$

for some $c > 1$ in order to achieve connectivity at each phase of the gradual deployment. Under the condition (4.7), we have $|\Sigma|_{\text{Avg},n} = 2K_n = O(\log n)$ whereas Theorem 3.3.5 yields $|\Sigma|_{\text{Max},n} = O(\log n)$. Thus, key rings with size $O(\log n)$ are sufficient to achieve connectivity of the pairwise scheme in the case of gradual deployment.

4.4 Security

Recall that conditions (1.2) and (1.3) are defined as the unassailability and unsplittability, respectively. These properties were investigated in [26] for the EG scheme and the main results can be summarized as follows:

Theorem 4.4.1 *Consider any scaling $\Sigma_{\mathrm{EG}}, P : \mathbb{N}_0 \to \mathbb{N}_0$ for the EG scheme which satisfies*

$$P_n = n^{\delta} \tag{4.8}$$

for some $\delta > 2$. Then, with high probability, (1.2) and (1.3) hold.

It is claimed in [26] that the same arguments also hold when $P_n \sim n \log n$, but no proof of is provided for that case. If indeed $P_n \sim n \log n$ is enough to ensure both unassailability and unsplittability, the key ring size must satisfy

$$|\Sigma|_{\mathrm{EG},n} \sim \log n$$

whenever (4.1) holds. Thus, key rings with size (at least) on the order $\log n$ is required to ensure unassailability and unsplittability of the EG scheme.

For the pairwise scheme, we have established in [44] the analogous version of Theorem 4.4.1:

Theorem 4.4.2 *Consider any scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$. With high probability, (1.2) always holds, whereas (1.3) is satisfied whenever*

$$\lim_{n \to \infty} K_n = \infty. \tag{4.9}$$

By Theorem 4.4.2, the pairwise scheme can ensure the unassailability of the network for any parameter K. In other words, the smallest key rings required to achieve unassailability satisfies

$$|\Sigma|_{\text{Avg}} = O(1).$$

On the other hand, as seen via Theorem 3.3.5, the maximal key ring size will satisfy

$$|\Sigma|_{\text{Max},n} = O(\log n)$$

under the minimum requirements for the unassailability of the network.

Now, pick any function $w : \mathbb{N}_0 \to \mathbb{N}_0$ satisfying $\lim_{n\to\infty} w_n = \infty$. It is clear from Theorem 4.4.2 that the unsplittability of the network (under the pairwise scheme) is ensured whenever

$$|\Sigma|_{\text{Avg},n} = O(w_n).$$

Clearly, one can select w_n such that $w_n = O(\log n)$. Thus, in view of Theorem 3.3.5, the pairwise scheme can ensure unsplittability with maximal key rings of size

$$|\Sigma|_{\text{Max},n} = O(\log n). \tag{4.10}$$

Recalling the discussion given above, we conclude that the pairwise scheme can ensure unassailability and unsplittability with key ring sizes that is comparable to what is required in the EG scheme.

Finally, we refer the reader to [7] for a discussion of the operational advantages of the pairwise scheme (over the EG scheme) such as node-to-node authentication, key revocation and perfect resiliency.

4.5 Summary and discussion

Table 4.5 gives a summary of the comparison between the EG scheme and the pairwise scheme. First of all, it is easy to see that the pairwise scheme yields connectivity in the full visibility case with much smaller key rings (on average) than the EG scheme, whereas the two schemes have the same memory requirement for achieving connectivity under the more realistic on-off channel model. On the other hand, both unassailability and unsplittability can be achieved by the pairwise scheme with key rings being much smaller (on average) than required by the EG scheme. Yet, for both schemes, key ring sizes needed to ensure connectivity under the on-off model yields already the unassailability and unsplittability of the network. In other words, both schemes require key rings of size $O\left(\frac{\log n}{p_n}\right)$ to achieve (at the same time) unassailability, unsplittability and connectivity under an on-off channel with parameter p_n. Therefore, operational advantages and disadvantages can play a key role in deciding the usefulness of one scheme over the other.

To dig into this further, we remind that the main advantage of the EG scheme is its *decentralized* operation in that no central authority is required for the deployment of the network. A major benefit of this characteristic is that the gradual deployment

of the network is always possible regardless of the particular parameter selection. However, as discussed in Chapter 3.3.3, the gradual deployment of the pairwise scheme is possible only when the maximal network size is set in advance and the scheme parameter is selected in a very specific manner. Even then, the pairwise scheme requires a central authority which maintains the sensor id's and key rings. Therefore, in WSN applications where decentralization is necessary, the EG scheme would be a better choice.

On the other hand, the pairwise scheme has the advantage of supporting node-to-node authentication so that sensors can identify the neighbors with which they are communicating. This is a major advantage in terms of network security since node-to-node authentication may help detect node misbehavior, and provides resistance against node replication attacks [7]. Also, since keys are unique to a particular communication link, if an adversary captures a group of sensors, it can compromise *only* the links that are incident to the captured sensors; this property is usually referred to as *perfect resiliency*. This is in contrast with the EG scheme where keys are drawn from a common key pool and the capture of sensor i can cause a link $j \sim k$ to be compromised; i.e., the EG scheme is not perfectly resilient. In short, if node capture and node replication attacks are major concerns in a WSN, then the pairwise scheme of Chan et al. would be a better choice than the EG scheme.

	EG Scheme	Pairwise Scheme
Connectivity (Full Visibility)	$\lvert\Sigma\rvert_{\mathrm{EG}} = O(\log n)$	$\lvert\Sigma\rvert_{\mathrm{Avg}} = O(1)$ $\lvert\Sigma\rvert_{\mathrm{Max}} = O(\log n)$
Connectivity (On-Off Model, p_n)	$\lvert\Sigma\rvert_{\mathrm{EG}} = O(\frac{\log n}{p_n})$	$\lvert\Sigma\rvert_{\mathrm{Avg}} = O(\frac{\log n}{p_n})$ $\lvert\Sigma\rvert_{\mathrm{Max}} = O(\frac{\log n}{p_n})$
Gradual Deployment	$\surd$	$\lvert\Sigma\rvert_{\mathrm{Avg}} = O(\log n)$ $\lvert\Sigma\rvert_{\mathrm{Max}} = O(\log n)$
Unassailability	$\lvert\Sigma\rvert_{\mathrm{EG}} = O(\log n)$	$\lvert\Sigma\rvert_{\mathrm{Avg}} = O(1)$ $\lvert\Sigma\rvert_{\mathrm{Max}} = O(\log n)$
Unsplittability	$\lvert\Sigma\rvert_{\mathrm{EG}} = O(\log n)$	$\lvert\Sigma\rvert_{\mathrm{Avg}} = O(w_n)$ $\lvert\Sigma\rvert_{\mathrm{Max}} = O(\log n)$
Perfect Resiliency	$\times$	$\surd$
Node Authentication	$\times$	$\surd$
Key Revocation	$\times$	$\surd$

Table 4.1: *Summary of the comparison between the EG scheme and the pairwise scheme given in terms of the required key ring size to ensure a desired property. Here, n denotes the number of sensor nodes in the network, p_n denotes the link existence probability under the on-off channel model, and w_n stands for any function satisfying $\lim_{n\to\infty} w_n = \infty$. The conditions for unassailability and unsplittability are derived so as to ensure also that the networks are a.a.s. connected (under full visibility).*

Chapter 5

Mathematical Tools

In this chapter we collect a number of technical facts and mathematical tools that will be used in establishing the main results of the dissertation.

5.1 Method of first and second moments

In the course of establishing zero-one laws for certain graph properties, we often rely on the method of first and second moments. The method of first moments [22, Eqn (3.10), p. 55] is a simple application of Markov's inequality for integer-valued rvs.

Lemma 5.1.1 *For any $\mathbb{N}$-valued rv Z, we have*

$$\mathbb{P}[Z = 0] \geq 1 - \mathbb{E}[Z]. \tag{5.1}$$

The method of second moment [22, Remark 3.1, p. 55] is a simple corollary of the Cauchy-Schwartz inequality.

Lemma 5.1.2 *For any* $\mathbb{N}$*-valued random variable* Z *with* $0 < \mathbb{E}[Z^2] < \infty$*, we have*

$$\mathbb{P}[Z = 0] \leq 1 - \frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]}. \tag{5.2}$$

Proof. By the Cauchy-Schwartz inequality, it is a simple matter to check that

$$\mathbb{E}[Z]^2 = \mathbb{E}[\mathbf{1}[Z \neq 0] Z]^2 \leq \mathbb{E}\left[\mathbf{1}[Z \neq 0]^2\right] \mathbb{E}\left[Z^2\right],$$

whence

$$\frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]} \leq \mathbb{P}[Z \neq 0]. \tag{5.3}$$

■

In due course, we often encounter the following situation: Given a collection $\{\chi_{n,i},\ n = 1, 2, \ldots,\ \ i = 1, \ldots, m_n\}$ of $\{0, 1\}$-valued rvs, the rvs $\{Z_n, n = 1, \ldots\}$ are defined by

$$Z_n = \sum_{i=1}^{m_n} \chi_{n,i}$$

where for each n, the rvs $\{\chi_{n,i}, i = 1, \ldots, m_n\}$ are identically distributed and the sequence $m : \mathbb{N}_0 \to \mathbb{N}_0$ satisfies

$$\lim_{n\to\infty} m_n = \infty. \tag{5.4}$$

Fix $n = 1, 2, \ldots$. The rvs $\chi_{n,1}, \ldots, \chi_{n,m_n}$ being identically distributed, we find

$$\mathbb{E}[Z_n] = m_n \mathbb{E}[\chi_{n,1}] \tag{5.5}$$

and

$$\mathbb{E}\left[Z_n^2\right] = m_n \mathbb{E}[\chi_{n,1}] + \sum_{1\le i,j\le m_n,\, i\neq j} \mathbb{E}[\chi_{n,i}\chi_{n,j}]$$

by the binary nature of the rvs involved. In some important special cases, we also have

$$\mathbb{E}[\chi_{n,i}\chi_{n,j}] = \mathbb{E}[\chi_{n,1}\chi_{n,2}], \qquad 1 \le i < j \le m_n, \tag{5.6}$$

whence

$$\frac{\mathbb{E}[Z_n^2]}{\mathbb{E}[Z_n]^2} = \frac{1}{m_n \mathbb{E}[\chi_{n,1}]} + \frac{m_n - 1}{m_n} \cdot \frac{\mathbb{E}[\chi_{n,1}\chi_{n,2}]}{(\mathbb{E}[\chi_{n,1}])^2}. \tag{5.7}$$

From (5.1) and (5.5) we see that the one-law $\lim_{n\to\infty} \mathbb{P}[Z_n = 0] = 1$ will then be established if we show that

$$\lim_{n\to\infty} m_n \mathbb{E}[\chi_{n,1}] = 0. \tag{5.8}$$

It is also plain from (5.2), (5.4) and (5.6) that the zero-law $\lim_{n\to\infty} \mathbb{P}[Z_n = 0] = 0$ holds if

$$\lim_{n\to\infty} m_n \mathbb{E}[\chi_{n,1}] = \infty \tag{5.9}$$

and

$$\limsup_{n\to\infty} \left(\frac{\mathbb{E}[\chi_{n,1}\chi_{n,2}]}{(\mathbb{E}[\chi_{n,1}])^2} \right) \leq 1. \tag{5.10}$$

Usually, the proof of a zero-one law passes through a number of technical propositions which establish (5.8), (5.9) and (5.10) under the appropriate conditions.

5.2 A basic union bound

One of our main objectives is to show that a random graph $\mathbb{G}(n;\upsilon)$ is connected with high probability (as n gets large) under certain conditions. In most cases, we accomplish this by first showing that $\mathbb{G}(n;\upsilon)$ does not contain any isolated nodes. Then, we derive an upper bound for the probability that $\mathbb{G}(n;\upsilon)$ is *not* connected and yet does not have any isolated nodes. The desired one-law follows by showing that the upper bound approaches to zero under the enforced assumptions. This method is similar to the one used for proving the one-law for connectivity in Erdős-Rényi graphs [4, p. 164] [32, p. 304] and will provide the basis of the approach we have followed in many cases.

For ease of presentation, here we derive the aforementioned bound for the general case: Consider a random graph $\mathbb{G}(n;\upsilon)$ on the vertex set $\{1,\ldots,m_n\}$ for some m_n :

$\mathbb{N}_0 \to \mathbb{N}_0$. Fix $m_n = 2, 3, \ldots$. For the time being, we omit the dependence of $\mathbb{G}$ (and the events related to it) on the parameters n and v. For any non-empty subset S of nodes, i.e., $S \subseteq \{1, \ldots, m_n\}$, we define the graph $\mathbb{G}(S)$ (with vertex set S) as the subgraph of $\mathbb{G}$ restricted to the nodes in S. We also say that S is *isolated* in $\mathbb{G}$ if there are no edges (in $\mathbb{G}$) between the nodes in S and the nodes in the complement $S^c = \{1, \ldots, m_n\} - S$. This is characterized by

$$(i \sim j)^c, \quad i \in S,\ j \in S^c$$

with $i \sim j$ denoting the event that there exists an edge between the nodes i and j (in $\mathbb{G}$).

With each non-empty subset S of nodes, we associate several events of interest: Let $C_n(S)$ denote the event that the subgraph $G(S)$ is itself connected. We also introduce the event $B_n(S)$ to capture the fact that S is isolated in G, i.e.,

$$B_n(S) := \left[(i \sim j)^c, \quad i \in S,\ j \in S^c\right].$$

Finally, we set

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

The starting point of the discussion is the following basic observation: If G is *not* connected and yet has *no* isolated nodes, then there must exist a subset S of nodes with $|S| \geq 2$ such that $G(S)$ is connected while S is isolated in G. This is captured

by the inclusion

$$C_n^c \cap I_n \subseteq \cup_{S \in \mathcal{N}:\ |S| \geq 2}\ A_n(S) \tag{5.11}$$

with $\mathcal{N}_{m_n}$ denoting the collection of all non-empty subsets of $\{1, \ldots, m_n\}$, C_n denoting the event that G is connected, and I_n denoting the event that G has no isolated nodes. A moment of reflection should convince the reader that this union need only be taken over all subsets S of $\{1, \ldots, m_n\}$ with $2 \leq |S| \leq \lfloor \frac{m_n}{2} \rfloor$. Then, a standard union bound argument immediately gives

$$\begin{aligned} \mathbb{P}\left[C_n^c \cap I_n\right] &\leq \sum_{S \in \mathcal{N}: 2 \leq |S| \leq \lfloor \frac{m_n}{2} \rfloor} \mathbb{P}\left[A_n(S)\right] \\ &= \sum_{r=2}^{\lfloor \frac{m_n}{2} \rfloor} \left(\sum_{S \in \mathcal{N}_r} \mathbb{P}\left[A_n(S)\right] \right) \end{aligned} \tag{5.12}$$

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

For each $r = 1, \ldots, m_n$, we can simplify the notation by writing $A_{n,r} := A_n(\{1, \ldots, r\})$, $B_{n,r} := B_n(\{1, \ldots, r\})$ and $C_{n,r} := C_n(\{1, \ldots, r\})$. As defined before, for $r = m_n$ we use C_n with a slight abuse of notation. For all random graphs under consideration, we have by exchangeability that

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

and the expression

$$\sum_{S\in\mathcal{N}_r} \mathbb{P}\left[A_n(S)\right] = \binom{m_n}{r} \mathbb{P}\left[A_{n,r}\right] \tag{5.13}$$

follows since $|\mathcal{N}_r| = \binom{m_n}{r}$. Substituting into (5.12) we obtain the key bound

$$\mathbb{P}\left[C_n^c \cap I_n\right] \leq \sum_{r=2}^{\lfloor \frac{m_n}{2} \rfloor} \binom{m_n}{r} \mathbb{P}\left[A_{n,r}\right]. \tag{5.14}$$

If the information that $\mathbb{G}$ does not have any isolated nodes is dropped, the union at (5.11) has to be taken over all subsets S with $|S| \geq 1$, and by the same arguments, we also have

$$\mathbb{P}\left[C_n^c\right] \leq \sum_{r=1}^{\lfloor \frac{m_n}{2} \rfloor} \binom{m_n}{r} \mathbb{P}\left[A_{n,r}\right]. \tag{5.15}$$

5.3 A useful decomposition

With $0 \leq x < 1$, it is a simple matter to check that

$$\log(1-x) = -\int_0^x \frac{1}{1-t} dt = -x - \Psi(x) \tag{5.16}$$

where we have set

$$\Psi(x) := \int_0^x \frac{t}{1-t} dt, \quad 0 \leq x < 1. \tag{5.17}$$

L'Hospital's rule yields

$$\lim_{x\downarrow 0} \frac{\Psi(x)}{x^2} = \frac{1}{2}, \tag{5.18}$$

while the decomposition (5.16) and the non-negativity of Ψ lead to the standard bound

$$1 - x \leq e^{-x}, \quad x \in [0, 1]. \tag{5.19}$$

5.4 Bounding the factorials

The following simple bounds will prove useful in a number of places.

Lemma 5.4.1 *For positive integers K, L and P such that $K + L \leq P$, we have*

$$\left(1 - \frac{L}{P - K}\right)^K \leq \frac{\binom{P-L}{K}}{\binom{P}{K}} \leq \left(1 - \frac{L}{P}\right)^K, \tag{5.20}$$

whence

$$\frac{\binom{P-L}{K}}{\binom{P}{K}} \leq e^{-K \cdot \frac{L}{P}}. \tag{5.21}$$

Proof. Under the condition $K + L \leq P$, the relation

$$\frac{\binom{P-L}{K}}{\binom{P}{K}} = \frac{(P-L)!}{(P-L-K)!} \cdot \frac{(P-K)!}{P!} \tag{5.22}$$

holds with

$$\frac{(P-jL)!}{(P-jL-K)!} = \prod_{\ell=0}^{K-1} (P - jL - \ell), \quad j = 0, 1.$$

Upon substituting we find

$$\frac{\binom{P-L}{K}}{\binom{P}{K}} = \prod_{\ell=0}^{K-1}\left(1-\frac{L}{P-\ell}\right) \tag{5.23}$$

and the bounds (5.20) are now immediate from the inequalities $P-K < P-\ell \le P$, $\ell = 0,\ldots,K-1$ and the easy bound (5.19). ■

Also, for $0 \le K \le x \le y$, we have

$$\frac{\binom{x}{K}}{\binom{y}{K}} = \prod_{\ell=0}^{K-1}\left(\frac{x-\ell}{y-\ell}\right) \le \left(\frac{x}{y}\right)^K \tag{5.24}$$

since $\frac{x-\ell}{y-\ell}$ decreases as ℓ increases from $\ell = 0$ to $\ell = K-1$.

Finally, the standard bounds

$$\binom{n}{r} \le \left(\frac{en}{r}\right)^r, \qquad \begin{matrix} r = 1,\ldots,n \\ n = 1,2,\ldots \end{matrix} \tag{5.25}$$

will be used throughout.

Chapter 6

Connectivity in random key graphs I: Node isolation

6.1 Introduction

As discussed in Chapter 2.3.1, the study of the connectivity properties of random key graphs resulted in the conjectures (2.11) and (2.12) (under (2.14)) that provide zero-one laws for the properties of absence of isolated nodes and connectivity, respectively. In this chapter, we start with the property of node isolation and establish the conjectured zero-one law (2.11). Namely, we will establish Theorem 2.3.1 which is restated here for the ease of exposition:

Theorem 2.3.1 *For any admissible pair of functions $P, K : \mathbb{N}_0 \to \mathbb{N}$, it holds that*

$$\lim_{n\to\infty} P^\star(n;\theta_n) = \begin{cases} 0 & \text{if } \lim_{n\to\infty} \alpha_n = -\infty \\ \\ 1 & \text{if } \lim_{n\to\infty} \alpha_n = +\infty \end{cases} \tag{6.1}$$

where the function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ is determined through (2.14).

6.2 An outline of the proof of Theorem 2.3.1

Consider $\theta = (K, P)$ with positive integers K and P such that $K \leq P$. Fix $n = 2, 3, \ldots$ and write

$$\chi_{n,i}(\theta) := \mathbf{1}\left[\text{Node } i \text{ is isolated in } \mathbb{K}_n(\theta)\right], \quad i = 1, \ldots, n \tag{6.2}$$

The number of isolated nodes in $\mathbb{K}(n;\theta)$ is simply given by

$$I(n;\theta) := \sum_{i=1}^{n} \chi_{n,i}(\theta).$$

The random graph $\mathbb{K}(n;\theta)$ has no isolated nodes if $I(n;\theta) = 0$, in which case

$$P^{\star}(n;\theta) = \mathbb{P}\left[I(n;\theta) = 0\right]. \tag{6.3}$$

The equivalence (6.3) provides the basis for proving Theorem 2.3.1 by means of the method of first and second moments; see Chapter 5.1. Here, observe that the rvs $\chi_{n,1}(\theta), \ldots, \chi_{n,n}(\theta)$ are exchangeable and therefore (5.6) holds. This yields (5.5) and (5.7) with Z_n replaced by the count variable $I(n;\theta)$, the index m_n replaced by n, and the indicator variables $\{\chi_{m_n,i}, i = 1, \ldots, m_n\}$ replaced by $\{\chi_{n,i}(\theta)$, i=1,. . . , n$\}$ as defined in (6.2). Thus, Theorem 2.3.1 will be established upon proving the next two technical lemmas which provide the appropriate versions of (5.8), (5.9) and (5.10).

Lemma 6.2.1 *For any pair of functions $P, K : \mathbb{N}_0 \to \mathbb{N}$, it holds that*

$$\lim_{n\to\infty} \mathbb{E}\left[I(n;\theta_n)\right] = \begin{cases} 0 & \text{if } \lim_{n\to\infty} \alpha_n = +\infty \\ \\ \infty & \text{if } \lim_{n\to\infty} \alpha_n = -\infty \end{cases} \tag{6.4}$$

where the function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ is determined through (2.14).

Lemma 6.2.1 will be established in Section 6.4.

Lemma 6.2.2 *For any pair of functions $P, K : \mathbb{N}_0 \to \mathbb{N}$, it holds that*

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[\chi_{n,1}(\theta_n)\chi_{n,2}(\theta_n)\right]}{\left(\mathbb{E}\left[\chi_{n,1}(\theta_n)\right]\right)^2} = 1 \tag{6.5}$$

whenever the function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ determined through (2.14) satisfies the condition

$$\lim_{n\to\infty} \alpha_n = -\infty. \tag{6.6}$$

A proof of Lemma 6.2.2 can be found in Section 6.5.

To complete the proof of Theorem 2.3.1, pick a pair of functions $P, K : \mathbb{N}_0 \to \mathbb{N}$ with function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ determined through (2.14) and recall the arguments of Section 5.1. Letting n go to infinity under the assumption $\lim_{n\to\infty} \alpha_n = \infty$, we get (5.8) (with Z_n, m_n and $\chi_{m_n,1}$ replaced by $I(n;\theta_n), n$ and $\chi_{n,1}(\theta_n)$, respectively) via Lemma

6.2.1, and the one-law $\lim_{n\to\infty} \mathbb{P}\left[I(n;\theta_n) = 0\right] = \lim_{n\to\infty} P^{\star}(n;\theta_n) = 1$ follows. Next, let n go to infinity under the condition (6.6): Lemma 6.2.1 already yields (5.9) with proper substitutions of the variables, while Lemma 6.2.2 leads (via (5.7)) to the appropriate version of (5.10). The conclusion $\lim_{n\to\infty} I(n;\theta_n) = \lim_{n\to\infty} P(n;\theta_n) = 0$ is now immediate by the arguments given in Chapter 5.1. This completes the proof of Theorem 2.3.1. ■

6.3 Some easy preliminaries

In this section we have collected for easy reference several technical facts that will be used repeatedly in the proofs of Lemma 6.2.1 and Lemma 6.2.2.

6.3.1 Simple consequences of the condition (6.6)

Pick a pair of functions $P, K : \mathbb{N}_0 \to \mathbb{N}$ with function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ determined through (2.14). Under condition (6.6), we have $\alpha_n < 0$ for n sufficiently large, in which case $\alpha_n = -|\alpha_n|$. By the non-negativity of $\frac{K_n^2}{P_n^2}$, hence of $\log n + \alpha_n$, we have $|\alpha_n| \leq \log n$ on that range and the inequality

$$\frac{K_n^2}{P_n} \leq \frac{\log n}{n} \tag{6.7}$$

follows, whence

$$\frac{K_n}{P_n} \leq \sqrt{\frac{1}{P_n} \cdot \frac{\log n}{n}} \leq \sqrt{\frac{\log n}{n}}. \tag{6.8}$$

As a result, we see that

$$\lim_{n\to\infty} \frac{K_n}{P_n} = 0 \tag{6.9}$$

and

$$\lim_{n\to\infty} \frac{K_n}{P_n - cK_n} = 0, \quad c \geq 0. \tag{6.10}$$

By virtue of (6.9), we note that for each $c > 0$, we have $P_n > cK_n$ for all n sufficiently large.

Next, for n sufficiently large, the first inequality in (6.8) gives

$$n\frac{K_n^3}{P_n^2} \leq \frac{n}{P_n^2} \cdot \left(\sqrt{\frac{\log n}{n}} P_n\right)^3 = n\left(\sqrt{\frac{\log n}{n}}\right)^3 \cdot \frac{\sqrt{P_n^3}}{P_n^2} \leq \sqrt{\frac{(\log n)^3}{n}},$$

and the conclusion

$$\lim_{n\to\infty} n\frac{K_n^3}{P_n^2} = 0 \tag{6.11}$$

is now immediate.

6.3.2 An easy technical fact

The next technical fact will help simplify the discussion in a number of places.

Lemma 6.3.1 *Consider a pair of functions $P, K : \mathbb{N}_0 \to \mathbb{N}$ such that the function*

$\alpha : \mathbb{N}_0 \to \mathbb{R}$ determined through (2.14) satisfies the condition (6.6). Then for any $c \geq 0$, we have

$$\lim_{n\to\infty} \left(1 - \frac{K_n}{P_n - cK_n}\right)^{K_n} = 1. \tag{6.12}$$

and

$$\lim_{n\to\infty} nK_n \Psi\left(\frac{K_n}{P_n - cK_n}\right) = 0 \tag{6.13}$$

where the function $\Psi(\cdot)$ is as defined in (5.16)-(5.17).

Proof. Fix $c \geq 0$ and recall (5.18) and (6.10). For each $n = 1, 2, \ldots$ sufficiently large, we can write

$$\begin{aligned} & K_n \Psi\left(\frac{K_n}{P_n - cK_n}\right) \\ &= K_n \left(\frac{K_n}{P_n - cK_n}\right)^2 \cdot \left(\frac{\Psi\left(\frac{K_n}{P_n - cK_n}\right)}{\left(\frac{K_n}{P_n - cK_n}\right)^2}\right) \\ &= K_n \left(\frac{K_n}{P_n - cK_n}\right)^2 \cdot \frac{1}{2}(1 + o(1)) \end{aligned}$$

with

$$\begin{aligned} \frac{K_n}{P_n - cK_n} &= \frac{K_n}{P_n} \cdot \left(1 - c\frac{K_n}{P_n}\right)^{-1} \\ &= \frac{K_n}{P_n}(1 + o(1)). \end{aligned}$$

Collecting these facts leads to

$$nK_n\Psi\left(\frac{K_n}{P_n - cK_n}\right) = \frac{n}{2}\frac{K_n^3}{P_n^2}(1 + o(1))$$

and (6.11) readily implies (6.13).

To establish (6.12), we need only show that

$$\lim_{n\to\infty} K_n \log\left(1 - \frac{K_n}{P_n - cK_n}\right) = 0. \tag{6.14}$$

For each $n = 2, 3, \ldots$, we rely on the decomposition (5.16) to write

$$\begin{aligned} & K_n \log\left(1 - \frac{K_n}{P_n - cK_n}\right) \\ &= -K_n\left(\frac{K_n}{P_n - cK_n} + \Psi\left(\frac{K_n}{P_n - cK_n}\right)\right) \\ &= -\frac{K_n^2}{P_n - cK_n} - K_n\Psi\left(\frac{K_n}{P_n - cK_n}\right). \end{aligned} \tag{6.15}$$

The arguments given in the first part of the proof also show that

$$K_n\Psi\left(\frac{K_n}{P_n - cK_n}\right) = \frac{K_n^3}{2P_n^2}(1 + o(1)) \tag{6.16}$$

and

$$\begin{aligned} \frac{K_n^2}{P_n - cK_n} &= \frac{K_n^2}{P_n}\left(1 - \frac{K_n}{P_n}\right)^{-1} \\ &= \frac{K_n^2}{P_n}(1 + o(1)). \end{aligned} \tag{6.17}$$

Let n go to infinity in (6.15): Making use of (6.16) and (6.17) we readily get (6.12) from the limits (6.7) and (6.11). ■

Finally, the bounds

$$\left(1-\frac{K}{P-K}\right)^K \leq \frac{\binom{P-K}{K}}{\binom{P}{K}} = q(\theta) \leq \left(1-\frac{K}{P}\right)^K \leq e^{-\frac{K^2}{P}} \tag{6.18}$$

are immediate upon applying Lemma 5.4.1 to the expression (2.3).

6.4 A proof of Lemma 6.2.1

Consider $\theta = (P, K)$ with positive integers K and P such that $2K < P$ and fix $n = 2, 3, \ldots$. Under the enforced independence assumptions, it is a simple matter to see that

$$\mathbb{E}\left[\chi_{n,i}(\theta)\right] = q(\theta)^{n-1}, \qquad i = 1, \ldots, , n \tag{6.19}$$

whence

$$\mathbb{E}\left[I(n;\theta)\right] = nq(\theta)^{n-1}. \tag{6.20}$$

Next, substitute in this expression θ by θ_n by means of an admissible pair of functions $P, K : \mathbb{N}_0 \to \mathbb{N}$. First we deal with the case $\lim_{n\to\infty} \alpha_n = \infty$: From (6.18) we obtain

$$nq(\theta_n)^{n-1} \leq e^{\alpha'_n}$$

for all $n = 1, 2, \ldots$ with

$$\begin{aligned} \alpha'_n &:= \log n - (n-1)\frac{K_n^2}{P_n} \\ &= \log n - \frac{(n-1)}{n} \cdot (\log n + \alpha_n) \\ &= \frac{\log n}{n} - \frac{n-1}{n}\alpha_n. \end{aligned}$$

We have $\lim_{n\to\infty} \alpha'_n = -\infty$ whenever $\lim_{n\to\infty} \alpha_n = \infty$, whence $\lim_{n\to\infty} nq(\theta_n)^{n-1} = 0$. The conclusion $\lim_{n\to\infty} \mathbb{E}\left[I(n;\theta_n)\right] = 0$ is now reached upon invoking (10.10). ■

In the case $\lim_{n\to\infty} \alpha_n = -\infty$ we note that the bounds (6.18) yield

$$n\left(1 - \frac{K_n}{P_n - K_n}\right)^{nK_n} \leq nq(\theta_n)^{n-1} \tag{6.21}$$

for all $n = 1, 2, \ldots$. We find it convenient to write the left handside of this last inequality as

$$n\left(1 - \frac{K_n}{P_n - K_n}\right)^{nK_n} = e^{\alpha''_n} \tag{6.22}$$

where

$$\begin{aligned} \alpha''_n &= \log n + nK_n \log\left(1 - \frac{K_n}{P_n - K_n}\right) \\ &= \log n - nK_n\left(\frac{K_n}{P_n - K_n} + \Psi\left(\frac{K_n}{P_n - K_n}\right)\right) \\ &= \log n - n\frac{K_n^2}{P_n - K_n} - nK_n\Psi\left(\frac{K_n}{P_n - K_n}\right) \end{aligned} \tag{6.23}$$

as we use the decomposition (5.16) in the second equality. The first two terms in (6.23) combine as

$$\begin{aligned}
&\log n - n\frac{K_n^2}{P_n - K_n} \\
&= \log n - n\frac{K_n^2}{P_n} \cdot \left(1 + \left(\frac{P_n}{P_n - K_n} - 1\right)\right) \\
&= \left(\log n - n\frac{K_n^2}{P_n}\right) - n\frac{K_n^2}{P_n} \cdot \frac{K_n}{P_n - K_n} \\
&= -\alpha_n - n\frac{K_n^2}{P_n} \cdot \frac{K_n}{P_n - K_n}.
\end{aligned}$$

Next, from (6.9) we observe that

$$\begin{aligned}
n\frac{K_n^2}{P_n}\frac{K_n}{P_n - K_n} &= n\frac{K_n^3}{P_n^2}\left(1 - \frac{K_n}{P_n}\right)^{-1} \\
&= n\frac{K_n^3}{P_n^2}\left(1 + o(1)\right).
\end{aligned}$$

Therefore, under the condition $\lim_{n\to\infty} \alpha_n = -\infty$, we get

$$\lim_{n\to\infty} n\frac{K_n^2}{P_n}\frac{K_n}{P_n - K_n} = 0$$

by virtue of (6.10) (with $c = 1$) and (6.11). We can then conclude to

$$\left(\log n - n\frac{K_n^2}{P_n - K_n}\right) = -\alpha_n(1 + o(1)), \tag{6.24}$$

and applying Lemma 6.3.1 (with $c = 1$) to the last term in (6.23) we find

$$\lim_{n\to\infty} nK_n\Psi\left(\frac{K_n}{P_n - K_n}\right) = 0. \tag{6.25}$$

Letting n go to infinity in (6.23), we conclude from (6.24) and (6.25) that $\lim_{n\to\infty}\alpha_n'' = -\lim_{n\to\infty}\alpha_n = \infty$ since $\lim_{n\to\infty}\alpha_n = -\infty$. It is now plain from (10.57) and (10.58) that $\lim_{n\to\infty}\mathbb{E}\left[I(n;\theta_n)\right] = \infty$ by virtue of (10.10). ■

6.5 A proof of Lemma 6.2.2

Consider $\theta = (P, K)$ with positive integers K and P such that $3K < P$, and write

$$b(\theta) := \frac{\binom{P-2K}{K}}{\binom{P}{K}}.$$

Fix $n = 2, 3, \ldots$. Under the enforced independence assumptions, it is a simple matter to check that

$$\mathbb{E}\left[\chi_{n,i}(\theta)\chi_{n,j}(\theta)\right] = q(\theta)b(\theta)^{n-2}$$

for distinct $i, j = 1, \ldots, n$, whence

$$\frac{\mathbb{E}\left[\chi_{n,1}(\theta)\chi_{n,2}(\theta)\right]}{\left(\mathbb{E}\left[\chi_{n,1}(\theta)\right]\right)^2} = \frac{q(\theta)b(\theta)^{n-2}}{q(\theta)^{2(n-1)}} = \frac{b(\theta)^{n-2}}{q(\theta)^{2n-3}}. \tag{6.26}$$

On the way to evaluating this ratio, we note that

$$\begin{aligned}\frac{b(\theta)^{n-2}}{q(\theta)^{2n-3}} &= \left(\frac{\binom{P-2K}{K}}{\binom{P}{K}}\right)^{n-2} \cdot \left(\frac{\binom{P}{K}}{\binom{P-K}{K}}\right)^{2n-3} \\ &= \left(\frac{\binom{P-2K}{K}}{\binom{P-K}{K}}\right)^{n-2} \cdot \left(\frac{\binom{P}{K}}{\binom{P-K}{K}}\right)^{n-1} \\ &= \frac{r(\theta)^{n-2}}{q(\theta)} \end{aligned} \tag{6.27}$$

where we have used the notation

$$r(\theta) := \frac{\binom{P-2K}{K}}{\binom{P-K}{K}} \cdot \frac{\binom{P}{K}}{\binom{P-K}{K}}.$$

Under the condition (6.6) we show below that

$$\lim_{n\to\infty} q(\theta_n) = 1 \tag{6.28}$$

and

$$\lim_{n\to\infty} r(\theta_n)^{n-2} = 1. \tag{6.29}$$

Once this is done, it is plain from (6.27) that

$$\lim_{n\to\infty} \frac{b(\theta_n)^{n-2}}{q(\theta_n)^{2n-3}} = 1 \tag{6.30}$$

and the desired result (6.5) follows from (6.26).

In order to establish (6.28) and (6.29) we proceed as in the proof of Lemma 6.2.1:

First, making use of (6.18) we obtain

$$\left(1-\frac{K_n}{P_n-K_n}\right)^{K_n} \leq q(\theta_n) \leq \left(1-\frac{K_n}{P_n}\right)^{K_n} \tag{6.31}$$

for all $n = 2, 3, \ldots$. Letting n go to infinity and using (6.12) (with $c = 0$ and $c = 1$) we get (6.28).

Next, using (6.18), this time with P_n replaced by $P_n - K_n$, we get

$$\left(1-\frac{K_n}{P_n-2K_n}\right)^{K_n} \leq \frac{\binom{P_n-2K_n}{K_n}}{\binom{P_n-K_n}{K_n}} \leq \left(1-\frac{K_n}{P_n-K_n}\right)^{K_n} \tag{6.32}$$

for all $n = 2, 3, \ldots$. Combining (6.31) and (6.32) readily gives

$$\left(\frac{1-\frac{K_n}{P_n-2K_n}}{1-\frac{K_n}{P_n}}\right)^{K_n} \leq r(\theta_n) \leq 1. \tag{6.33}$$

It is now plain that the convergence (6.29) will be established if we can show that

$$\lim_{n\to\infty}(n-2)K_n \log\left(\frac{1-\frac{K_n}{P_n-2K_n}}{1-\frac{K_n}{P_n}}\right) = 0. \tag{6.34}$$

To that end, for each $n = 2, 3, \ldots$ we note that

$$\begin{aligned}
&(n-2)K_n \cdot \log\left(\frac{1-\frac{K_n}{P_n-2K_n}}{1-\frac{K_n}{P_n}}\right) \\
&= -(n-2)K_n \cdot \left(\frac{K_n}{P_n-2K_n} + \Psi\left(\frac{K_n}{P_n-2K_n}\right)\right) \\
&\quad + (n-2)K_n \cdot \left(\frac{K_n}{P_n} + \Psi\left(\frac{K_n}{P_n}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{2(n-2)K_n^3}{P_n(P_n-2K_n)} \\
&\quad -(n-2)K_n\cdot\left(\Psi\left(\frac{K_n}{P_n-2K_n}\right)-\Psi\left(\frac{K_n}{P_n}\right)\right) \\
&= -\frac{2(n-2)K_n^3}{P_n^2}\cdot\left(1-\frac{2K_n}{P_n}\right)^{-1} \\
&\quad -(n-2)K_n\cdot\left(\Psi\left(\frac{K_n}{P_n-2K_n}\right)-\Psi\left(\frac{K_n}{P_n}\right)\right).
\end{aligned}
$$

Applying Lemma 6.3.1 (with $c=0$ and $c=2$) yields

$$
\lim_{n\to\infty}(n-2)K_n\cdot\left(\Psi\left(\frac{K_n}{P_n-2K_n}\right)-\Psi\left(\frac{K_n}{P_n}\right)\right)=0,
$$

while (6.9) and (6.11) together lead to

$$
\lim_{n\to\infty}\frac{(n-2)K_n^3}{P_n^2}\left(1-\frac{2K_n}{P_n}\right)^{-1}=0.
$$

The convergence (6.34) now follows and Lemma 6.2.2 is now established. ■

Chapter 7

Connectivity in random key graphs II: Graph connectivity

7.1 Introduction

As discussed in Chapter 2.3.1, the study of the connectivity properties of random key graphs resulted in the conjectures (2.11) and (2.12) (under (2.14)). In Chapter 6, we established the conjectured zero-one law (2.11) for the absence of isolated nodes. In this chapter, we report on the results regarding the conjectured zero-one law (2.12) for graph connectivity under (2.14). We will establish Theorem 2.3.2 which is restated here for the ease of exposition.

Theorem 7.1.1 *For any admissible scaling $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ such that*

$$\frac{K_n^2}{P_n} = \frac{\log n + \alpha_n}{n}, \quad n = 1, 2, \ldots \tag{7.1}$$

for some sequence $\alpha : \mathbb{N}_0 \to \mathbb{R}$*, it follows that*

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{K}(n;\theta_n) \text{ is connected}\right] = \begin{cases} 0 & \textit{if } \lim_{n\to\infty} \alpha_n = -\infty \\ \\ 1 & \textit{if } \lim_{n\to\infty} \alpha_n = +\infty \end{cases} \tag{7.2}$$

whenever there exists some $\sigma > 0$ *such that*

$$\sigma n \leq P_n \tag{7.3}$$

for all $n = 1, 2, \ldots$ *sufficiently large.*

7.1.1 Related work

Recent results concerning the conjectured zero-one law (7.1)-(7.2) are now surveyed: Di Pietro et al. have shown [10, Thm. 4.6] that for large n, the random key graph will be connected with very high probability if P_n and K_n are selected such that

$$K_n \geq 5, \ P_n \geq n \quad \text{and} \quad \frac{K_n^2}{P_n} \sim c \, \frac{\log n}{n} \tag{7.4}$$

as soon as $c \geq 16$.[1] They also observe that for large n, the random key graph will be disconnected with very high probability if the scaling satisfies

$$\frac{K_n^2}{P_n} = o\left(\frac{\log n}{n}\right).$$

[1] In the conference version of this work [9, Thm. 4.6] the result is claimed to hold for $c > 8$.

As mentioned earlier, the zero-law in (7.2) follows from the zero-law in Theorem 2.3.1 which was established independently by Blackburn and Gerke [3], and by Yağan and Makowski [35]. In both papers, it was shown that

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{K}(n;(K_n,P_n)) \text{ contains no isolated nodes}\right] = 0$$

whenever $\lim_{n\to\infty} \alpha_n = -\infty$ in (7.1), a result which clearly implies the conjectured zero-law.

Blackburn and Gerke [3] also succeeded in generalizing the one-law result by Di Pietro et al. in a number of directions: Under the additional conditions

$$K_n \geq 2 \quad \text{and} \quad P_n \geq n, \quad n = 1, 2, \ldots, \tag{7.5}$$

they showed [3, Thm. 5] that

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{K}(n;(K_n,P_n)) \text{ is connected}\right] = 1 \quad \text{if} \quad \liminf_{n\to\infty} \frac{K_n^2}{P_n}\frac{n}{\log n} > 1. \tag{7.6}$$

This result is weaker than the one-law in the conjecture (7.1)-(7.2). However, in the process of establishing (7.6), they also show [3, Thm. 3] that the conjecture does hold in the special case $K_n = 2$ for all $n = 1, 2, \ldots$ *without* any constraint on the size of the key pools, say $P_n \leq n$ or $n \leq P_n$. Specifically, the one-law in (7.2) is shown to

hold whenever the scaling is done according to

$$K_n = 2, \quad \frac{4}{P_n} = \frac{\log n + \alpha_n}{n}, \quad n = 1, 2, \ldots \tag{7.7}$$

with $\lim_{n\to\infty} \alpha_n = \infty$. For each $n = 2, 3, \ldots$, a simple coupling argument yields

$$P(n; (2, P)) \leq P(n; (K, P)) \tag{7.8}$$

whenever $2 \leq K \leq P$. Thus, it is now easy to conclude that the one-law in (7.2) holds whenever $2 \leq K_n \leq P_n$ and $P_n = o\left(\frac{n}{\log n}\right)$; this corresponds to requiring $P_n \ll n$.

7.1.2 Contributions

We complement existing results concerning the conjecture (7.1)-(7.2) in several ways: We establish the conjecture (7.1)-(7.2) whenever $K_n \geq 2$ and $P_n \geq \sigma n$ for some $\sigma > 0$. The condition (7.3) is sometimes expressed as $P_n = \Omega(n)$ and is weaker than the growth condition at (7.5) used by Blackburn and Gerke [3]. It is also easy to check that Theorem 7.1.1 implies the one-law (7.6). Therefore, Theorem 7.1.1 already improves on the one-law (7.6) obtained by Blackburn and Gerke [3] under the condition (7.5). Moreover, as discussed earlier, these authors have established the conjectured one-law in (7.2) under conditions on the scalings very different from the ones used here, i..e., either $K_n = 2$ or $K_n \geq 2$ with $P_n = o\left(\frac{n}{\log n}\right)$. In practical WSN scenarios it is expected that the size of the key pool will be large compared to the

number of participating nodes [10, 14] and that key rings will contain more than two keys. In this context, our results concerning the full conjecture (7.1)-(7.2) are given under more realistic conditions than in earlier work.

Finally, the one-law in Theorem 7.1.1 cannot hold if the condition (2.18) fails. This is a simple consequence of the following observation.

Lemma 7.1.2 *For any mapping $P : \mathbb{N}_0 \to \mathbb{N}_0$ for which the limit $\lim_{n\to\infty} P_n$ exists (possibly infinite), we have*

$$\lim_{n\to\infty} P(n;(1,P_n)) = \begin{cases} 0 & \textit{if } \lim_{n\to\infty} P_n > 1 \\ \\ 1 & \textit{if } \lim_{n\to\infty} P_n = 1. \end{cases} \tag{7.9}$$

Proof. For $n = 2, 3, \ldots$ and any positive integer P_n, the graph $\mathbb{K}(n;(1,P_n))$ is connected if and only if all nodes choose the *same* key, an event which happens with probability $P_n^{-(n-1)}$. The conclusion is now immediate once we observe that the condition $\lim_{n\to\infty} P_n = 1$ (resp. $\lim_{n\to\infty} P_n > 1$) requires $P_n = 1$ (resp. $P_n \geq 2$) for all $n = 1, 2, \ldots$ sufficiently large owing to P_n being integer. ■

The proof of the main result is lengthy and technically involved. However, in

a parallel development, we have also shown in [38] that when $P_n = O(n^\delta)$ with $0 < \delta < \frac{1}{2}$, the so-called small key pool case, elementary arguments can be used to establish a one-law for connectivity. This is an easy byproduct of the observation that connectivity is achieved in the random key graph whenever *all* possible key rings have been distributed to the participating nodes; see Section 7.2 for details.

7.1.3 The structure

The chapter is organized as follows: Section 7.2 is devoted to establishing the conjecture (7.1)-(7.2) under special cases not covered by Theorem 7.1.1. In Section 7.3 we give a roadmap to the proof of Theorem 7.1.1. The approach is similar to the one used for proving the one-law for graph connectivity in Erdős-Rényi graphs [4, p. 164] [32, p. 304]; see (2.7)-(2.9). Here as well, we focus on the probability that the random key graph is not connected and yet has no isolated nodes. We then seek to show that this probability becomes vanishingly small as n grows large under the appropriate scaling. As in the classical case this is achieved through a combination of judicious bounding arguments, the starting point being the well-known bound (5.14) on the probability of interest. However, in order for these arguments to successfully go through, we found it necessary to restrict attention to a subclass of structured scalings (referred throughout as strongly admissible scalings). In Section 7.4.1 a reduction argument shows that we need only establish the desired one-law for such strongly admissible scalings. The explanation of the right handside of (7.1) as a proxy for link assignment in the limiting regime is revealed through a useful equivalence developed

in Section 7.4.2.

With these technical prerequisites in place, the needed bounding arguments are then developed in Section 7.4.3, Section 7.4.4 and Section 7.4.5, and the final steps of the proof of Theorem 7.1.1 are outlined in Section 7.4.6. The final sections of the chapter, namely Section 7.4.7 through Section 7.4.12, are devoted to the various technical steps needed to complete the arguments outlined in Section 7.4.6.

7.2 Simple proofs for special cases

As will become apparent soon, the proof of Theorem 7.1.1 is rather long and technically involved. In this section, we discuss a number of situations for which the conjectured one-law can be easily recovered when the key pool P_n is "small" compared to n. The basic idea behind these shorter proofs is the following simple observation: The key graph *is* automatically connected if *all* possible key rings have been distributed to the nodes.

The arguments which we develop on the basis of this fact clearly indicate the interplay that exists in random key graphs between the size of the key pool and the number of nodes; this is reflected in the additional assumptions under which the results were given in this dissertation.

7.2.1 A basic observation

Assume given a pair of positive integers K and P such that $K \leq P$, and pick $n = 2, 3, \ldots$. We define the events

$$C_n(\theta) := [\mathbb{K}(n;\theta) \text{ is connected}]$$

and

$$Q_n(\theta) := \left[\begin{array}{c} \text{All key rings of size } K \\ \\ \text{have been distributed} \end{array} \right].$$

The event $Q_n(\theta)$ is always empty under the condition

$$n < \binom{P}{K}. \tag{7.10}$$

The next observation provides an easy condition for graph connectivity in random key graphs.

Lemma 7.2.1 *For any given pair $\theta = (K, P)$ with $2 \leq K \leq P$, it is always the case that $\mathbb{K}(n;\theta)$ is connected whenever all the key rings of size K have been distributed, i.e.,*

$$Q_n(\theta) \subseteq C_n(\theta). \tag{7.11}$$

Proof. Fix $2 \leq K \leq P$ and let ω be a sample that belongs to the event $Q_n(\theta)$. Pick two distinct nodes, say $i, j = 1, \ldots, n$. We need to show that there is path between them in $\mathbb{K}(n;\theta)(\omega)$. If the key rings $\Sigma_i(\theta)(\omega)$ and $\Sigma_j(\theta)(\omega)$ have a non-empty intersection, then the two nodes are adjacent and there is a one hop path between them. On the other hand, if these key rings do not intersect, then it is necessarily the case that $2K \leq P$. Under these conditions it is possible to construct an element S of $\mathcal{P}_K$ such that $S \cap \Sigma_i(\theta)(\omega) \neq \emptyset$ and $S \cap \Sigma_j(\theta)(\omega) \neq \emptyset$. Note that such an argument could not be made for the case $K = 1$. Since all the key rings have been distributed in $\mathbb{K}(n;\theta)(\omega)$ it follows that there exists a node, say ℓ (possibly dependent on ω), distinct from both i and j, such that $K_\ell(\theta)(\omega) = S$. As a result, nodes i and j are connected by a two-hop path passing through ℓ. ■

By virtue of Lemma 7.2.1, it is now natural to look for conditions under which the event $Q_n(\theta)$ occurs with high probability. For this purpose we first consider its complement which corresponds to the event that *some* key ring of size K has *not* been distributed, namely

$$Q_n(\theta)^c = \cup_{S \in \mathcal{P}_K} \left[\Sigma_1(\theta) \neq S, \ldots, \Sigma_n(\theta) \neq S\right].$$

By a union bound argument, we get

$$\begin{aligned}
\mathbb{P}\left[Q_n(\theta)^c\right] &\leq \sum_{S\in\mathcal{P}_K} \mathbb{P}\left[\Sigma_1(\theta)\neq S,\ldots,\Sigma_n(\theta)\neq S\right] \\
&= \sum_{S\in\mathcal{P}_K}\left(\prod_{i=1}^{n}\mathbb{P}\left[\Sigma_i\neq S\right]\right) \\
&= \sum_{S\in\mathcal{P}_K}\mathbb{P}\left[\Sigma_1\neq S\right]^n \\
&= \binom{P}{K}\left(1-\frac{1}{\binom{P}{K}}\right)^n \qquad (7.12)
\end{aligned}$$

under the enforced probabilistic assumptions on key ring selection.

7.2.2 An easy one-law

Lemma 7.2.1 and the calculations following it suggest a very simple strategy to obtain versions of the one-law in random key graphs. Consider an admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ such that

$$\binom{P_n}{K_n} \leq n \qquad (7.13)$$

for all $n = 1, 2, \ldots$ sufficiently large. On that range, it follows from (7.12) that

$$\begin{aligned}
\mathbb{P}\left[Q_n(\theta_n)^c\right] &\leq \binom{P_n}{K_n}\left(1-\frac{1}{\binom{P_n}{K_n}}\right)^n \\
&\leq \binom{P_n}{K_n}e^{-\frac{n}{\binom{P_n}{K_n}}} \qquad (7.14)
\end{aligned}$$

by standard bounding arguments. The conclusion

$$\lim_{n\to\infty} \mathbb{P}\left[Q_n(\theta_n)^c\right] = 0 \tag{7.15}$$

then follows *provided* the condition

$$\lim_{n\to\infty} \binom{P_n}{K_n} e^{-\frac{n}{\binom{P_n}{K_n}}} = 0 \tag{7.16}$$

holds under (7.13). This observation readily leads to the following one-law.

Lemma 7.2.2 *Consider an admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ such that (7.13) holds for all $n = 1, 2, \ldots$ sufficiently large. We have $\lim_{n\to\infty} P(n;\theta_n) = 1$ provided the condition (7.16) holds*

In the next three sections we use Lemma 7.2.2 to derive several one-laws under specific sets of assumptions.

7.2.3 Fixed values of K and P

The next result has a well-known analog for Erdős-Renyi graphs.

Lemma 7.2.3 *For any given pair $\theta = (K, P)$ with $2 \le K \le P$, we have $\lim_{n\to\infty} P(n;\theta) = 1$.*

The pair $\theta = (K, P)$ with $2 \leq K \leq P$ corresponds to a scaling whose deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ is given by

$$\alpha_n := n\frac{K^2}{P} - \log n, \quad n = 1, 2, \ldots$$

so that $\lim_{n\to\infty} \alpha_n = \infty$. However, the conclusion of Lemma 7.2.3 does not follow either from the result (7.6) by Blackburn and Gerke [3] or from Theorem 7.1.1 since conditions (7.3) and (7.5) are not satisfied with $P_n \equiv P$. The result is nevertheless a consequence of [3, Thm. 3] since condition (7.7) holds with $\lim_{n\to\infty} \alpha_n = \infty$.

Proof. It follows from (7.12) that

$$\mathbb{P}\left[Q_n(\theta)\right] \geq 1 - \binom{P}{K}\left(1 - \frac{1}{\binom{P}{K}}\right)^n$$

for all $n = 1, 2, \ldots$ sufficiently large to ensure $\binom{P}{K} \leq n$. The conclusion $\lim_{n\to\infty} \mathbb{P}\left[Q_n(\theta)\right] = 1$ is now immediate and we get the result by making use of the inclusion (7.11). ■

7.2.4 The case $\limsup_{n\to\infty} P_n < \infty$

Lemma 7.2.3 leads to a proof of the conjectured one-law for scalings $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ satisfying the property

$$\bar{P} := \limsup_{n\to\infty} P_n = \inf_{n\geq 1} \left(\sup_{m\geq n} P_m \right) < \infty. \tag{7.17}$$

Lemma 7.2.4 *For any admissible scaling $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ satisfying (7.17). we have* $\lim_{n\to\infty} P(n, \theta_n) = 1$.

Proof. Under the finiteness condition (7.17) we have

$$\limsup_{n\to\infty} \binom{P_n}{K_n} < \infty.$$

Hence, both conditions (7.13) and (7.16) hold, and the result follows from Lemma 7.2.2. ■

7.2.5 Small key pools with $K_n = 2$

With $K_n = 2$, we note that

$$\binom{P_n}{2} = \frac{P_n(P_n - 1)}{2} \leq P_n^2, \quad n = 1, 2, \ldots.$$

Therefore, the condition (7.13) holds whenever

$$P_n^2 \leq n \tag{7.18}$$

for all $n = 1, 2, \ldots$ sufficiently large. Since the mapping $t \to te^{-\frac{n}{t}}$ is increasing on $(0, \infty)$, the convergence condition (7.16) is implied whenever

$$\lim_{n\to\infty} P_n^2 e^{-\frac{n}{P_n^2}} = 0. \tag{7.19}$$

This observation leads to the following one-law.

Lemma 7.2.5 *Consider an admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ satisfying (7.7) such that*

$$P_n = O(n^\delta) \tag{7.20}$$

for some δ in $(0, \frac{1}{2})$. Then we have

$$\lim_{n\to\infty} P(n; (2, P_n)) = 1. \tag{7.21}$$

This is of course a weaker version of Theorem [3, Thm. 3] (which establishes (7.21) whenever $P_n = o\left(\frac{n}{\log n}\right)$) but as the discussion below shows, its proof is much simpler and comes with the additional benefit of pointing out the underlying reason for connectivity when P_n is *much smaller* than n – In that case it is very likely that

all the possible key rings are eventually assigned!

Proof. Condition (7.20) implies the existence of a constant $C > 0$ such that

$$P_n \leq Cn^{\delta}, \quad n \geq n^{\star}$$

for some finite integer $n^{\star}$. Therefore, (7.18) is automatically satisfied for δ in the prescribed range and condition (7.13) holds.

Next, by the aforementioned monotonicity, we also get

$$P_n^2 e^{-\frac{n}{P_n^2}} \leq C^2 n^{2\delta} e^{-C^{-2} n^{1-2\delta}}, \quad n \geq n^{\star}$$

and the convergence (7.19) follows since we have $2\delta < 1$ here. The desired conclusion is now immediate by Lemma 7.2.2. ■

As before, it follows from (7.8) and Lemma 7.2.5 that $\lim_{n\to\infty} P(n;\theta_n) = 1$ whenever

$$2 \leq K_n \leq P_n$$

for all $n = 1, 2, \ldots$ sufficiently large under the condition (7.20) with δ in $(0, \frac{1}{2})$.

7.3 A roadmap for the proof of Theorem 7.1.1

Fix $n = 2, 3, \ldots$ and consider positive integers K and P such that $2 \leq K \leq P$. As before, we define the events

$$C_n(\theta) := [\mathbb{K}(n;\theta) \text{ is connected}]$$

and

$$I_n(\theta) := [\mathbb{K}(n;\theta) \text{ contains no isolated nodes}]\,.$$

If the random key graph $\mathbb{K}(n;\theta)$ is connected, then it does not contain isolated nodes, whence $C_n(\theta)$ is a subset of $I_n(\theta)$, and the conclusions

$$\mathbb{P}\left[C_n(\theta)\right] \leq \mathbb{P}\left[I_n(\theta)\right] \tag{7.22}$$

and

$$\mathbb{P}\left[C_n(\theta)^c\right] = \mathbb{P}\left[C_n(\theta)^c \cap I_n(\theta)\right] + \mathbb{P}\left[I_n(\theta)^c\right] \tag{7.23}$$

obtain.

Taken together with Theorem 2.3.1, the relations (7.22) and (10.42) pave the way to proving Theorem 7.1.1. Indeed, pick an admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ with deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$. If $\lim_{n\to\infty} \alpha_n = -\infty$, then $\lim_{n\to\infty} \mathbb{P}\left[I_n(\theta_n)\right] = 0$ by the zero-law in Theorem 2.3.1, whence $\lim_{n\to\infty} \mathbb{P}\left[C_n(\theta_n)\right] = 0$ with the help of (7.22). If $\lim_{n\to\infty} \alpha_n = \infty$, then $\lim_{n\to\infty} \mathbb{P}\left[I_n(\theta_n)\right] = 1$ by the one-law in Theorem 2.3.1, and

the desired conclusion $\lim_{n\to\infty} \mathbb{P}\left[C_n(\theta_n)\right] = 1$ (or equivalently, $\lim_{n\to\infty} \mathbb{P}\left[C_n(\theta_n)^c\right] = 0$) will follow via (10.42) if we show that

$$\lim_{n\to\infty} \mathbb{P}\left[C_n(\theta_n)^c \cap I_n(\theta_n)\right] = 0. \tag{7.24}$$

We shall do this by finding a sufficiently tight upper bound on the probability in (7.24) and then showing that it goes to zero as well. While the additional condition (7.3) plays a crucial role in carrying out this argument, a number of additional assumptions will be imposed on the admissible scaling under consideration. This is done mostly for technical reasons in that it leads to simpler proofs. Eventually these additional conditions will be removed to ensure the desired final result, namely $\lim_{n\to\infty} \mathbb{P}\left[C_n(\theta_n)\right] = 1$ under (7.3), e.g., see Section 7.4.1 for details.

With this in mind, the admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ is said to be *strongly admissible* if its deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies the additional growth condition

$$\alpha_n = o(n). \tag{7.25}$$

Strong admissibility has the following useful implications: Under (7.25) it is always the case from (7.1) that

$$\lim_{n\to\infty} \frac{K_n^2}{P_n} = 0. \tag{7.26}$$

Since $1 \leq K_n \leq K_n^2$ for all $n = 1, 2, \ldots$, this last convergence implies

$$\lim_{n\to\infty} \frac{K_n}{P_n} = 0 \tag{7.27}$$

and

$$\lim_{n\to\infty} P_n = \infty. \tag{7.28}$$

As a result,

$$2K_n \leq P_n \tag{7.29}$$

for all $n = 1, 2, \ldots$ sufficiently large, and the random key graph does not degenerate into a complete graph under a strongly admissible scaling. We shall also make use of the fact that (7.27) is equivalent to

$$\lim_{n\to\infty} \frac{P_n}{K_n} = \infty. \tag{7.30}$$

Finally in Lemma 7.4.4 we show that (7.26) suffices to imply

$$1 - q(\theta_n) \sim \frac{K_n^2}{P_n}. \tag{7.31}$$

This is discussed in Section 7.4.2, and provides the appropriate version of (2.13).

7.4 A proof of Theorem 7.1.1

7.4.1 A reduction step

The relevance of the notion of strong admissibility flows from the following fact.

Lemma 7.4.1 *Consider an admissible scaling $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ whose deviation sequence $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies*

$$\lim_{n\to\infty} \alpha_n = \infty. \tag{7.32}$$

Assume there exists some $\sigma > 0$ such that (7.3) holds for all $n = 1, 2, \ldots$ sufficiently large. Then, there always exists an admissible scaling $\tilde{K}, \tilde{P} : \mathbb{N}_0 \to \mathbb{N}_0$ with

$$\tilde{K}_n \leq K_n \quad \text{and} \quad \tilde{P}_n = P_n, \quad n = 1, 2, \ldots \tag{7.33}$$

whose deviation function $\tilde{\alpha} : \mathbb{N}_0 \to \mathbb{R}$ satisfies both conditions

$$\lim_{n\to\infty} \tilde{\alpha}_n = \infty \tag{7.34}$$

and

$$\tilde{\alpha}_n = o(n). \tag{7.35}$$

In other words, the scaling $\tilde{K}, \tilde{P} : \mathbb{N}_0 \to \mathbb{N}_0$ defined at (7.33) is strongly admissible and still satisfies the condition (7.3).

Proof. For each $n = 1, 2, \ldots$, we set

$$K_n^\star := \sqrt{P_n \cdot \frac{\log n + \alpha_n^\star}{n}}.$$

where

$$\alpha_n^\star := \min(\alpha_n, \log n)$$

The properties

$$\lim_{n\to\infty} \alpha_n^\star = \infty \tag{7.36}$$

and

$$\alpha_n^\star = o(n) \tag{7.37}$$

are immediate by construction.

Now define the scaling $\tilde{K}, \tilde{P} : \mathbb{N}_0 \to \mathbb{N}_0$ by

$$\tilde{K}_n := \lceil K_n^\star \rceil, \ \tilde{P}_n = P_n, \quad n = 1, 2, \ldots \tag{7.38}$$

We get $K_n^\star \leq K_n$ for all $n = 1, 2, \ldots$ since $\alpha_n^\star \leq \alpha_n$, whence $\tilde{K}_n \leq K_n$ by virtue of the fact that K_n is always an integer. This establishes (7.33).

Next, observe that $\tilde{K}_n = 1$ if and only $K_n^\star \leq 1$, a condition which occurs only when

$$P_n \left(\log n + \alpha_n^\star\right) \leq n. \tag{7.39}$$

This last inequality can only hold for a finite number of values of n. Otherwise, there would exist a countably infinite subset N of $\mathbb{N}_0$ such that both (7.3) and (7.39) simultaneously hold on N. In that case, we conclude that

$$\sigma \left(\log n + \alpha_n^\star\right) \leq 1, \quad n \in N$$

and this is a clear impossibility in view of (7.36). Together with (7.33) this establishes the admissibility of the scaling $\tilde{K}, \tilde{P} : \mathbb{N}_0 \to \mathbb{N}_0$.

Fix $n = 1, 2, \ldots$. The definitions imply $K_n^\star \leq \tilde{K}_n < 1 + K_n^\star$ and upon squaring we get the inequalities

$$P_n \cdot \frac{\log n + \alpha_n^\star}{n} \leq \tilde{K}_n^2 \tag{7.40}$$

and

$$\tilde{K}_n^2 < 1 + 2\sqrt{P_n \cdot \frac{\log n + \alpha_n^\star}{n}} + P_n \cdot \frac{\log n + \alpha_n^\star}{n}. \tag{7.41}$$

The deviation sequence $\tilde{\alpha} : \mathbb{N}_0 \to \mathbb{R}$ of the newly defined scaling (7.33) is determined through

$$\frac{\tilde{K}_n^2}{\tilde{P}_n} = \frac{\log n + \tilde{\alpha}_n}{n}, \quad n = 1, 2, \ldots.$$

Using (7.40) and (7.41) we then conclude that

$$\alpha_n^\star \leq \tilde{\alpha}_n \tag{7.42}$$

and

$$\frac{\tilde{\alpha}_n}{n} < \frac{1}{P_n} + 2\sqrt{\frac{1}{P_n} \cdot \frac{\log n + \alpha_n^\star}{n}} + \frac{\alpha_n^\star}{n}. \tag{7.43}$$

It is now plain from (7.36) and (7.42) that (7.34) holds. Next, by combining (7.42) and (7.43) we get

$$\frac{\alpha_n^\star}{n} \leq \frac{\tilde{\alpha}_n}{n} < \frac{1}{P_n} + 2\sqrt{\frac{1}{P_n} \cdot \frac{\log n + \alpha_n^\star}{n}} + \frac{\alpha_n^\star}{n}. \tag{7.44}$$

Letting n go to infinity in (7.44) and using (7.37) we conclude to (7.35) since $\lim_{n\to\infty} P_n = \infty$ by virtue of (7.3). ■

This construction also works with

$$\alpha_n^\star = \min\left(\alpha_n, \omega_n\right), \quad n = 1, 2, \ldots$$

for any sequence $\omega : \mathbb{N}_0 \to \mathbb{R}_+$ such that $\lim_{n\to\infty} \omega_n = \infty$ and $\omega_n = o(n)$, e.g., $\omega_n = n^\delta$ for some $0 < \delta < 1$.

We close with a key technical consequence of Lemma 7.4.1: By construction the scaling $\tilde{K}, \tilde{P} : \mathbb{N}_0 \to \mathbb{N}_0$ *is* a strongly admissible scaling and an easy coupling argument

based on (7.33) implies

$$P(n;\tilde{\theta}_n) \leq P(n;\theta_n), \quad n = 2, 3, \ldots$$

Thus, we need only show the one law in Theorem 7.1.1 for strongly admissible scalings. As a result, in view of the discussion leading to (7.24) it suffices to establish the following result, to which the remainder of the chapter is devoted.

Proposition 7.4.2 *Consider any strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ whose deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies $\lim_{n\to\infty} \alpha_n = \infty$. Under the condition (7.3), we have*

$$\lim_{n\to\infty} \mathbb{P}\left[C_n(\theta_n)^c \cap I_n(\theta_n)\right] = 0. \tag{7.45}$$

Proposition 7.4.2 shows that in random key graphs, graph connectivity is asymptotically equivalent to the absence of isolated nodes under any strongly admissible scaling whose deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies $\lim_{n\to\infty} \alpha_n = \infty$ under the condition (7.3).

7.4.2 The equivalence (7.31)

To establish the key equivalence (7.31) we start by applying Lemma 5.4.1 to the expression (2.3). This yields the following bounds:

Lemma 7.4.3 *With positive integers K and P such that $2K \leq P$, we have*

$$1 - e^{-\frac{K^2}{P}} \leq 1 - q(\theta) \leq \frac{K^2}{P-K}. \tag{7.46}$$

Proof. Lemma 5.4.1 (with $L = K$) yields the bounds

$$1 - e^{-\frac{K^2}{P}} \leq 1 - q(\theta) \leq 1 - \left(1 - \frac{K}{P-K}\right)^K. \tag{7.47}$$

The conclusion (7.46) is now immediate once we note that

$$1 - \left(1 - \frac{K}{P-K}\right)^K = \int_{1-\frac{K}{P-K}}^1 K t^{K-1} dt \leq \frac{K^2}{P-K}$$

by a crude bounding argument. ■

A little bit more than (7.31) can be said.

Lemma 7.4.4 *For any scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$, we have*

$$\lim_{n\to\infty} q(\theta_n) = 1 \tag{7.48}$$

if and only if

$$\lim_{n\to\infty} \frac{K_n^2}{P_n} = 0, \tag{7.49}$$

and under either condition the asymptotic equivalence

$$1 - q(\theta_n) \sim \frac{K_n^2}{P_n} \tag{7.50}$$

holds.

On several occasions, we will rely on (7.50) through the following equivalent formulation: For every δ in $(0,1)$ there exists a finite integer $n^\star(\delta)$ such that

$$(1-\delta)\frac{K_n^2}{P_n} \le 1 - q(\theta_n) \le (1+\delta)\frac{K_n^2}{P_n} \tag{7.51}$$

whenever $n \ge n^\star(\delta)$.

Proof. As noted already at the end of Section 7.3, condition (7.49) (which holds for any strongly admissible scaling) implies

$$2K_n \le P_n \tag{7.52}$$

for all n sufficiently large. On that range Lemma 7.4.3 yields

$$1 - e^{-\frac{K_n^2}{P_n}} \le 1 - q(\theta_n) \le \frac{K_n^2}{P_n - K_n}. \tag{7.53}$$

Multiply (7.53) by $\frac{P_n}{K_n^2}$ and let n go to infinity in the resulting set of inequalities. Under (7.49), we get

$$\lim_{n\to\infty} \frac{P_n}{K_n^2} \cdot \left(1 - e^{-\frac{K_n^2}{P_n}}\right) = 1 \tag{7.54}$$

from the elementary fact $\lim_{t\downarrow 0} \frac{1-e^{-t}}{t} = 1$, while

$$\lim_{n\to\infty} \frac{P_n}{K_n^2} \cdot \frac{K_n^2}{P_n - K_n} = \lim_{n\to\infty} \frac{P_n}{P_n - K_n} = 1 \tag{7.55}$$

by virtue of (7.27) (which is implied by (7.49)). The asymptotic equivalence (7.50) follows, and the validity of (7.48) is immediate.

Conversely, under the condition $\lim_{n\to\infty} q(\theta_n) = 1$, we have $0 < q(\theta_n) < 1$ for all n sufficiently large (by the comment following (2.4)), and the constraint (7.52) necessarily holds. On that range, (7.53) being valid, we conclude to $\lim_{n\to\infty} e^{-\frac{K_n^2}{P_n}} = 1$ under (7.48). The convergence (7.49) now follows and the asymptotic equivalence (7.50) is given by the first part of the proof. ■

7.4.3 The union bound (5.14)

Proposition 7.4.2 will be established with the help of a bound for the probability appearing at (7.45). Indeed, an appropriate version of the union bound (5.14) (established in Chapter 5.2) will suffice: Fix $n = 2, 3, \ldots$ and consider positive integers K and P such that $2K \leq P$. Consider the definitions given in Chapter 5.2 with $\mathbb{G}(n; \upsilon)$

replaced by $\mathbb{K}(n;\theta)$, m_n replaced by n, I_n replaced by $I_n(\theta)$ and for each $r = 1, \ldots, n$, $A_{n,r}$ replaced by $A_{n,r}(\theta)$, $B_{n,r}$ replaced by $B_{n,r}(\theta)$ and $C_{n,r}$ replaced by $C_{n,r}(\theta)$. Here, $C_{n,r}(\theta)$ does not depend on n and we will use $C_r(\theta)$ for simplicity. For $r = n$ this notation now becomes consistent with $C_n(\theta)$ as defined in Section 7.3. The arguments of Chapter 5.2 now lead to the key bound:

$$\mathbb{P}\left[C_n(\theta)^c \cap I_n(\theta)\right] \leq \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta)\right] \tag{7.56}$$

upon using (5.14) with the aforementioned substitutions.

Consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ as in the statement of Proposition 7.4.2. In the right hand side of (7.56) we substitute θ by θ_n by means of this strongly admissible scaling. The proof of Proposition 7.4.2 will be completed once we show that

$$\lim_{n\to\infty} \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] = 0 \tag{7.57}$$

under the appropriate conditions. This approach was used to establish the one-law in Erdős-Rényi graphs [4], [32] where simple bounds can be derived for the probability terms in (7.57). Our situation is technically more involved and requires more delicate bounding arguments as will become apparent in the forthcoming sections.

7.4.4 Bounding the probabilities $\mathbb{P}\left[A_{n,r}(\theta)\right](r = 1, \ldots, n)$

Consider positive integers K and P such that $2K \leq P$. Fix $n = 2, 3, \ldots$ and pick $r = 1, \ldots, n-1$. In the course of evaluating $\mathbb{P}\left[A_{n,r}(\theta)\right]$, we shall make use of the rv $U_r(\theta)$ given by

$$U_r(\theta) := \left|\cup_{i=1}^{r} \Sigma_i(\theta)\right|. \tag{7.58}$$

The rv $U_r(\theta)$ counts the number of *distinct* keys issued to the nodes $1, \ldots, r$.

It is always the case that $U_r(\theta) \leq P$. However, the equivalence

$$B_{n,r}(\theta) = [(\cup_{i=1}^{r} \Sigma_i(\theta)) \cap \Sigma_j(\theta) = \emptyset,\ j = r+1, \ldots n]$$

implies that the set of nodes $\{1, \ldots, r\}$ *cannot* be isolated in $\mathbb{K}(n; \theta)$ if $P - U_r(\theta) < K$, i.e.,

$$B_{n,r}(\theta) \cap [P - U_r(\theta) < K] = \emptyset.$$

Hence, under the enforced assumptions on the rvs $\Sigma_1(\theta), \ldots, \Sigma_n(\theta)$, we readily obtain the expression

$$\mathbb{P}\left[B_{n,r}(\theta) | \Sigma_i(\theta),\ i = 1, \ldots, r\right] = \left(\frac{\binom{P-U_r(\theta)}{K}}{\binom{P}{K}}\right)^{n-r} \quad a.s.$$

on the event $[U_r(\theta) \leq P - K]$.

Upon conditioning on the rvs $\Sigma_1(\theta), \ldots, \Sigma_r(\theta)$ (which determine the event $C_r(\theta)$),

we conclude that

$$\begin{aligned}\mathbb{P}\left[A_{n,r}(\theta)\right] &= \mathbb{P}\left[C_r(\theta) \cap B_{n,r}(\theta)\right] \\ &= \mathbb{E}\left[\mathbf{1}\left[C_r^{\star}(\theta)\right] \cdot \left(\frac{\binom{P-U_r(\theta)}{K}}{\binom{P}{K}}\right)^{n-r}\right] \end{aligned} \tag{7.59}$$

with

$$C_r^{\star}(\theta) := C_r(\theta) \cap [U_r(\theta) \leq P - K].$$

The bound

$$\mathbb{P}\left[A_{n,r}(\theta)\right] \leq \mathbb{E}\left[\mathbf{1}\left[C_r^{\star}(\theta)\right] \cdot e^{-(n-r)\frac{K}{P}\cdot U_r(\theta)}\right] \tag{7.60}$$

follows by applying (5.21)(with $L = U_r(\theta)$) in Lemma 5.4.1.

The constraints

$$K \leq U_r(\theta) \leq \min\left(rK, P\right) \tag{7.61}$$

automatically imply $U_r(\theta) \leq P - K$ whenever $rK \leq P - K$, i.e., $(r+1)K \leq P$. Thus,

$$C_r^{\star}(\theta) = C_r(\theta), \quad r = 1, \ldots, r_n(\theta) \tag{7.62}$$

where we have set

$$r_n(\theta) := \min\left(r(\theta), \left\lfloor \frac{n}{2} \right\rfloor\right) \quad \text{with} \quad r(\theta) := \left\lfloor \frac{P}{K} \right\rfloor - 1.$$

This discussion already brings out a number of items that are likely to require some

attention: We will need good bounds for the probabilities $\mathbb{P}[C_r(\theta)]$ and $\mathbb{P}[C_r^\star(\theta)]$. Also, some of the distributional properties of the rv $U_r(\theta)$ are expected to play a role. Finally, different arguments are probably needed for the ranges $1 \leq r \leq r_n(\theta)$ and $r_n(\theta) < r \leq \lfloor \frac{n}{2} \rfloor$.

The next result is crucial to showing that for each $r = 2. \ldots, n$, the probability of the event $C_r(\theta)$ can be given an upper bound in terms of known quantities. First some notation: For each $r = 2, \ldots, n$, let $\mathbb{K}_r(n;\theta)$ stand for the subgraph $\mathbb{K}(n;\theta)(S)$ when $S = \{1, \ldots, r\}$. Also let $\mathcal{T}_r$ denote the collection of all spanning trees on the vertex set $\{1, \ldots, r\}$.

Lemma 7.4.5 *For each $r = 2, \ldots, n$, we have*

$$\mathbb{P}[T \subset \mathbb{K}_r(n;\theta)] = (1 - q(\theta))^{r-1}, \quad T \in \mathcal{T}_r \tag{7.63}$$

where the notation $T \subset \mathbb{K}_r(n;\theta)$ indicates that the tree T is a subgraph spanning $\mathbb{K}_r(n;\theta)$.

This last expression is analogous to the one found in Erdős-Rényi graphs [4] with $1 - q(\theta)$ playing the role of probability of link assignment, and this is in spite of possible correlations between some link assignments.

Proof. We shall prove the result by induction on $r = 2, \ldots, n$. For $r = 2$ the conclusion (7.63) is nothing more than (2.3) since $\mathcal{T}_2$ contains exactly one tree, and

this establishes the basis step.

Next, we consider the following induction step: Pick $r = 2, \ldots, n-1$ and assume that for each $s = 2, \ldots, r$, it is already known that

$$\mathbb{P}\left[T \subset \mathbb{K}_s(n;\theta)\right] = (1 - q(\theta))^{s-1}, \quad T \in \mathcal{T}_s. \tag{7.64}$$

We now show that (7.64) also holds for each $s = 2, \ldots, , r+1$. To that end, pick a tree T in $\mathcal{T}_{r+1}$ and identify its root.[2] Let i denote a node that is farthest from the root of T – There might be several such nodes. Also denote by p its unique parent, and let $D(p)$ denote the set of children of p. Obviously $D(p)$ is not empty as it contains node i; set $|D(p)| = d$. Next we construct a new tree $T^\star$ from T by removing from T all the edges from node p to the nodes in $D(p)$. By exchangeability, there is no loss of generality in assuming (as we do from now on) that the tree is rooted at node 1, that the unique parent p of the farthest node selected has label $r - d + 1$, and that its children have been labelled $r - d + 2, \ldots, r + 1$. With this convention, the tree $T^\star$ is defined on the set of nodes $\{1, \ldots, r - d + 1\}$.

It is plain that $T \subseteq \mathbb{K}_{r+1}(n, ; \theta)$ occurs if and only if the two sets of conditions

$$\Sigma_{r-d+1}(\theta) \cap \Sigma_\ell(\theta) \neq \emptyset, \quad \ell = r - d + 2, \ldots, r + 1$$

[2] As we are considering undirected graphs, all nodes can act as a root for the (undirected) tree T, in which case any one will do for the forthcoming discussion.

and

$$T^\star \subseteq \mathbb{K}_{r-d+1}(n;\theta)$$

both hold. Under the enforced independence assumptions we get

$$\mathbb{P}\left[\begin{array}{l} \Sigma_{r-d+1}(\theta)\cap\Sigma_\ell(\theta)\neq\emptyset, \\ \ell = r-d+2,\ldots,r+1 \end{array}\Big|\Sigma_1(\theta),\ldots,\Sigma_{r-d+1}(\theta)\right] = (1-q(\theta))^d.$$

Therefore, upon conditioning with respect to the rvs $\Sigma_1(\theta),\ldots,\Sigma_{r-d+1}(\theta)$ we readily find

$$\begin{aligned} \mathbb{P}\left[T\subseteq\mathbb{K}_{r+1}(n,;\theta)\right] &= (1-q(\theta))^d\mathbb{P}\left[T^\star\subseteq\mathbb{K}_{r-d+1}(n;\theta)\right] \\ &= (1-q(\theta))^d(1-q(\theta))^{r-d} \\ &= (1-q(\theta))^r \end{aligned} \tag{7.65}$$

as we use the induction hypothesis (7.64) for evaluating the probability of the event $[T^\star\subseteq\mathbb{K}_{r-d+1}(n;\theta)]$. This establishes the induction step. ■

The bound below now follows as in Erdős-Rényi graphs [4].

Lemma 7.4.6 *For each $r=2,\ldots,n$, we have*

$$\mathbb{P}\left[C_r(\theta)\right] \leq r^{r-2}\left(1-q(\theta)\right)^{r-1}. \tag{7.66}$$

Proof. Fix $r = 2, \ldots, n$. If $\mathbb{K}_r(n;\theta)$ is a connected graph, then it must contain a spanning tree on the vertex set $\{1, \ldots .r\}$, and a union bound argument yields

$$\mathbb{P}\left[C_r(\theta)\right] \leq \sum_{T \in \mathcal{T}_r} \mathbb{P}\left[T \subset \mathbb{K}(n;\theta)(S)\right]. \tag{7.67}$$

By Cayley's formula [6, 25] there are r^{r-2} trees on r vertices, i.e., $|\mathcal{T}_r| = r^{r-2}$, and (7.66) follows upon making use of (7.63). ■

The bound (7.60) and the inequality $U_r(\theta) \geq K$ together imply

$$\begin{aligned} \mathbb{P}\left[A_{n,r}(\theta)\right] &\leq \mathbb{P}\left[C_r(\theta)\right] \cdot e^{-(n-r)\frac{K^2}{P}} \\ &\leq r^{r-2}\left(1 - q(\theta)\right)^{r-1} \cdot e^{-(n-r)\frac{K^2}{P}} \end{aligned} \tag{7.68}$$

as we make use of Lemma 7.4.6 in the last step. Unfortunately, this bound turns out to be too loose for our purpose. As this can be traced to the crude lower bound used for $U_r(\theta)$, we expect that improvements are possible if we take into account the distributional properties of the rv $U_r(\theta)$. This step is taken in the next section.

7.4.5 The tail of the rv $U_r(\theta)$ and improved bounds

Consider positive integers K and P such that $K \leq P$. Rough estimates will suffice to get the needed information regarding the distribution of the rv $U_r(\theta)$. This is the content of the next result.

Lemma 7.4.7 *For all $r = 1, 2, \ldots$, the bound*

$$\mathbb{P}\left[U_r(\theta) \leq x\right] \leq \binom{P}{x} \left(\frac{\binom{x}{K}}{\binom{P}{K}} \right)^r \tag{7.69}$$

holds whenever $x = K, \ldots, \min(rK, P)$.

Proof. For a given x in the prescribed range, we note that $U_r(\theta) \leq x$ implies that $\cup_{i=1}^r \Sigma_i(\theta)$ is contained in some set S of size x, so that

$$[U_r(\theta) \leq x] \subseteq \bigcup_{S \in \mathcal{P}_x} [\cup_{i=1}^r \Sigma_i(\theta) \subseteq S].$$

A standard union bound argument gives

$$\begin{aligned}
\mathbb{P}\left[U_r(\theta) \leq x\right] &\leq \sum_{S \in \mathcal{P}_x} \mathbb{P}\left[\cup_{i=1}^r \Sigma_i(\theta) \subseteq S\right] \\
&= \sum_{S \in \mathcal{P}_x} \mathbb{P}\left[\Sigma_i(\theta) \subseteq S,\ i = 1, \ldots, r\right] \\
&= \sum_{S \in \mathcal{P}_x} \prod_{i=1}^r \mathbb{P}\left[\Sigma_i(\theta) \subseteq S\right] \\
&= \sum_{S \in \mathcal{P}_x} \left(\mathbb{P}\left[\Sigma_1(\theta) \subseteq S\right]\right)^r
\end{aligned} \tag{7.70}$$

under the enforced assumptions on the rvs $\Sigma_1(\theta), \ldots, \Sigma_n(\theta)$.

Since every set S of size x contain $\binom{x}{K}$ further subsets of size K, we get

$$\mathbb{P}\left[\Sigma_1(\theta) \subseteq S\right] = \frac{\binom{x}{K}}{\binom{P}{K}}, \quad S \in \mathcal{P}_x.$$

Reporting this fact into (7.70) we readily obtain (7.69) from the fact $|\mathcal{P}_x| = \binom{P}{x}$. ■

Reporting (5.24) into (7.69) we conclude to a somewhat looser but simpler bound.

Lemma 7.4.8 *For all $r = 1, 2, \ldots$, the bounds*

$$\mathbb{P}\left[U_r(\theta) \leq x\right] \leq \binom{P}{x} \left(\frac{x}{P}\right)^{rK} \tag{7.71}$$

holds whenever $x = K, \ldots, \min(rK, P)$.

The bounds (7.69) and (7.71) trivially hold with $\mathbb{P}\left[U_r(\theta) \leq x\right] = 0$ when $x = 1, \ldots, K-1$ since we always have $U_r(\theta) \geq K$. We shall make repeated use of this fact as follows: For all $n, r = 1, 2, \ldots$, with $r < n$, we have

$$\begin{aligned} \binom{n}{r} \mathbb{P}\left[U_r(\theta) \leq x\right] &\leq \binom{n}{r}\binom{P}{x}\left(\frac{x}{P}\right)^{rK} \\ &\leq \binom{\lfloor P/\sigma \rfloor}{r}\binom{P}{x}\left(\frac{x}{P}\right)^{rK} \end{aligned} \tag{7.72}$$

on the range $x = 1, \ldots, \min(rK, P)$ *whenever* $\sigma n \leq P$ for some $\sigma > 0$, a condition which is needed only for the last step and which implies $n \leq \lfloor \frac{P}{\sigma} \rfloor$ since n is an integer.

We are now in a position to improve on the bound (7.68): Fix $n = 2, 3, \ldots$ and pick $r = 2, \ldots, n-1$. For each positive integer x, the decomposition

$$\begin{aligned} \mathbb{P}\left[A_{n,r}(\theta)\right] &= \mathbb{P}\left[C_r(\theta) \cap B_{n,r}(\theta)\right] \\ &= \mathbb{P}\left[C_r(\theta) \cap B_{n,r}(\theta) \cap E_r(\theta; x)\right] \\ &\quad + \mathbb{P}\left[C_r(\theta) \cap B_{n,r}(\theta) \cap E_r(\theta; x)^c\right] \end{aligned} \tag{7.73}$$

holds where the event $E_r(\theta; x)$ is given by

$$E_r(\theta; x) := [U_r(\theta) \leq x].$$

The arguments leading to (7.59) also yield

$$\begin{aligned} &\mathbb{P}\left[C_r(\theta) \cap B_{n,r}(\theta) \cap E_r(\theta; x)\right] \\ &= \mathbb{E}\left[\mathbf{1}\left[C_r^\star(\theta)\right] \mathbf{1}\left[E_r(\theta; x)\right] \left(\frac{\binom{P-U_r(\theta)}{K}}{\binom{P}{K}}\right)^{n-r}\right] \\ &\leq \mathbb{E}\left[\mathbf{1}\left[C_r^\star(\theta)\right] \mathbf{1}\left[E_r(\theta; x)\right] e^{-(n-r)\frac{K}{P}U_r(\theta)}\right] \\ &\leq \mathbb{P}\left[C_r^\star(\theta) \cap E_r(\theta; x)\right] e^{-(n-r)\frac{K^2}{P}} \end{aligned} \tag{7.74}$$

given that $U_r(\theta) \geq K$. In a similar way we obtain

$$\mathbb{P}\left[C_r(\theta) \cap B_{n,r}(\theta) \cap E_r(\theta;x)^c\right] \leq \mathbb{P}\left[C_r^\star(\theta) \cap E_r(\theta;x)^c\right] e^{-(n-r)\frac{K}{P}(x+1)} \tag{7.75}$$

since $U_r(\theta) \geq x+1$ on the complement $E_r(\theta;x)^c$. Reporting (7.74) and (7.75) into (7.73) leads to the following fact.

Lemma 7.4.9 *Consider positive integers K and P such that $K \leq P$. With $n = 2, 3, \ldots$ and $r = 1, \ldots, n$, we have*

$$\mathbb{P}\left[A_{n,r}(\theta)\right] \leq \mathbb{P}\left[E_r(\theta;x)\right] e^{-(n-r)\frac{K^2}{P}} + \mathbb{P}\left[C_r(\theta)\right] e^{-(n-r)\frac{K}{P}(x+1)} \tag{7.76}$$

for each positive integer x.

This decomposition combines with Lemma 7.4.6 to provide bounds which are tighter than (7.68).

7.4.6 Outlining the proof of Proposition 7.4.2

It is now clear how to proceed: Consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ as in the statement of Proposition 7.4.2. Under (7.25) we necessarily have $\lim_{n\to\infty} \frac{P_n}{K_n} = \infty$ as discussed at the end of Section 7.3. As a result, $\lim_{n\to\infty} r_n(\theta_n) =$

∞, and for any given integer $R \geq 2$ we have

$$R < r_n(\theta_n), \quad n \geq n^\star(R) \tag{7.77}$$

for some finite integer $n^\star(R)$.

For the time being, pick an integer $R \geq 2$ (to be specified in Section 7.4.8), and on the range $n \geq n^\star(R)$ consider the decomposition

$$\begin{aligned} \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] &= \sum_{r=2}^{R} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] \\ &\quad + \sum_{r=R+1}^{r_n(\theta)} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] \\ &\quad + \sum_{r=r_n(\theta_n)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right]. \end{aligned} \tag{7.78}$$

Let n go to infinity: The desired convergence (7.57) will be established if we show

$$\lim_{n\to\infty} \sum_{r=2}^{R} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] = 0, \tag{7.79}$$

$$\lim_{n\to\infty} \sum_{r=R+1}^{r_n(\theta_n)} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] = 0 \tag{7.80}$$

and

$$\lim_{n\to\infty} \sum_{r=r_n(\theta_n)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] = 0. \tag{7.81}$$

The next sections are devoted to proving the validity of (7.79), (7.80) and (7.81) by repeated applications of Lemma 7.4.9. We address these three cases by making

use of the bounds (7.76) with

$$x = \lfloor (1+\varepsilon) K_n \rfloor, \quad \varepsilon \in (0, \frac{1}{2}),$$

$$x = \lfloor \lambda r K_n \rfloor, \quad \lambda \in (0,1),$$

and

$$x = \lfloor \mu P_n \rfloor, \quad \mu \in (0,1),$$

respectively. Finally, we note by convexity that the inequality

$$(x+y)^p \leq 2^{p-1}(x^p + y^p), \qquad \begin{matrix} x, y \geq 0 \\ p \geq 1 \end{matrix} \tag{7.82}$$

holds.

Before getting on the way, we close this section by highlighting key differences between our approach and the one used in the papers [3, 9]. The observation yielding (7.56), which forms the basis of our discussion, is also used in some form as the starting point in both these references. However, these authors did not take advantage of the fact that the sufficiently tight bound (7.66) is available for the probability of the event $C_r(\theta)$, a consequence of the *exact* expression (7.63). Through this bound, we can leverage strong admissibility (via (7.31)) to get

$$(1 - q(\theta_n)) \leq (1+\delta) \cdot \frac{K_n^2}{P_n}$$

for n sufficiently large with any $0 < \delta < 1$, in which case

$$\mathbb{P}\left[C_r(\theta_n)\right] \leq r^{r-2} \left((1+\delta) \cdot \frac{K_n^2}{P_n} \right)^{r-1}$$

for each $r = 2, 3, \ldots, n$. This opens the way to using the properties of the scaling by means of its deviation function defined by (7.1) – Such a line of arguments cannot be made if the scaling is merely admissible.

The bound (7.76) arises from the need to efficiently bound the rv $U_r(\theta_n)$. Indeed, if it were the case that $U_r(\theta_n) = rK_n$ for each $r = 1, \ldots, \lfloor \frac{n}{2} \rfloor$, then the conjecture (7.1)-(7.2) would readily follow as in Erdős-Rényi graphs [4] by simply making use of the bound (7.68), e.g., see the arguments in [4, 32]. In addition, the constraint $U_r(\theta_n) \leq \min(rK_n, P_n)$ already suggests that the cases $rK_n \leq P_n$ and $P_n < rK_n$ be considered separately, with a different decomposition (7.76) on each range – This was also the approach taken in the references [3] and [9]. Interestingly enough, a further decomposition of the range $r = 1, \ldots, \lfloor \frac{P_n}{K_n} \rfloor$ is needed to establish Theorem 7.1.1. In particular, using the bound (7.76) with $x = \lfloor \lambda r K_n \rfloor$ for sufficiently small λ in $(0, 1)$ across the entire range $r = 1, \ldots, \lfloor \frac{P_n}{K_n} \rfloor$ would not suffice for very small values of r: In that range the obvious bound $U_r(\theta_n) \geq K_n$ might be tighter than $U_r(\theta_n) \geq \lfloor \lambda r K_n \rfloor$, and another form of the bound (7.76) is needed to obtain the desired results, hence (7.78).

7.4.7 Establishing (7.79)

Consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ whose deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies $\lim_{n\to\infty} \alpha_n = \infty$. According to this scaling, for each $r = 2, 3, \ldots$ and $n = r+1, r+2, \ldots$, replace θ by θ_n in Lemma 7.4.9 with $x = \lfloor (1+\varepsilon)K_n \rfloor$ for some ε in $(0, \frac{1}{2})$. For an arbitrary integer $R \geq 2$, the convergence (7.79) will follow if we show that

$$\lim_{n\to\infty} \binom{n}{r} \mathbb{P}\left[C_r(\theta_n)\right] e^{-(n-r)\frac{K_n}{P_n}(\lfloor (1+\varepsilon)K_n \rfloor + 1)} = 0 \tag{7.83}$$

and

$$\lim_{n\to\infty} \binom{n}{r} \mathbb{P}\left[E_r\left(\theta_n; \lfloor (1+\varepsilon)K_n \rfloor\right)\right] e^{-(n-r)\frac{K_n^2}{P_n}} = 0 \tag{7.84}$$

for *each* $r = 2, 3, \ldots$. These two convergence statements are established below in Proposition 7.4.10 and Proposition 7.4.11, respectively.

Proposition 7.4.10 *Consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ whose deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies $\lim_{n\to\infty} \alpha_n = \infty$. With $\varepsilon > 0$, the convergence (7.83) holds for each $r = 2, 3, \ldots$.*

Proof. Pick $r = 2, 3, \ldots$ and $\varepsilon > 0$, and consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$. We combine the bounds (5.25) and (7.66) to write

$$\begin{aligned} &\binom{n}{r} \mathbb{P}\left[C_r(\theta_n)\right] e^{-(n-r)\frac{K_n}{P_n}(\lfloor (1+\varepsilon)K_n \rfloor + 1)} \\ &\leq \left(\frac{en}{r}\right)^r r^{r-2} \left(1 - q(\theta_n)\right)^{r-1} e^{-(n-r)\frac{K_n}{P_n}(\lfloor (1+\varepsilon)K_n \rfloor + 1)} \end{aligned}$$

$$\leq \quad \left(\frac{e^r}{r^2}\right) n^r \left(1 - q(\theta_n)\right)^{r-1} e^{-(n-r)\frac{K_n^2}{P_n}(1+\varepsilon)} \tag{7.85}$$

for all $n = r+1, r+2, \ldots$. Thus, it follows from Lemma 7.4.4 (via (7.50)) that the convergence (7.83) will be established if we show that

$$\lim_{n\to\infty} n^r \left(\frac{K_n^2}{P_n}\right)^{r-1} e^{-(n-r)\frac{K_n^2}{P_n}(1+\varepsilon)} = 0. \tag{7.86}$$

This step relies on the strong admissibility of the scaling.

On the range where (7.85) holds, we find with the help of (7.1) that

$$\begin{aligned}
& n^r \left(\frac{K_n^2}{P_n}\right)^{r-1} e^{-(n-r)\frac{K_n^2}{P_n}(1+\varepsilon)} \\
&= \quad n^r \left(\frac{\log n + \alpha_n}{n}\right)^{r-1} \cdot e^{-(n-r)\frac{\log n + \alpha_n}{n}(1+\varepsilon)} \\
&= \quad n \left(\log n + \alpha_n\right)^{r-1} \cdot e^{-(1+\varepsilon)(1-\frac{r}{n})\log n} \cdot e^{-(1+\varepsilon)(1-\frac{r}{n})\alpha_n} \\
&= \quad n^{1-(1+\varepsilon)(1-\frac{r}{n})} \cdot \left(\log n + \alpha_n\right)^{r-1} \cdot e^{-(1+\varepsilon)(1-\frac{r}{n})\alpha_n} \\
&= \quad n^{-\varepsilon+(1+\varepsilon)\frac{r}{n}} \cdot \left(\log n + \alpha_n\right)^{r-1} \cdot e^{-(1+\varepsilon)(1-\frac{r}{n})\alpha_n}.
\end{aligned} \tag{7.87}$$

Under the condition $\lim_{n\to\infty} \alpha_n = \infty$ it is plain that

$$\lim_{n\to\infty} n^{-\varepsilon+(1+\varepsilon)\frac{r}{n}} (\log n)^{r-1} e^{-(1+\varepsilon)(1-\frac{r}{n})\alpha_n} = 0$$

and

$$\lim_{n\to\infty} n^{-\varepsilon+(1+\varepsilon)\frac{r}{n}} \alpha_n^{r-1} e^{-(1+\varepsilon)(1-\frac{r}{n})\alpha_n} = 0.$$

Letting n go to infinity in (7.87) we readily get (7.86) by making use of (7.82). ■

Proposition 7.4.11 *Consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ whose deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies $\lim_{n\to\infty} \alpha_n = \infty$. For every ε in $(0, \frac{1}{2})$, the convergence (7.84) holds for each $r = 2, 3, \ldots$.*

Proof. Pick $r = 2, 3, \ldots$ and ε in $(0, \frac{1}{2})$, and consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$. For n sufficiently large, we use (7.71) with $x = \lfloor (1+\varepsilon)K_n \rfloor$ to obtain

$$\begin{aligned}
&\binom{n}{r} \mathbb{P}\left[E_r(\theta_n; \lfloor (1+\varepsilon)K_n \rfloor)\right] \\
&\leq \binom{n}{r} \binom{P_n}{\lfloor K_n(1+\varepsilon) \rfloor} \left(\frac{\lfloor K_n(1+\varepsilon) \rfloor}{P_n} \right)^{rK_n} \\
&\leq n^r \left(\frac{eP_n}{\lfloor K_n(1+\varepsilon) \rfloor} \right)^{\lfloor K_n(1+\varepsilon) \rfloor} \left(\frac{\lfloor K_n(1+\varepsilon) \rfloor}{P_n} \right)^{rK_n} \\
&\leq n^r \left(e^{\frac{\lfloor K_n(1+\varepsilon) \rfloor}{rK_n - \lfloor K_n(1+\varepsilon) \rfloor}} \frac{\lfloor K_n(1+\varepsilon) \rfloor}{P_n} \right)^{rK_n - \lfloor K_n(1+\varepsilon) \rfloor}.
\end{aligned}$$

The condition $r \geq 2$ implies the inequalities

$$\frac{\lfloor K_n(1+\varepsilon) \rfloor}{rK_n - \lfloor K_n(1+\varepsilon) \rfloor} \leq \frac{1+\varepsilon}{r - (1+\varepsilon)} \leq \frac{1+\varepsilon}{1-\varepsilon}$$

and

$$rK_n - \lfloor K_n(1+\varepsilon) \rfloor \geq K_n \left(r - (1+\varepsilon)\right) > 0.$$

Thus, upon setting

$$\Gamma(\varepsilon) := (1+\varepsilon)e^{\frac{1+\varepsilon}{1-\varepsilon}},$$

we conclude by strong admissibility (in view of (7.27)) that $\Gamma(\varepsilon)\cdot\frac{K_n}{P_n} < 1$ for all n sufficiently large, whence

$$e^{\frac{\lfloor K_n(1+\varepsilon)\rfloor}{rK_n-\lfloor K_n(1+\varepsilon)\rfloor}}\frac{\lfloor K_n(1+\varepsilon)\rfloor}{P_n} \leq \Gamma(\varepsilon)\cdot\frac{K_n}{P_n} < 1$$

on that range.

There we can write

$$\begin{aligned}
\binom{n}{r}\mathbb{P}\left[E_r(\theta_n;\lfloor(1+\varepsilon)K_n\rfloor)\right] &\leq n^r\left(\Gamma(\varepsilon)\cdot\frac{K_n}{P_n}\right)^{rK_n-\lfloor K_n(1+\varepsilon)\rfloor} \\
&\leq n^r\left(\Gamma(\varepsilon)\cdot\frac{K_n}{P_n}\right)^{K_n(r-1-\varepsilon)} \\
&\leq n^r\left(\Gamma(\varepsilon)\cdot\frac{K_n}{P_n}\right)^{2(r-1-\varepsilon)} \qquad (7.88) \\
&\leq n^r\left(\Gamma(\varepsilon)\cdot\frac{K_n^2}{P_n}\right)^{2(r-1-\varepsilon)} \\
&= n^r\left(\Gamma(\varepsilon)\cdot\frac{\log n+\alpha_n}{n}\right)^{2(r-1-\varepsilon)} \\
&= n^{-r+2+2\varepsilon}\left(\Gamma(\varepsilon)\cdot(\log n+\alpha_n)\right)^{2(r-1-\varepsilon)} \qquad (7.89)
\end{aligned}$$

where we obtain (7.88) upon using the fact $K_n \geq 2$. On the other hand we also have

$$e^{-(n-r)\frac{K_n^2}{P_n}} = e^{-(n-r)\frac{\log n+\alpha_n}{n}} = n^{-(1-\frac{r}{n})}\cdot e^{-\frac{n-r}{n}\alpha_n}. \qquad (7.90)$$

Therefore, upon multiplying (7.89) and (7.90) we see that Proposition 7.4.11 will follow if we show that

$$\lim_{n\to\infty} n^{-r+1+2\varepsilon+\frac{r}{n}} \cdot (\log n + \alpha_n)^{2(r-1-\varepsilon)} \cdot e^{-\frac{n-r}{n}\alpha_n} = 0. \tag{7.91}$$

The choice of ε and r ensures that $r - 1 - \varepsilon > 0$ and $-r + 1 + 2\varepsilon + \frac{r}{n} < 0$ for all n sufficiently large. The condition $\lim_{n\to\infty} \alpha_n = \infty$ now yields

$$\lim_{n\to\infty} n^{-r+1+2\varepsilon+\frac{r}{n}} \cdot (\log n)^{2(r-1-\varepsilon)} \cdot e^{-\frac{n-r}{n}\alpha_n} = 0 \tag{7.92}$$

and

$$\lim_{n\to\infty} n^{-r+1+2\varepsilon+\frac{r}{n}} \cdot \alpha_n^{2(r-1-\varepsilon)} \cdot e^{-\frac{n-r}{n}\alpha_n} = 0. \tag{7.93}$$

The desired conclusion (7.91) follows by making use of (7.92) and (7.93) with the help of the inequality (7.82). ■

Note that neither of these two results made use of the condition (7.3).

7.4.8 Establishing (7.80)

In order to establish (7.80) we will need two technical facts which are presented in Proposition 7.4.12 and Proposition 7.4.13.

Proposition 7.4.12 *Consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ whose*

deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ *satisfies* $\lim_{n\to\infty} \alpha_n = \infty$. *Then, with* $0 < \lambda < 1$ *and integer* $R \geq 2$, *we have*

$$\lim_{n\to\infty} \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[C_r(\theta_n)\right] e^{-(n-r)\frac{K_n}{P_n}(\lfloor \lambda r K_n \rfloor + 1)} = 0 \tag{7.94}$$

whenever λ *and* R *are selected so that*

$$2 < \lambda(R+1). \tag{7.95}$$

Proposition 7.4.12 is proved in Section 7.4.10. Next, with λ in $(0, \frac{1}{2})$ and $\sigma > 0$, we write

$$C(\lambda; \sigma) := \left(\frac{e^2}{\sigma}\right)^{\frac{\lambda}{1-2\lambda}}. \tag{7.96}$$

Proposition 7.4.13 *Consider a strongly admissible scaling* $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ *whose deviation function* $\alpha : \mathbb{N}_0 \to \mathbb{R}$ *satisfies* $\lim_{n\to\infty} \alpha_n = \infty$. *If there exists some* $\sigma > 0$ *such that (7.3) holds for all* $n = 1, 2, \ldots$ *sufficiently large, then we have*

$$\lim_{n\to\infty} \sum_{r=1}^{r_n(\theta_n)} \binom{n}{r} \mathbb{P}\left[E_r(\theta_n; \lfloor \lambda r K_n \rfloor)\right] = 0 \tag{7.97}$$

whenever λ *in* $(0, \frac{1}{2})$ *is selected small enough so that*

$$\max\left(2\lambda\sigma, \lambda^{1-2\lambda}, \lambda C(\lambda; \sigma)\right) < 1. \tag{7.98}$$

A proof of Proposition 7.4.13 can be found in Section 7.4.11. Note that for any $\sigma > 0$, $\lim_{\lambda \downarrow 0} \lambda C(\lambda; \sigma) = 0$ and that $\lim_{\lambda \downarrow 0} \lambda^{1-2\lambda} = 0$ so that the condition (7.98) can always be met by suitably selecting $\lambda > 0$ small enough.

We now turn to the proof of (7.80): Keeping in mind Proposition 7.4.12 and Proposition 7.4.13, we select λ sufficiently small in $(0, \frac{1}{2})$ to meet the condition (7.98) and then pick any integer $R \geq 2$ sufficiently large to ensure (7.95). Next consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ whose deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies the condition $\lim_{n \to \infty} \alpha_n = \infty$. Then, for each $n \geq n^\star(R)$ (with $n^\star(R)$ as specified at (7.77)), replace θ by θ_n according to this scaling, and for each $r = R+1, \ldots, r_n(\theta_n)$, set $x = \lfloor \lambda r K_n \rfloor$ in Lemma 7.4.9 with λ as specified earlier.

With these preliminaries in place, we see from Lemma 7.4.9 that (7.80) holds if both limits

$$\lim_{n \to \infty} \sum_{r=R+1}^{r_n(\theta_n)} \binom{n}{r} \mathbb{P}\left[C_r(\theta_n)\right] e^{-(n-r)\frac{K_n}{P_n}(\lfloor \lambda r K_n \rfloor + 1)} = 0$$

and

$$\lim_{n \to \infty} \sum_{r=R+1}^{r_n(\theta_n)} \binom{n}{r} \mathbb{P}\left[E_r(\theta_n; \lfloor \lambda r K_n \rfloor)\right] e^{-(n-r)\frac{K_n^2}{P_n}} = 0$$

hold. However, under (7.95) and (7.98), these two convergence statements are immediate from Proposition 7.4.12 and Proposition 7.4.13, respectively. ■

7.4.9 Establishing (7.81)

The following two results are needed to establish (7.81). The first of these results is given next with a proof available in Section 7.4.12.

Proposition 7.4.14 *Consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ whose deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies $\lim_{n\to\infty} \alpha_n = \infty$. If there exists some $\sigma > 0$ such that (7.3) holds for all $n = 1, 2, \ldots$ sufficiently large, then we have*

$$\lim_{n\to\infty} \sum_{r=r_n(\theta_n)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[E_r(\theta_n; \lfloor \mu P_n \rfloor)\right] = 0 \tag{7.99}$$

whenever μ in $(0, \frac{1}{2})$ is selected so that

$$\max\left(2\left(\sqrt{\mu}\left(\frac{e}{\mu}\right)^{\mu}\right)^{\sigma}, \sqrt{\mu}\left(\frac{e}{\mu}\right)^{\mu}\right) < 1. \tag{7.100}$$

We have $\lim_{\mu\downarrow 0}\left(\frac{e}{\mu}\right)^{\mu} = 1$, whence $\lim_{\mu\downarrow 0}\sqrt{\mu}\left(\frac{e}{\mu}\right)^{\mu} = 0$, and (7.100) can be made to hold for any $\sigma > 0$ by taking $\mu > 0$ sufficiently small. The next proposition is established in Section 7.4.13.

Proposition 7.4.15 *Consider an admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ whose deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies $\lim_{n\to\infty} \alpha_n = \infty$. If there exists some $\sigma > 0$ such that*

(7.3) holds for all $n = 1, 2, \ldots$ sufficiently large, then we have

$$\lim_{n\to\infty} \sum_{r=r_n(\theta_n)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[C_r(\theta_n)\right] e^{-(n-r)\frac{K_n}{P_n}(\lfloor \mu P_n \rfloor + 1)} = 0 \tag{7.101}$$

for each μ in $(0, 1)$.

The proof of (7.81) is now within easy reach: Consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ whose deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies $\lim_{n\to\infty} \alpha_n = \infty$. On the range where (7.3) holds, for each $n \geq n^\star(R)$ (with $n^\star(R)$ as specified at (7.77) where R and λ still satisfy (7.95) and (7.98)), replace θ by θ_n according to this scaling, and set $x = \lfloor \mu P_n \rfloor$ in Lemma 7.4.9 with μ as specified by (7.100). We get (7.81) as a direct consequence of Proposition 7.4.14 and Proposition 7.4.15. ■

7.4.10 A proof of Proposition 7.4.12

Let λ and R be as in the statement of Proposition 7.4.12, and pick a positive integer n such that $2(R + 1) < n$. Arguments similar to the ones leading to (7.85) yield

$$\binom{n}{r} \mathbb{P}\left[C_r(\theta_n)\right] e^{-(n-r)\frac{K_n}{P_n}(\lfloor \lambda r K_n \rfloor + 1)} \leq \left(\frac{e^r}{r^2}\right) n^r e^{-\lambda r (n-r)\frac{K_n^2}{P_n}} \left(1 - q(\theta_n)\right)^{r-1}$$

for all $r = 1, \ldots, n$. Thus, in order to establish (7.94), we need only show

$$\lim_{n\to\infty} \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \frac{e^r}{r^2} n^r e^{-\lambda r(n-r)\frac{K_n^2}{P_n}} \left(1 - q(\theta_n)\right)^{r-1} = 0. \tag{7.102}$$

As in the proof of Proposition 7.4.11, by the strong admissibility of the scaling (with the help of (7.51)), it suffices to show

$$\lim_{n\to\infty} \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \frac{e^r}{r^2} n^r e^{-\lambda r(n-r)\frac{K_n^2}{P_n}} \left((1+\delta)\frac{K_n^2}{P_n}\right)^{r-1} = 0 \tag{7.103}$$

with $0 < \delta < 1$.

Fix $n = 2, 3, \ldots$. For each $r = 1, \ldots, \lfloor \frac{n}{2} \rfloor$, we get

$$\begin{aligned}
&\left(\frac{e^r}{r^2}\right) n^r e^{-\lambda r(n-r)\frac{K_n^2}{P_n}} \left((1+\delta)\frac{K_n^2}{P_n}\right)^{r-1} \\
&= \left(\frac{e^r}{r^2}\right) n^r e^{-\lambda r(n-r)\frac{\log n + \alpha_n}{n}} \left((1+\delta)\frac{\log n + \alpha_n}{n}\right)^{r-1} \\
&= n\left(\frac{e^r}{r^2}\right) e^{-\lambda r(n-r)\frac{\log n + \alpha_n}{n}} \left((1+\delta)(\log n + \alpha_n)\right)^{r-1} \\
&\leq ne^r e^{-\lambda r(1-\frac{r}{n})(\log n + \alpha_n)} \left((1+\delta)(\log n + \alpha_n)\right)^{r-1} \\
&\leq ne^r e^{-\frac{\lambda}{2} r(\log n + \alpha_n)} \left((1+\delta)(\log n + \alpha_n)\right)^{r-1} \\
&= n\left(e^{1-\frac{\lambda}{2}(\log n + \alpha_n)}\right)^r \left((1+\delta)(\log n + \alpha_n)\right)^{r-1}
\end{aligned} \tag{7.104}$$

as we note that

$$1 - \frac{r}{n} \geq \frac{1}{2}, \quad r = 1, \ldots, \left\lfloor \frac{n}{2} \right\rfloor. \tag{7.105}$$

Next, we set

$$\Gamma_n(\lambda) := ne^{1-\frac{\lambda}{2}(\log n+\alpha_n)}$$

and

$$a_n(\lambda) := e^{1-\frac{\lambda}{2}(\log n+\alpha_n)}(1+\delta)(\log n+\alpha_n).$$

With this notation we conclude that

$$\begin{aligned}
&\sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{e^r}{r^2}\right) n^r e^{-\lambda r(n-r)\frac{K_n^2}{P_n}} \left((1+\delta)\frac{K_n^2}{P_n}\right)^{r-1} \\
&\leq \Gamma_n(\lambda) \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} a_n(\lambda)^{r-1} \\
&\leq \Gamma_n(\lambda) \sum_{r=R}^{\infty} a_n(\lambda)^r. \qquad (7.106)
\end{aligned}$$

Obviously, $\lim_{n\to\infty} a_n(\lambda) = 0$ under the condition $\lim_{n\to\infty} \alpha_n = \infty$, so that $a_n(\lambda) < 1$ for all n sufficiently large. On that range, the geometric series at (7.106) converges to a finite limit with

$$\sum_{r=R}^{\infty} a_n(\lambda)^r = \frac{a_n(\lambda)^R}{1-a_n(\lambda)}.$$

Thus,

$$\begin{aligned}
&\sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{e^r}{r^2}\right) n^r e^{-\lambda r(n-r)\frac{K_n^2}{P_n}} \left((1+\delta)\frac{K_n^2}{P_n}\right)^{r-1} \\
&\leq \Gamma_n(\lambda) \cdot \frac{a_n(\lambda)^R}{1-a_n(\lambda)} \\
&= C_{n,R}(\delta) \cdot n^{1-\frac{\lambda}{2}(R+1)} \cdot e^{-\frac{\lambda}{2}(R+1)\alpha_n} \cdot (\log n+\alpha_n)^R
\end{aligned}$$

with

$$C_{n,R}(\delta) := \frac{e^{R+1}(1+\delta)^R}{1-a_n(\lambda)}.$$

Under (7.95), the condition $\lim_{n\to\infty} \alpha_n = \infty$ implies

$$\lim_{n\to\infty} n^{1-\frac{\lambda}{2}(R+1)} \cdot e^{-\frac{\lambda}{2}(R+1)\alpha_n} \cdot (\log n)^R = 0$$

and

$$\lim_{n\to\infty} n^{1-\frac{\lambda(R+1)}{2}} \cdot e^{-\frac{\lambda(R+1)}{2}\alpha_n} \cdot \alpha_n^R = 0.$$

The desired conclusion (7.103) is now immediate with the help of the inequality (7.82).

■

Condition (7.3) played no role.

7.4.11 A proof of Proposition 7.4.13

We begin by providing bounds on the probabilities of interest entering (7.97). Recall the definitions of the quantities introduced before the statement of Proposition 7.4.13.

Proposition 7.4.16 *Consider positive integers K, P and n such that $2 \leq K \leq P$*

and $\sigma n \leq P$ for some $\sigma > 0$. For any λ in $(0, \frac{1}{2})$ small enough to ensure

$$\max\left(2\lambda\sigma, \lambda C(\lambda;\sigma)\right) < 1, \tag{7.107}$$

we have

$$\binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \lambda r K \rfloor)\right] \leq B(\lambda;\sigma;K)^r \tag{7.108}$$

for all $r = 1, \ldots, r_n(\theta)$ where we have set

$$\max\left(\lambda^{(1-2\lambda)K}, \left(\lambda^{1-2\lambda}\left(\frac{e^2}{\sigma}\right)^{\lambda}\right)^K, \frac{e^2}{\sigma K^{K-2}}\right). \tag{7.109}$$

Proof. Pick positive integers K, P and n as in the statement of Proposition 7.4.16. For each $r = 1, 2, \ldots, n$, we use (7.72) with $x = \lfloor \lambda r K \rfloor$ to find

$$\binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \lambda r K \rfloor)\right] \leq \binom{\lfloor \frac{P}{\sigma} \rfloor}{r} \binom{P}{\lfloor \lambda r K \rfloor} \left(\frac{\lfloor \lambda r K \rfloor}{P}\right)^{rK}. \tag{7.110}$$

On the range

$$r = 1, \ldots, r_n(\theta), \tag{7.111}$$

the inequalities

$$r \leq \left\lfloor \frac{P}{K} \right\rfloor - 1 < \frac{P}{K} \tag{7.112}$$

hold, whence $r < \frac{P}{2}$ since $K \geq 2$. Now if λ is selected in $(0, \frac{1}{2})$ such that $2\lambda\sigma < 1$, it then follows from (7.112) that $\lambda rK < \lambda P < \frac{P}{2\sigma}$ so that

$$\lfloor \lambda rK \rfloor \leq \left\lfloor \frac{P}{2\sigma} \right\rfloor \leq \frac{1}{2} \left\lfloor \frac{P}{\sigma} \right\rfloor. \tag{7.113}$$

Under these circumstances, we also have

$$rK - \lfloor 2\lambda rK \rfloor \geq (1 - 2\lambda) rK > 0. \tag{7.114}$$

Two possibilities arise:

Case I: $r \leq \lfloor \lambda rK \rfloor$ – Since $r \leq \lfloor \lambda rK \rfloor \leq \frac{1}{2} \left\lfloor \frac{P}{\sigma} \right\rfloor$ by (7.113), we can use (7.110) to get

$$\begin{aligned}
&\binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \lambda rK \rfloor)\right] \\
&\leq \binom{\lfloor \frac{P}{\sigma} \rfloor}{\lfloor \lambda rK \rfloor} \binom{P}{\lfloor \lambda rK \rfloor} \left(\frac{\lfloor \lambda rK \rfloor}{P} \right)^{rK} \\
&\leq \left(\frac{e \lfloor \frac{P}{\sigma} \rfloor}{\lfloor \lambda rK \rfloor} \right)^{\lfloor \lambda rK \rfloor} \left(\frac{eP}{\lfloor \lambda rK \rfloor} \right)^{\lfloor \lambda rK \rfloor} \left(\frac{\lfloor \lambda rK \rfloor}{P} \right)^{rK} \\
&\leq \left(\frac{e}{\sigma} \frac{P}{\lfloor \lambda rK \rfloor} \right)^{\lfloor \lambda rK \rfloor} \left(\frac{eP}{\lfloor \lambda rK \rfloor} \right)^{\lfloor \lambda rK \rfloor} \left(\frac{\lfloor \lambda rK \rfloor}{P} \right)^{rK} \\
&= \left(\frac{e^2}{\sigma} \right)^{\lfloor \lambda rK \rfloor} \left(\frac{\lfloor \lambda rK \rfloor}{P} \right)^{rK - 2\lfloor \lambda rK \rfloor} \\
&= \left(\left(\frac{e^2}{\sigma} \right)^{\frac{\lfloor \lambda rK \rfloor}{rK - 2\lfloor \lambda rK \rfloor}} \cdot \frac{\lfloor \lambda rK \rfloor}{P} \right)^{rK - 2\lfloor \lambda rK \rfloor} \\
&\leq \left(\max\left(1, C(\lambda; \sigma)\right) \cdot \frac{\lfloor \lambda rK \rfloor}{P} \right)^{rK - 2\lfloor \lambda rK \rfloor}
\end{aligned} \tag{7.115}$$

with $C(\lambda;\sigma)$ given by (7.96) – In the last step we made use of (7.114) together with the fact that

$$\frac{\lfloor \lambda r K \rfloor}{rK - 2\lfloor \lambda r K \rfloor} \leq \frac{\lambda r K}{rK - 2\lambda r K} = \frac{\lambda}{1-2\lambda}$$

since $\lfloor \lambda r K \rfloor \leq \lambda r K$.

On the range (7.111), we have $rK \leq P$ from (7.112) and reporting this fact into (7.115) yields

$$\binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \lambda r K \rfloor)\right] \leq \left(\lambda \cdot \max\left(1, C(\lambda;\sigma)\right)\right)^{rK - 2\lfloor \lambda r K \rfloor}.$$

In particular, if λ in $(0, \frac{1}{2})$ were selected such that $\lambda C(\lambda;\sigma) < 1$, then we have $\lambda \max\left(1, C(\lambda;\sigma)\right) < 1$ and we get

$$\binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \lambda r K \rfloor)\right] \leq \left(\lambda \cdot \max\left(1, C(\lambda;\sigma)\right)\right)^{(1-2\lambda)rK}$$

by recalling (7.114). Such a selection also implies that

$$\left(\lambda \cdot \max\left(1, C(\lambda;\sigma)\right)\right)^{(1-2\lambda)K} = \max\left(\lambda^{(1-2\lambda)K}, \left(\lambda^{1-2\lambda}\left(\frac{e^2}{\sigma}\right)^{\lambda}\right)^{K}\right)$$

and the conclusion

$$\binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \lambda r K \rfloor)\right] \leq \max\left(\lambda^{(1-2\lambda)K}, \left(\lambda^{1-2\lambda}\left(\frac{e^2}{\sigma}\right)^{\lambda}\right)^{K}\right)^{r} \tag{7.116}$$

follows.

Case II: $\lfloor \lambda r K \rfloor \leq r$ – On the range (7.111), we have $\lfloor \lambda r K \rfloor \leq r \leq \frac{P}{2}$ by virtue of (7.112). Using (7.110) we find

$$\begin{aligned}
\binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \lambda r K \rfloor)\right] &\leq \binom{\lfloor \frac{P}{\sigma} \rfloor}{r} \binom{P}{r} \left(\frac{\lfloor \lambda r K \rfloor}{P}\right)^{rK} \\
&\leq \left(\frac{e}{r}\left\lfloor \frac{P}{\sigma} \right\rfloor\right)^r \left(\frac{eP}{r}\right)^r \left(\frac{\lfloor \lambda r K \rfloor}{P}\right)^{rK} \\
&\leq \left(\frac{eP}{r\sigma}\right)^r \left(\frac{eP}{r}\right)^r \left(\frac{\lfloor \lambda r K \rfloor}{P}\right)^{rK}. \qquad (7.117)
\end{aligned}$$

The condition $\lfloor \lambda r K \rfloor \leq r$ now implies via (7.117) that

$$\begin{aligned}
\binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \lambda r K \rfloor)\right] &\leq \left(\frac{eP}{r\sigma}\right)^r \left(\frac{eP}{r}\right)^r \left(\frac{r}{P}\right)^{rK}. \\
&= \left(\frac{e^2}{\sigma}\right)^r \left(\frac{r}{P}\right)^{r(K-2)} \\
&= \left(\frac{e^2}{\sigma} \cdot \left(\frac{r}{P}\right)^{(K-2)}\right)^r \\
&\leq \left(\frac{e^2}{\sigma K^{K-2}}\right)^r \qquad (7.118)
\end{aligned}$$

since $r \leq \frac{P}{K}$ upon using (7.112). The proof of Proposition 7.4.16 is completed by combining the inequalities (7.116) and (7.118). ■

We can now turn to the proof of Proposition 7.4.13: Consider positive integers K, P and n as in the statement of Proposition 7.4.16. Pick λ in $(0, \frac{1}{2})$ which satisfies

(7.98) and note that (7.107) is also valid under this selection. We get

$$\sum_{r=1}^{r_n(\theta)} \binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \lambda r K \rfloor)\right] \;\leq\; \sum_{r=1}^{r_n(\theta)} B(\lambda; \sigma; K)^r \tag{7.119}$$

as we invoke Proposition 7.4.16. If it is the case that $B(\lambda; \sigma; K) < 1$, the geometric series is summable and

$$\sum_{r=1}^{r_n(\theta)} B(\lambda; \sigma; K)^r \leq \sum_{r=1}^{\infty} B(\lambda; \sigma; K)^r = \frac{B(\lambda; \sigma; K)}{1 - B(\lambda; \sigma; K)},$$

so that

$$\sum_{r=1}^{r_n(\theta)} \binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \lambda r K \rfloor)\right] \leq \frac{B(\lambda; \sigma; K)}{1 - B(\lambda; \sigma; K)}. \tag{7.120}$$

Now, consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ whose deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies $\lim_{n\to\infty} \alpha_n = \infty$. On the range where (7.3) holds, replace θ by θ_n in the last inequality according to this admissible scaling. From (7.1) we see that

$$K_n^2 = \frac{P_n}{n}(\log n + \alpha_n) \geq \sigma(\log n + \alpha_n),$$

so that $\lim_{n\to\infty} K_n = \infty$, whence

$$\lim_{n\to\infty} \left(\frac{e^2}{\sigma K_n^{K_n - 2}} \right) = 0. \tag{7.121}$$

Moreover, for any λ in $(0, \frac{1}{2})$ we have $\lambda^{1-2\lambda} < 1$, and any λ in the interval $(0, \frac{1}{2})$ satisfying (7.98) also satisfies the condition $\lambda C(\lambda; \sigma) < 1$, so that

$$\lambda^{1-2\lambda} \left(\frac{e^2}{\sigma} \right)^{\lambda} = (\lambda C(\lambda; \sigma))^{1-2\lambda} < 1.$$

As a result, under (7.98) we have

$$\lim_{n \to \infty} \max \left(\lambda^{1-2\lambda}, \lambda^{1-2\lambda} \left(\frac{e^2}{\sigma} \right)^{\lambda} \right)^{K_n} = 0$$

since $\lim_{n \to \infty} K_n = \infty$ via (7.3). Combining with (7.121), we now find

$$\lim_{n \to \infty} B(\lambda; \sigma; K_n) = 0 \tag{7.122}$$

so that $B(\lambda; \sigma; K_n) < 1$ for all n sufficiently large. Therefore, on that range (7.120) is valid under the enforced assumptions with θ is replaced by θ_n. Letting n go to infinity in (7.120) (with θ replaced by θ_n) we immediately get (7.97) via (7.122). ■

7.4.12 A proof of Proposition 7.4.14

Proposition 7.4.14 is an easy consequence of the following bound.

Proposition 7.4.17 *Consider positive integers K and P such that $2 \leq K$ and $2K \leq$*

P. For each μ in $(0, \frac{1}{2})$, we have

$$\sum_{r=r_n(\theta)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \mu P \rfloor)\right] \leq 2^n \left(\sqrt{\mu}\left(\frac{e}{\mu}\right)^{\mu}\right)^P \tag{7.123}$$

for all $n = 2, 3, \ldots$.

Proof. Fix $n = 2, 3, \ldots$. In establishing (7.123) we need only consider the case $r_n(\theta) < \lfloor \frac{n}{2} \rfloor$ (for otherwise (7.123) trivially holds), so that $r_n(\theta) = r(\theta)$ and $r_n(\theta)+1 = \lfloor \frac{P}{K} \rfloor$. The range $r_n(\theta) + 1 \leq r \leq \lfloor \frac{n}{2} \rfloor$ is then equivalent to

$$\left\lfloor \frac{P}{K} \right\rfloor \leq r \leq \left\lfloor \frac{n}{2} \right\rfloor,$$

hence

$$rK \geq \left(\frac{P}{K} - 1\right) K \geq \frac{P}{2}$$

as we make use of the condition $2K \leq P$ in the last step.

With μ in the interval $(0, \frac{1}{2})$ it follows that

$$\lfloor \mu P \rfloor \leq \frac{P}{2} \leq \min(rK, P) \tag{7.124}$$

and the bound (7.71) applies with $x = \lfloor \mu P \rfloor$ for all $r = r(\theta) + 1, \ldots, \lfloor \frac{n}{2} \rfloor$.

With this in mind, recall (7.105). We then get

$$\begin{aligned}
&\sum_{r=r_n(\theta)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \mu P \rfloor)\right] \\
&\leq \sum_{r=r(\theta)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \binom{P}{\lfloor \mu P \rfloor} \left(\frac{\lfloor \mu P \rfloor}{P}\right)^{rK} \\
&\leq \sum_{r=r(\theta)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \left(\frac{eP}{\lfloor \mu P \rfloor}\right)^{\lfloor \mu P \rfloor} \left(\frac{\lfloor \mu P \rfloor}{P}\right)^{rK} \\
&\leq \sum_{r=r(\theta)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} e^{\lfloor \mu P \rfloor} \left(\frac{\lfloor \mu P \rfloor}{P}\right)^{rK-\lfloor \mu P \rfloor} \\
&\leq \sum_{r=r(\theta)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} e^{\lfloor \mu P \rfloor} \mu^{rK-\lfloor \mu P \rfloor} \qquad (7.125) \\
&\leq \left(\frac{e}{\mu}\right)^{\lfloor \mu P \rfloor} \left(\sum_{r=r(\theta)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r}\right) \mu^{\frac{P}{2}}
\end{aligned}$$

since $\frac{P}{2} \leq rK$ for all $r = r(\theta)+1, \ldots, \lfloor \frac{n}{2} \rfloor$ as pointed out earlier. The passage to (7.125) made use of the fact that $rK - \lfloor \mu P \rfloor \geq 0$. The binomial formula now implies

$$\sum_{r=r(\theta)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \leq 2^n, \qquad (7.126)$$

so that

$$\sum_{r=r_n(\theta)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[E_r(\theta; \lfloor \mu P \rfloor)\right] \leq 2^n \left(\frac{e}{\mu}\right)^{\mu P} \mu^{\frac{P}{2}}$$

and the desired conclusion (7.123) follows. ■

To conclude the proof of Proposition 7.4.14 consider a strongly admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$. On the range where (7.3) holds, replace θ by θ_n in (7.123) according to this scaling. Observe that, under the condition $\sigma n \leq P_n$ for some $\sigma > 0$, the inequality

$$\left(\sqrt{\mu}\left(\frac{e}{\mu}\right)^{\mu}\right)^{P} \leq \left(\sqrt{\mu}\left(\frac{e}{\mu}\right)^{\mu}\right)^{\sigma n}$$

follows as soon as

$$\sqrt{\mu}\left(\frac{e}{\mu}\right)^{\mu} < 1, \tag{7.127}$$

and (7.123) takes the more compact form

$$\sum_{r=r_n(\theta_n)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[E_r(\theta_n; \lfloor \mu P_n \rfloor)\right] \leq \left(2\left(\sqrt{\mu}\left(\frac{e}{\mu}\right)^{\mu}\right)^{\sigma}\right)^{n}.$$

Letting n go to infinity in this last inequality, we readily get the desired conclusion (7.99) as we note that (7.127) is implied by (7.100). ■

We note that this result does not make use of the condition $\lim_{n\to\infty} \alpha_n = \infty$.

7.4.13 A proof of Proposition 7.4.15

Consider positive integers K and P such that $2 \leq K \leq P$, and pick μ in the interval $(0,1)$. For each $n = 2, 3, \ldots$, crude bounding arguments yield

$$\begin{aligned}
&\sum_{r=r_n(\theta)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[C_r(\theta)\right] \cdot e^{-(n-r)\frac{K}{P}(\lfloor \mu P \rfloor + 1)} \\
&\leq \sum_{r=r_n(\theta)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} e^{-(n-r)\frac{K}{P}(\mu P)} \\
&\leq \left(\sum_{r=r_n(\theta)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \right) e^{-\frac{n}{2} K \mu} \\
&\leq 2^n e^{-\frac{n}{2} K \mu} \qquad (7.128)
\end{aligned}$$

where we have used (7.105) and (7.126).

To complete the proof of Proposition 7.4.15, consider an admissible scaling $P, K : \mathbb{N}_0 \to \mathbb{N}_0$ whose deviation function $\alpha : \mathbb{N}_0 \to \mathbb{R}$ satisfies $\lim_{n\to\infty} \alpha_n = \infty$. Replace θ by θ_n in (7.128) according to this admissible scaling so that

$$\sum_{r=r_n(\theta_n)+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[C_{n,r}(\theta_n)\right] e^{-(n-r)\frac{K_n}{P_n}\lfloor \mu P_n \rfloor} \leq \left(2e^{-\frac{\mu K_n}{2}}\right)^n. \qquad (7.129)$$

The condition (7.3) implies

$$K_n^2 = \frac{\log n + \alpha_n}{n} \cdot P_n \ \geq \sigma \left(\log n + \alpha_n\right)$$

for $n = 1, 2, \ldots$ sufficiently large, whence $\lim_{n\to\infty} K_n = \infty$ since the assumed con-

dition $\lim_{n\to\infty} \alpha_n = \infty$ ensures that eventually $\alpha_n \geq 0$ for all n sufficiently large. Consequently,

$$\lim_{n\to\infty} \left(2e^{-\frac{\mu K_n}{2}}\right) = 0$$

and the conclusion (7.101) follows upon letting n go to infinity in (7.129). ■

Chapter 8

Connectivity in Random Pairwise Graphs

8.1 Introduction

The main goal of this chapter 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. In the original paper of Chan et al. [7] (as in the reference [21]), 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 (as in Chapter 2.3.1) 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.

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 that 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).

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 [4], random geometric random graphs [27] and random key graphs. As seen in Chapters 6 and 7, this equivalence, when it holds, can be 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. 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.

We close this chapter by discussing the number of keys that is required to be kept in the memory module of each sensor to achieve secure connectivity. The key rings produced by the pairwise scheme 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 $2K$ (in some probabilistic sense). We also show that the *maximum* key ring size is on the order $\log n$ with very high probability provided $K = O(\log n)$.

8.2 A proof of Theorem 3.3.1

In this section we establish Theorem 3.3.1, namely that if $K \geq 2$, then

$$\mathbb{P}\left[\mathbb{H}(n;K) \text{ is connected}\right] := P(n;K) \geq 1 - \frac{(K+1)^{K^2-1}}{2} \cdot n^{-(K^2-2)} \tag{8.1}$$

for all $n \geq \lceil e(K+1) \rceil$.

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

$$2 \leq K \quad \text{and} \quad e(K+1) < n \tag{8.2}$$

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.

Consider the definitions given in Chapter 5.2 with $\mathbb{G}(n;v)$ replaced by $\mathbb{H}(n;K)$, m_n replaced by n, I_n replaced by $I_n(K)$, and for each $r = 1, \ldots, n$, $A_{n,r}$ replaced by $A_{n,r}(K)$, $B_{n,r}$ replaced by $B_{n,r}(K)$ and $C_{n,r}$ replaced by $C_{n,r}(K)$ – For $r = n$, we use $C_n(K)$ (with a slight abuse of notation) in order to denote the event that $\mathbb{H}(n;K)$ is connected. The arguments of Chapter 5.2 yield

$$\mathbb{P}\left[C_n(K)^c \cap I_n(K)\right] \leq \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(K)\right]. \tag{8.3}$$

upon using (5.14) with the aforementioned substitutions.

Since each node in $\mathbb{H}(n;K)$ is connected to at least K other nodes, no node is ever isolated in $\mathbb{H}(n;K)$ and the event $I_n(K)$ is always in effect. Since a subset S of nodes can be isolated in $\mathbb{H}(n;K)$ only if $|S| \geq K+1$, we get

$$\mathbb{P}\left[C_n(K)^c\right] \leq \sum_{r=K+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(K)\right], \tag{8.4}$$

whence the key bound

$$\mathbb{P}\left[C_n(K)^c\right] \leq \sum_{r=K+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[B_{n,r}(K)\right] \tag{8.5}$$

readily follows 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

$$\mathbb{P}\left[B_{n,r}(K)\right] = \left(\frac{\binom{r-1}{K}}{\binom{n-1}{K}} \right)^r \cdot \left(\frac{\binom{n-r-1}{K}}{\binom{n-1}{K}} \right)^{n-r}. \tag{8.6}$$

Reporting (8.6) into (8.5) we get

$$\mathbb{P}\left[C_n(K)^c\right] \leq \sum_{r=K+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \left(\frac{\binom{r-1}{K}}{\binom{n-1}{K}} \right)^r \left(\frac{\binom{n-r-1}{K}}{\binom{n-1}{K}} \right)^{n-r}. \tag{8.7}$$

Invoking (5.24) and (5.25) into (8.7), we conclude that

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

$$\leq \sum_{r=K+1}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{ne}{r}\right)^r \left(\frac{r}{n}\right)^{rK} e^{-rK\frac{(n-r)}{n}}$$
$$= \sum_{r=K+1}^{\lfloor \frac{n}{2} \rfloor} \left(\left(\frac{r}{n}\right)^{K-1} e^{1-K\frac{(n-r)}{n}}\right)^r. \tag{8.8}$$

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 (8.8) we find

$$\mathbb{P}\left[C_n(K)^c\right] \leq \sum_{r=K+1}^{\lfloor \frac{n}{2} \rfloor} \left(\frac{r}{n}\right)^{r(K-1)}. \tag{8.9}$$

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

$$\left(\frac{x}{n}\right)^{x(K-1)} = e^{(K-1)f_n(x)}, \quad x \geq 1 \tag{8.10}$$

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 (8.2)). Using (8.10) together with the fact that

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

we obtain the equality

$$\left(\left(\frac{r}{n}\right)^{r(K-1)} : \ r = K+1, \ldots, \left\lfloor \frac{n}{2} \right\rfloor\right) = \left(\frac{K+1}{n}\right)^{K^2-1} \tag{8.11}$$

for all $n \geq n(K)$. Reporting (8.11) into (8.9), 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 (8.1) is established since $P(n,K) = 1 - \mathbb{P}\left[C_n(K)^c\right]$. ■

8.3 A proof of the zero-law in Theorem 3.3.2

In this section we establish the zero-law of the Theorem 3.3.2 by showing that

$$\lim_{n\to\infty} P(n;K) = 0 \quad \text{if } K = 1. \tag{8.12}$$

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

$$P(n;1) = \frac{\sum_{\boldsymbol{\gamma}} \mathbf{1}\left[H\boldsymbol{\gamma}(n) \text{ is connected }\right]}{(n-1)^n} \tag{8.13}$$

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

$$\lim_{n\to\infty} \frac{\sum_{\boldsymbol{\gamma}} \mathbf{1}\left[H_{\boldsymbol{\gamma}}(n) \text{ is connected }\right]}{(n-1)^n} = 0. \tag{8.14}$$

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, and two possibilities arise:

   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

$$\begin{aligned}
&\mathbb{P}\left[\mathbb{H}(n;1) \text{ is connected and has } n-1 \text{ edges}\right] \\
&= \frac{1}{(n-1)^n} \cdot \sum\nolimits_{\boldsymbol{\gamma}} \mathbf{1}\left[\begin{array}{c} H_{\boldsymbol{\gamma}}(n) \text{ is connected and} \\ \\ \text{has one bi-directional edge} \end{array}\right] \\
&= \frac{1}{(n-1)^n} \cdot \sum\nolimits_{\boldsymbol{\gamma}} \sum_{T \in \mathcal{T}_n} \mathbf{1}\left[H^{\star}_{\boldsymbol{\gamma}}(n) = T\right] \\
&= \frac{1}{(n-1)^n} \cdot (n-1) \cdot n^{n-2} \\
&= \frac{1}{n} \cdot \left(\frac{n}{n-1}\right)^{n-1}. \qquad (8.15)
\end{aligned}$$

It is now clear that

$$\lim_{n\to\infty} \mathbb{P}\left[\begin{array}{c} \mathbb{H}(n;1) \text{ is connected} \\ \\ \text{and has } n-1 \text{ edges} \end{array}\right] = 0. \qquad (8.16)$$

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 $\{2,3,4,1,\gamma_5,\gamma_6,\dots\gamma_n\}$ and $\{4,1,2,3,\gamma_5,\gamma_6,\dots\gamma_n\}$. 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

$$\begin{aligned}
&\mathbb{P}\left[\mathbb{H}(n;1) \text{ is connected and has } n \text{ edges}\right] \\
&= \frac{1}{(n-1)^n} \cdot \sum\nolimits_{\boldsymbol{\gamma}} \mathbf{1}\left[\begin{array}{c} H_{\boldsymbol{\gamma}}(n) \text{ is connected and} \\ \text{has exactly one cycle} \end{array}\right] \\
&= \frac{1}{(n-1)^n} \cdot \sum\nolimits_{\boldsymbol{\gamma}} \sum_{G \in \mathcal{T}_n^+} \mathbf{1}\left[H^\star_{\boldsymbol{\gamma}}(n) = G\right] \\
&= \frac{1}{(n-1)^n} \cdot 2 \cdot |\mathcal{T}_n^+|. \qquad (8.17)
\end{aligned}$$

However, it is known [16, p. 133-134] that

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

and reporting this fact into (8.17) gives

$$\begin{aligned}
\mathbb{P}\left[\mathbb{H}(n;1) \text{ is connected and has } n \text{ edges}\right] &\sim \frac{\sqrt{2\pi}}{2}\left(\frac{n}{n-1}\right)^n n^{-\frac{1}{2}} \\
&\sim \frac{\sqrt{2\pi}e}{2} n^{-\frac{1}{2}}. \qquad (8.18)
\end{aligned}$$

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 (8.16) and Facts 2-3, we now conclude that (8.14) holds. ■

8.4 Key ring sizes associated with the pairwise scheme: A proof of Lemma 3.3.4

The next two sections are devoted to obtaining conditions on a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ which ensure that all key rings have size of the order $\log n$. The proofs of both Lemma 3.3.4 and Theorem 3.3.5 are based on the following observations: Fix $n = 2, 3, \ldots$ and positive integer K with $K < n$. For each $i = 1, \ldots, n$ it is easy to see that

$$|\Sigma_{n,i}(K)| = K + B_{n,i}(K) \tag{8.19}$$

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

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

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

$$\mathbb{E}\left[B_{n,i}(K)\right] = (n-1) \cdot \frac{K}{n-1} = K. \tag{8.20}$$

Of particular relevance here is the following well-known concentration result for binomial rvs [27, Lemma 1.1, p. 16]: With $H(t) := 1 - t + t \log t$ $(t > 0)$, we have the concentration inequalities

$$\mathbb{P}\left[B_{n,1}(K) > K + t\right] \leq e^{-K \cdot H(\frac{K+t}{K})} \tag{8.21}$$

and

$$\mathbb{P}\left[B_{n,1}(K) < K - t\right] \leq e^{-K \cdot H(\frac{K-t}{K})} \tag{8.22}$$

where the additional condition $0 < t < K$ is required for (8.22) to hold. Simple calculations yield

$$K \cdot H\left(\frac{K \pm t}{K}\right) = K a\left(\pm \frac{t}{K}\right) \tag{8.23}$$

on the appropriate ranges as we make use of (3.21).

Taking the derivative of (3.21) we find

$$\frac{d}{d\tau} a(\tau) = \log\left(1 + \tau\right), \quad \tau > -1. \tag{8.24}$$

Therefore, the mapping $\tau \to a(\tau)$ is convex on $(-1, \infty)$, first strictly decreasing on $(-1, 0)$ and then strictly increasing on $(0, \infty)$ with $\lim_{\tau \downarrow -1} a(\tau) = 1$, $a(0) = 0$ and

$\lim_{\tau\to\infty} a(\tau) = \infty$. Therefore, $a(\tau) > 0$ on $(-1, 0) \cup (0, \infty)$. Since

$$\frac{d}{d\tau}\left(a(-\tau) - a(\tau)\right) = -\log(1-\tau^2) > 0, \quad 0 < \tau < 1,$$

it is easy to check that

$$a(\tau) < a(-\tau), \quad 0 < \tau < 1. \tag{8.25}$$

Fix the positive integers $n = 2, 3, \ldots$ and K with $K < n$. To take advantage of (8.21)-(8.22) we note from (8.19) that

$$|\Sigma_{n,1}(K)| - 2K = B_{n,1}(K) - K, \tag{8.26}$$

and for each $t > 0$, it follows that

$$\begin{aligned} &\mathbb{P}\left[||\Sigma_{n,1}(K)| - 2K| > t\right] \qquad (8.27)\\ &= \mathbb{P}\left[B_{n,1}(K) > K + t\right] + \mathbb{P}\left[B_{n,1}(K) < K - t\right]. \end{aligned}$$

Using (8.21)-(8.22) and the definition (3.22), we then conclude that

$$\mathbb{P}\left[||\Sigma_{n,1}(K)| - 2K| > t\right] \le b\left(\frac{t}{K}\right) e^{-Ka\left(\frac{t}{K}\right)} \tag{8.28}$$

as we recall (8.25) and the fact $\mathbb{P}\left[B_{n,1}(K) < K - t\right] = 0$ for $t \ge K$. The desired conclusion (3.23) follows as we replace t by cK in (8.28). ■

8.5 A proof of Theorem 3.3.5

Fix the positive integers $n = 2, 3, \ldots$ and K with $K < n$. Again using (8.19) we get

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

Since every key appears in exactly two different key rings it follows that

$$\sum_{i=1}^{n} |\Sigma_{n,i}(K)| = 2nK$$

by construction, whence $M_n(K) \geq 2K$.

As in the proof of Lemma 3.3.4, for any given $t > 0$, we now find

$$\begin{aligned} \mathbb{P}\left[|M_n(K) - 2K| > t \right] &= \mathbb{P}\left[M_n(K) - 2K > t \right] \\ &= \mathbb{P}\left[\max_{i=1,\ldots,n} \left(B_{n,i}(K) - K \right) > t \right] \\ &= \mathbb{P}\left[\max_{i=1,\ldots,n} B_{n,i}(K) > K + t \right]. \end{aligned} \tag{8.29}$$

A simple union argument shows that

$$\begin{aligned} \mathbb{P}\left[\max_{i=1,\ldots,n} B_{n,i}(K) > K + t \right] &= \mathbb{P}\left[\cup_{i=1}^{n} \left[B_{n,i}(K) > K + t \right] \right] \\ &\leq \sum_{i=1}^{n} \mathbb{P}\left[B_{n,i}(K) > K + t \right] \\ &= n\mathbb{P}\left[B_{n,1}(K) > K + t \right] \end{aligned} \tag{8.30}$$

since the rvs $B_{n,1}(K), \ldots, B_{n,n}(K)$ are identically distributed (but not independent). By the first concentration inequality (8.21) we then conclude that

$$\mathbb{P}\left[\max_{i=1,\ldots,n} B_{n,i}(K) > K + t\right] \leq e^{\log n - Ka\left(\frac{t}{K}\right)} \tag{8.31}$$

Next, consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ satisfying (3.26) for some $\lambda > 0$, and select the sequence $t : \mathbb{N}_0 \to \mathbb{R}_+$ given by

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

with $c > 0$. On that range, replacing K and t accordingly by K_n and t_n in (8.31), we conclude from (8.29) that

$$\mathbb{P}\left[|M_n(K_n) - 2K_n| > cK_n\right] \leq e^{-(-\log n + K_n a(c))}.$$

Under the enforced assumptions (3.26) it is easy to check that

$$-\log n + K_n a\left(c\right) = \left(-1 + \lambda a(c) + o(1)\right) \cdot \log n \tag{8.32}$$

and the bound (3.27) follows.

Finally pick $\lambda > 0$, and note that $\lim_{c \downarrow 0} \lambda a\left(c\right) = 0$ while

$$\lim_{c \uparrow 1} \lambda a\left(c\right) = \lambda\left(2\log 2 - 1\right) = \frac{\lambda}{\lambda^\star} \tag{8.33}$$

where

$$\lambda^\star := (2\log 2 - 1)^{-1} \simeq 2.6. \tag{8.34}$$

By the strict monotonicity of the mapping $a : \mathbb{R}_+ \to \mathbb{R}_+$, the equation

$$\lambda a\,(c) = 1, \quad c > 0$$

has a unique solution hereafter denoted $c(\lambda)$ – It is plain from (8.33) that taking $\lambda > \lambda^\star$ will ensure $c(\lambda) < 1$. By construction it is clear that

$$1 < \lambda a\,(c)\,, \quad c(\lambda) < c.$$

This last statement being equivalent to (3.28), the proof is now completed. ■

8.6 A proof of Theorem **??**

Fix integers K and n such that $K < n$ and pick $t > 0$. We start by observing that

$$\mathbb{P}\left[M_n > 2K + t\right] = \mathbb{P}\left[\max_{i=1,\ldots,n} B_{n,i} > K + t\right] \leq n \cdot \mathbb{P}\left[B_{n,1} > K + t\right] \tag{8.35}$$

upon recalling (8.30) in the last step. Next, we use a standard Chernoff-Hoeffding bound for binomial random variables: It is well-known [12, Thm. 1.1, p. 6] that

$$\mathbb{P}\left[B_{n,1} > (1+\delta)\mathbb{E}\left[B_{n,1}\right]\right] \leq \exp\left\{-\frac{\delta^2}{3}\mathbb{E}\left[B_{n,1}\right]\right\},$$

for any $\delta > 0$. Thus, we get

$$\mathbb{P}\left[B_{n,1} > K + \delta K\right] \leq \exp\left\{-\frac{\delta^2}{3}K\right\} \tag{8.36}$$

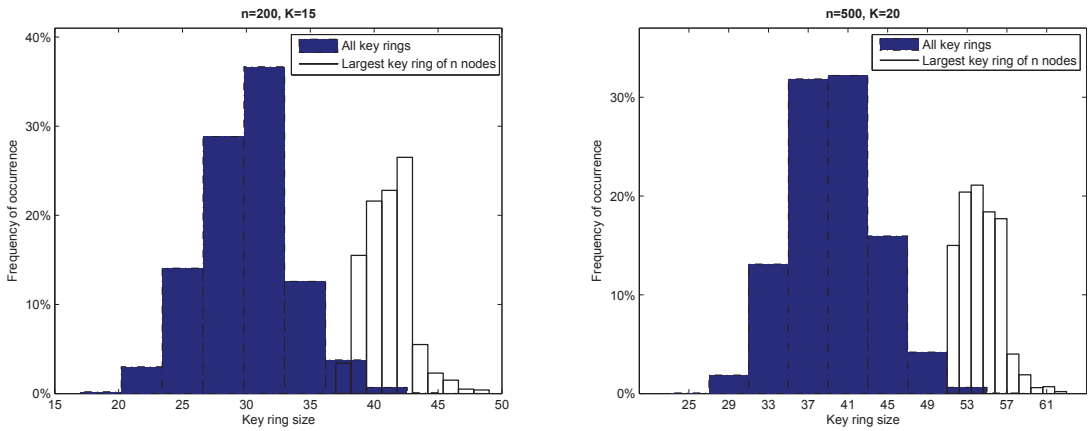
by using (8.20). Now consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ and report (8.36) into (8.35) by replacing δ with $\frac{t}{K}$. That yields

$$\mathbb{P}\left[M_n > 2K_n + t\right] \leq n \exp\left\{-\frac{t^2}{3K_n}\right\}$$

for any $t > 0$ and the desired conclusion (**??**) follows upon setting $t = \sqrt{(3+\epsilon)\log n \cdot K_n}$ for any given $\epsilon > 0$. ■

8.7 Simulation study

We now present experimental results that validate Lemma 3.3.4 and Theorem 3.3.5: For fixed values of n and K we have constructed key rings according to the mechanism presented in Section 3.1. For each pair of parameters n and K, the

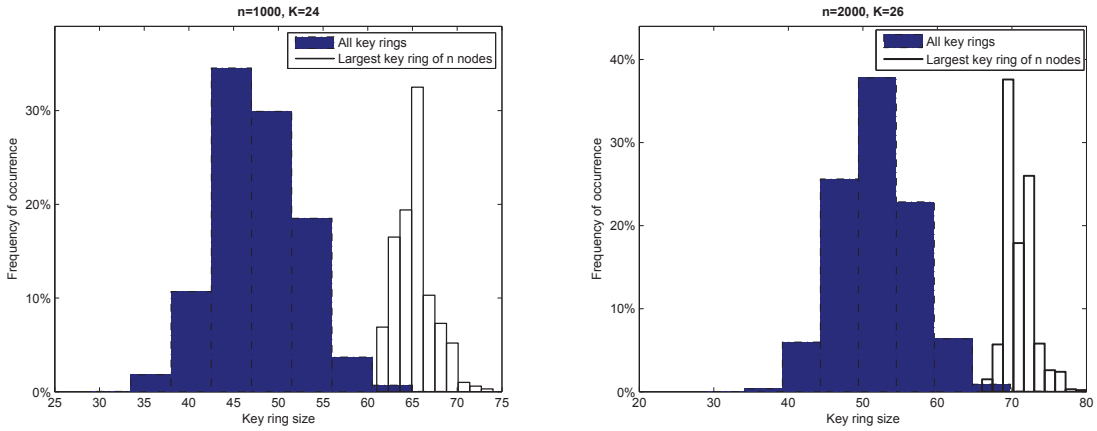


(a) (b)

Figure 8.1: *a) Key ring sizes observed in* $1,000$ *experiments for* $n = 200$ *and* $K = 15$ *– Only 2% of the key rings are larger than* $3K$ *and the largest key ring has size* 49. *b) Key ring sizes observed in* $1,000$ *experiments for* $n = 500$ *and* $K = 20$ *– Out of the* $500,000$ *key rings produced only* 9 *happened to be larger than* $3K$ *while the largest size observed is* 63.

experiments have been repeated $1,000$ times yielding $1,000 \times n$ key rings for each parameter pair. The results are depicted in Figures 8.1(a)-8.2(b) 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 8.1(a)-8.2(b) that the key ring sizes tend to concentrate around $2K$, validating the claim of Lemma 3.3.4. 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 3.3.5.



(a) (b)

Figure 8.2: *a) Key ring sizes observed in* $1,000$ *experiments for* $n = 1,000$ *and* $K = 24$ *– Although* $1,000,000$ *key rings are produced, only* 3 *of them happened to be larger than* $3K$ *and the largest observed key ring size is* 74. *b) 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.*

Chapter 9

Gradually deploying the pairwise scheme

9.1 Introduction

In this chapter we study the connectivity properties of the pairwise scheme in the setting where sensor nodes are deployed gradually over time. Under the implementation model presented in Section 3.3.3, we comment on how the parameter K needs to scale with n large in order to ensure that connectivity is *maintained* a.a.s. throughout gradual deployment. Using the results from Chapter 8.4, we also discuss the number of keys needed in the memory module of each sensor to achieve secure connectivity at every step of the gradual deployment.

Along these lines the key contributions of this chapter can be stated as follows: Let $\mathbb{H}_\gamma(n;K)$ denote the subgraph of $\mathbb{H}(n;K)$ restricted to the nodes $1,\ldots,\lfloor \gamma n \rfloor$. We first present scaling laws for the absence of isolated nodes in the form of a full zero-one law, and use these results to formulate conditions under which $\mathbb{H}_\gamma(n;K)$ is a.a.s. *not* connected. Then, with $0 < \gamma_1 < \gamma_2 < \ldots < \gamma_\ell < 1$, we give conditions on

n, K and γ_1 so that $\mathbb{H}_{\gamma_i}(n;K)$ is a.a.s. connected for each $i = 1, 2, \ldots, \ell$. We show that connectivity can be achieved a.a.s. when the number of keys to be stored in the memory modules is $O(\log n)$.

9.2 Establishing Theorem 3.3.6

Theorem 3.3.6 will be established in two steps. First we find conditions on γ, K and n such that the subgraph $\mathbb{H}_\gamma(n;K)$ of $\mathbb{H}(n;K)$ will be a.a.s. connected. Recalling the definitions

$$P_\gamma(n;K) := \mathbb{P}\left[\mathbb{H}_\gamma(n;K) \text{ is connected}\right] = \mathbb{P}\left[C_{n,\gamma}(K)\right]$$

introduced in Chapter 3, we have the following one law.

Theorem 9.2.1 *With γ in the unit interval $(0,1)$ and $c > 0$, consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that*

$$K_n \sim c\,\frac{\log n}{\gamma}. \tag{9.1}$$

Then, we have $\lim_{n\to\infty} P_\gamma(n;K_n) = 1$ whenever $c > 1$.

Theorem 9.2.1 is established in Section 9.3.

Next, in order to obtain conditions on γ, K and n such that $\mathbb{H}_\gamma(n;K)$ is a.a.s. *not* connected, we investigate the node isolation property in $\mathbb{H}_\gamma(n;K)$. Observe that the random graphs $\mathbb{H}(n;K)$ and $\mathbb{H}_\gamma(n;K)$ have very different neighborhood structures.

For example, any node in $\mathbb{H}(n;K)$ has degree at least K, so that no node is isolated in $\mathbb{H}(n;K)$. However, there is a positive probability that isolated nodes exist in $\mathbb{H}_\gamma(n;K)$. In fact, with

$$P^\star_\gamma(n;K_n) := \mathbb{P}\left[\mathbb{H}_\gamma(n;K) \text{ contains no isolated nodes}\right],$$

we have the following zero-one law which is established in Section 9.4.

Theorem 9.2.2 *With γ in the unit interval $(0,1)$, consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that (9.1) holds for some $c > 0$. Then, we have*

$$\lim_{n\to\infty} P^\star_\gamma(n;K_n) = \begin{cases} 0 & \text{if } \ c < r(\gamma) \\ \\ 1 & \text{if } c > r(\gamma) \end{cases} \tag{9.2}$$

where the threshold $r(\gamma)$ is given by

$$r(\gamma) := \left(1 - \frac{\log(1-\gamma)}{\gamma}\right)^{-1}. \tag{9.3}$$

As can be seen from Figure 9.2, $r(\gamma)$ is decreasing on the interval $[0,1]$ with $\lim_{\gamma\downarrow 0} r(\gamma) = \frac{1}{2}$ and $\lim_{\gamma\uparrow 1} r(\gamma) = 0$. Since a connected graph has no isolated nodes, Theorem 9.2.2 yields $\lim_{n\to\infty} \mathbb{P}\left[\mathbb{H}_\gamma(n;K_n) \text{ is connected}\right] = 0$ if the scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ satisfies (9.1) with $c < r(\gamma)$. Combining this with Theorem 9.2.1, we immediately

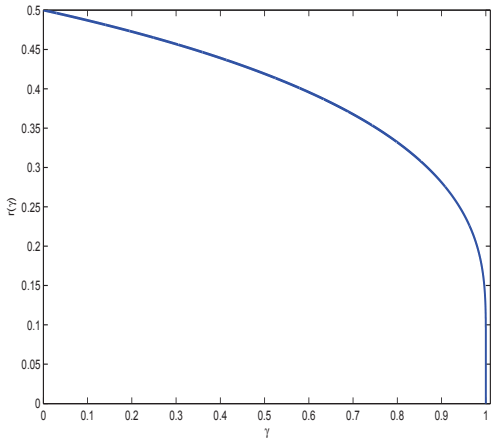


Figure 9.1: $r(\gamma)$ vs γ.

obtain Theorem 3.3.6.

9.3 A proof of Theorem 9.2.1

Fix $n = 2, 3, \ldots$ and γ in the interval $(0, 1)$, and consider a positive integer $K \geq 2$. Throughout the discussion, n is sufficiently large so that the conditions

$$2(K+1) < n, \quad K+1 \leq n - \lfloor \gamma n \rfloor \quad \text{and} \quad 2 < \gamma n \tag{9.4}$$

are all enforced; these conditions are made in order 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.

Consider the definitions given in Chapter 5.2 with $\mathbb{G}(n; \upsilon)$ replaced by $\mathbb{H}_\gamma(n; K)$, m_n replaced by $\lfloor \gamma n \rfloor$, I_n replaced by $I_{n,\gamma}(K)$, and for each $r = 1, \ldots, n$, $A_{n,r}$ re-

placed by $A_{n,\gamma,r}(K)$, $B_{n,r}$ replaced by $B_{n,\gamma,r}(K)$ and $C_{n,r}$ replaced by $C_{n,\gamma,r}(K)$ – For $r = \lfloor \gamma n \rfloor$, the notation $C_{n,\gamma,\lfloor \gamma n \rfloor}(K)$ coincides with $C_{n,\gamma}(K)$ as defined earlier. The arguments of Chapter 5.2 yield

$$\mathbb{P}\left[C_n(K)^c\right] \leq \sum_{r=1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(K)\right] \tag{9.5}$$

upon using (5.15) with the aforementioned replacements.

Recalling the obvious inclusion $A_{n,\gamma,r}(K) \subseteq B_{n,\gamma,r}(K)$, we now get

$$\mathbb{P}\left[C_{n,\gamma}(K)^c\right] \leq \sum_{r=1}^{\lfloor \frac{\gamma n}{2} \rfloor} \binom{\lfloor \gamma n \rfloor}{r} \mathbb{P}\left[B_{n,\gamma,r}(K)\right]. \tag{9.6}$$

Under the enforced assumptions, we have

$$\mathbb{P}\left[B_{n,\gamma,r}(K)\right] = \left(\frac{\binom{n-\lfloor \gamma n \rfloor + r - 1}{K}}{\binom{n-1}{K}} \right)^r \cdot \left(\frac{\binom{n-r-1}{K}}{\binom{n-1}{K}} \right)^{\lfloor \gamma n \rfloor - r}. \tag{9.7}$$

To see why this last relation holds, recall that for the set $\{1, \ldots, r\}$ to be isolated *in* $\mathbb{H}_\gamma(n; K)$ we need that (i) each of the nodes $r+1, \ldots, \lfloor \gamma n \rfloor$ are adjacent only to nodes *outside* the set of nodes $\{1, \ldots, r\}$; and (ii) none of the nodes $1, \ldots, r$ are adjacent with any of the nodes $r+1, \ldots, \lfloor \gamma n \rfloor$ – This last requirement does not preclude adjacency with any of the nodes $\lfloor \gamma n \rfloor + 1, \ldots, n$. Reporting (9.7) into (9.6), we conclude that

$$\mathbb{P}\left[C_{n,\gamma}(K)^c\right] \leq \sum_{r=1}^{\lfloor \frac{\gamma n}{2} \rfloor} \binom{\lfloor \gamma n \rfloor}{r} \left(\frac{\binom{n-\lfloor \gamma n \rfloor + r - 1}{K}}{\binom{n-1}{K}} \right)^r \cdot \left(\frac{\binom{n-r-1}{K}}{\binom{n-1}{K}} \right)^{\lfloor \gamma n \rfloor - r} \tag{9.8}$$

with conditions (9.4) ensuring that the binomial coefficients are well defined.

The remainder of the proof consists in bounding each of the terms in (9.8). To do so we make use of several standard bounds. Recall (5.24) and the well-known bound (5.25) which we use here in the form

$$\binom{\lfloor \gamma n \rfloor}{r} \leq \left(\frac{\lfloor \gamma n \rfloor e}{r} \right)^r, \quad r = 1, \ldots, \lfloor \gamma n \rfloor.$$

Now pick $r = 1, \ldots, \lfloor \gamma n \rfloor$. Under (9.4) we can apply these bounds to obtain

$$\begin{aligned}
&\binom{\lfloor \gamma n \rfloor}{r} \left(\frac{\binom{n-\lfloor \gamma n \rfloor + r - 1}{K}}{\binom{n-1}{K}} \right)^r \cdot \left(\frac{\binom{n-r-1}{K}}{\binom{n-1}{K}} \right)^{\lfloor \gamma n \rfloor - r} \\
&\leq \left(\frac{\lfloor \gamma n \rfloor e}{r} \right)^r \cdot \left(\frac{n - \lfloor \gamma n \rfloor + r - 1}{n-1} \right)^{rK} \cdot \left(\frac{n-r-1}{n-1} \right)^{K(\lfloor \gamma n \rfloor - r)} \\
&\leq \left(\frac{\gamma n e}{r} \right)^r \left(1 - \frac{\lfloor \gamma n \rfloor - r}{n-1} \right)^{rK} \left(1 - \frac{r}{n-1} \right)^{K(\lfloor \gamma n \rfloor - r)} \\
&\leq (\gamma n e)^r \cdot \left(1 - \frac{\lfloor \gamma n \rfloor - r}{n} \right)^{rK} \cdot \left(1 - \frac{r}{n} \right)^{K(\lfloor \gamma n \rfloor - r)} \\
&\leq (\gamma n e)^r \cdot e^{-\left(\frac{\lfloor \gamma n \rfloor - r}{n} \right) rK} \cdot e^{-\left(\frac{r}{n} \right)(\lfloor \gamma n \rfloor - r)K}.
\end{aligned}$$

It is plain that

$$\begin{aligned}
\mathbb{P}\left[C_{n,\gamma}(K)^c \right] &\leq \sum_{r=1}^{\lfloor \frac{\gamma n}{2} \rfloor} (\gamma n e)^r \cdot e^{-2\left(\frac{\lfloor \gamma n \rfloor - r}{n} \right) rK} \\
&\leq \sum_{r=1}^{\lfloor \frac{\gamma n}{2} \rfloor} \left(\gamma n e \cdot e^{-2\left(\frac{\lfloor \gamma n \rfloor - \lfloor \frac{\gamma n}{2} \rfloor}{n} \right) K} \right)^r \qquad (9.9)
\end{aligned}$$

as we note that

$$\frac{\lfloor \gamma n \rfloor - r}{n} \geq \frac{\lfloor \gamma n \rfloor - \lfloor \frac{\gamma n}{2} \rfloor}{n}, \quad r = 1, \ldots, \lfloor \frac{\gamma n}{2} \rfloor.$$

Next, consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that (9.1) holds for some $c > 1$, and replace K by K_n in (9.9) according to this scaling. Using the form (9.1) of the scaling we get,

$$a_n := \gamma n e \cdot e^{-2\left(\frac{\lfloor \gamma n \rfloor - \lfloor \frac{\gamma n}{2} \rfloor}{n}\right) K_n} = (\gamma e) \cdot n^{1 - 2c_n \left(\frac{\lfloor \gamma n \rfloor - \lfloor \frac{\gamma n}{2} \rfloor}{\gamma n}\right)}$$

for each $n = 1, 2, \ldots$, with $\lim_{n\to\infty} c_n = c$. It is a simple matter to check that

$$\lim_{n\to\infty} \left(2c_n \left(\frac{\lfloor \gamma n \rfloor - \lfloor \frac{\gamma n}{2} \rfloor}{\gamma n} \right) \right) = c,$$

so that by virtue of the fact that $c > 1$, we have

$$\lim_{n\to\infty} a_n = 0. \tag{9.10}$$

From (9.9) we conclude that

$$\mathbb{P}\left[C_{n,\gamma}(K_n)^c\right] \leq \sum_{r=1}^{\lfloor \frac{\gamma n}{2} \rfloor} (a_n)^r \leq \sum_{r=1}^{\infty} (a_n)^r = \frac{a_n}{1 - a_n}$$

where for n sufficiently large the summability of the geometric series is guaranteed by (9.10). The conclusion $\lim_{n\to\infty} \mathbb{P}\left[C_{n,\gamma}(K)^c\right] = 0$ is now a straightforward consequence of the last bound, again by virtue of (9.10). ■

9.4 A proof of Theorem 9.2.2

Fix $n = 2, 3, \ldots$ and consider γ in $(0, 1)$ and positive integer K such that $K < n$. We write

$$\chi_{n,\gamma,i}(K) := \mathbf{1}\left[\text{Node } i \text{ is isolated in } \mathbb{H}_\gamma(n; K)\right]$$

for each $i = 1, \ldots, \lfloor \gamma n \rfloor$. The number of isolated nodes in $\mathbb{H}_\gamma(n; K)$ is simply given by

$$I_\gamma(n; K) := \sum_{i=1}^{\lfloor \gamma n \rfloor} \chi_{n,\gamma,i}(K),$$

whence the random graph $\mathbb{H}_\gamma(n; K)$ has no isolated nodes if $I_\gamma(n; K) = 0$.

Theorem 9.2.2 will be established by applying the method of first and second moments (as discussed in Chapter 5.1) to the count variable $I_\gamma(n; K)$. Here, the exchangeability of the rvs $\chi_{n,\gamma,1}(K), \ldots, \chi_{n,\gamma,\lfloor \gamma n \rfloor}(K)$ ensure (5.6). Thus, we have (5.5) and (5.7) with Z_n replaced by $I_\gamma(n; K)$, the index m_n replaced by $\lfloor \gamma n \rfloor$, and the indicator variables $\{\chi_{m_n,i}, i = 1, \ldots, m_n\}$ replaced by $\{\chi_{n,\gamma,i}(\theta), i = 1, \ldots, \lfloor \gamma n \rfloor\}$. Thus, Theorem 9.2.1 will follow upon proving the next two technical lemmas which provide the appropriate versions of (5.8), (5.9) and (5.10).

Lemma 9.4.1 *Consider γ in $(0, 1)$ and a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that (9.1) holds*

for some $c > 0$. We have

$$\lim_{n\to\infty} n\mathbb{E}\left[\chi_{n,\gamma,1}(K_n)\right] = \begin{cases} 0 & \text{if } c > r(\gamma) \\ \\ \infty & \text{if } c < r(\gamma) \end{cases} \tag{9.11}$$

with $r(\gamma)$ specified via (9.3).

Lemma 9.4.2 *Consider γ in $(0,1)$ and a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that (9.1) holds for some $c > 0$. We have*

$$\limsup_{n\to\infty} \left(\frac{\mathbb{E}\left[\chi_{n,\gamma,1}(K_n)\chi_{n,\gamma,2}(K_n)\right]}{\left(\mathbb{E}\left[\chi_{n,\gamma,1}(K_n)\right]\right)^2} \right) \leq 1. \tag{9.12}$$

Proofs of Lemma 9.4.1 and Lemma 9.4.2 can be found in Section 9.4.1 and Section 9.4.2, respectively. To complete the proof of Theorem 9.2.2, pick a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that (9.1) holds for some $c > 0$. Under the condition $c > r(\gamma)$ we get (5.8) (with Z_n, m_n and $\chi_{m_n,1}$ replaced by $I_\gamma(n; K_n)$, $\lfloor \gamma n \rfloor$ and $\chi_{n,\gamma,1}(K_n)$, respectively) from Lemma 9.4.1, and the one-law $\lim_{n\to\infty} \mathbb{P}\left[I_\gamma(n; K_n) = 0\right] = 1$ follows. Next, assume the condition $c < r(\gamma)$. We obtain (5.9) and (5.10) (with m_n, $\chi_{m_n,1}$ and $\chi_{m_n,2}$ replaced by $\lfloor \gamma n \rfloor$, $\chi_{n,\gamma,1}(K_n)$ and $\chi_{n,\gamma,2}(K_n)$, respectively) with the help of Lemmas 9.4.1 and 9.4.2, respectively. The conclusion $\lim_{n\to\infty} \mathbb{P}\left[I_\gamma(n; K_n) = 0\right] = 0$ is now immediate from the arguments given in Section 5.1 and Theorem 9.2.1 is established. ■

9.4.1 A proof of Lemma 9.4.1

Fix $n = 2, 3, \ldots$ and γ in $(0, 1)$, and consider a positive integer K such that $K < n$. Here as well there is no loss of generality in assuming $n - \lfloor \gamma n \rfloor \geq K$ and $\lfloor \gamma n \rfloor > 1$. Under the enforced assumptions, we get

$$\begin{aligned} \mathbb{E}\left[\chi_{n,\gamma,1}(K)\right] &= \frac{\binom{n-\lfloor \gamma n \rfloor}{K}}{\binom{n-1}{K}} \left(\frac{\binom{n-2}{K}}{\binom{n-1}{K}} \right)^{\lfloor \gamma n \rfloor - 1} \\ &= a(n;K) \cdot \left(1 - \frac{K}{n-1}\right)^{\lfloor \gamma n \rfloor - 1} \end{aligned} \tag{9.13}$$

with

$$a(n;K) := \frac{(n - \lfloor \gamma n \rfloor)!}{(n - \lfloor \gamma n \rfloor - K)!} \cdot \frac{(n-1-K)!}{(n-1)!}.$$

Now pick a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that (9.1) holds for some $c > 0$ and replace K by K_n in (9.13) with respect to this scaling. Applying Stirling's formula

$$m! \sim \sqrt{2\pi m} \left(\frac{m}{e}\right)^m \quad (m \to \infty)$$

to the factorials appearing in (9.13), we readily get

$$\begin{aligned} a(n;K_n) &\sim \sqrt{\frac{(n - \lfloor \gamma n \rfloor)(n-1-K_n)}{(n - \lfloor \gamma n \rfloor - K_n)(n-1)}} \cdot \alpha_n \beta_n \\ &\sim \alpha_n \beta_n \end{aligned} \tag{9.14}$$

under the enforced assumptions on the scaling with

$$\begin{aligned}\alpha_n &:= \frac{(n-K_n-1)^{n-K_n-1}}{(n-1)^{n-1}} \\ &= \left(1-\frac{K_n}{n-1}\right)^{n-1}\cdot(n-K_n-1)^{-K_n}\end{aligned}$$

and

$$\begin{aligned}\beta_n &:= \frac{(n-\lfloor\gamma n\rfloor)^{n-\lfloor\gamma n\rfloor}}{(n-\lfloor\gamma n\rfloor-K_n)^{n-\lfloor\gamma n\rfloor-K_n}} \\ &= \left(1-\frac{K_n}{n-\lfloor\gamma n\rfloor}\right)^{-(n-\lfloor\gamma n\rfloor)}\cdot(n-\lfloor\gamma n\rfloor-K_n)^{K_n}\,.\end{aligned}$$

In obtaining the asymptotic behavior of (9.14) we rely on the following technical fact: For any sequence $m:\mathbb{N}_0\to\mathbb{N}_0$ with $m_n=\Omega(n)$, we have

$$\left(1-\frac{K_n}{m_n}\right)^{m_n}\sim e^{-K_n}. \tag{9.15}$$

whenever the scaling $K:\mathbb{N}_0\to\mathbb{N}_0$ satisfies (9.1). To see why (9.15) holds, recall the elementary decomposition (5.16) with $x=\frac{K_n}{m_n}$. Using this, we get

$$\left(1-\frac{K_n}{m_n}\right)^{m_n}=e^{-K_n}\cdot e^{-m_n\Psi\left(\frac{K_n}{m_n}\right)} \tag{9.16}$$

for all $n=1,2,\ldots$.

Under the enforced assumptions we have $m_n = \Omega(n)$ and $K_n = O(\log n)$, so that

$$\lim_{n\to\infty} \frac{K_n}{m_n} = 0 \quad \text{and} \quad \lim_{n\to\infty} m_n \left(\frac{K_n}{m_n}\right)^2 = 0.$$

It is now plain that

$$\lim_{n\to\infty} m_n \Psi\left(\frac{K_n}{m_n}\right) = 0$$

as we note (5.18). This establishes (9.15) via (9.16).

Using (9.15), first with $m_n = n - 1$, then with $m_n = n - \lfloor \gamma n \rfloor$, we obtain

$$\left(1 - \frac{K_n}{n-1}\right)^{n-1} \sim e^{-K_n}$$

and

$$\left(1 - \frac{K_n}{n - \lfloor \gamma n \rfloor}\right)^{-(n-\lfloor \gamma n \rfloor)} \sim \left(e^{-K_n}\right)^{-1} = e^{K_n},$$

whence

$$\alpha_n \beta_n \sim \left(\frac{n - \lfloor \gamma n \rfloor - K_n}{n - K_n - 1}\right)^{K_n}. \tag{9.17}$$

With the help of (9.13) and (9.14) we now conclude that

$$n\mathbb{E}\left[\chi_{n,\gamma,1}(K_n)\right] \sim n\left(1 - \frac{K_n}{n-1}\right)^{\lfloor \gamma n \rfloor - 1} \cdot \left(\frac{n - \lfloor \gamma n \rfloor - K_n}{n - K_n - 1}\right)^{K_n}. \tag{9.18}$$

A final application of (9.15), this time with $m_n = n-1$, gives

$$\begin{aligned}\left(1-\frac{K_n}{n-1}\right)^{\lfloor \gamma n\rfloor -1} &= \left(\left(1-\frac{K_n}{n-1}\right)^{n-1}\right)^{\frac{\lfloor \gamma n\rfloor -1}{n-1}} \\ &\sim e^{-\frac{\lfloor \gamma n\rfloor -1}{n-1}K_n} \end{aligned} \tag{9.19}$$

since $\lim_{n\to\infty}\frac{\lfloor \gamma n\rfloor -1}{n-1}=\gamma$. Reporting (9.19) into (9.18) we obtain

$$n\mathbb{E}\left[\chi_{n,\gamma,1}(K_n)\right] \sim e^{\zeta_n} \tag{9.20}$$

with

$$\zeta_n := \log n - \left(\frac{\lfloor \gamma n\rfloor -1}{n-1} + \log\left(\frac{n-\lfloor \gamma n\rfloor -K_n}{n-K_n-1}\right)\right)K_n$$

for all $n=1,2,\ldots$. Finally, from the condition (9.1) on the scaling, we see that

$$\lim_{n\to\infty}\frac{\zeta_n}{\log n} = 1-c+c\frac{\log(1-\gamma)}{\gamma} = 1-\frac{c}{r(\gamma)}.$$

Thus, $\lim_{n\to\infty}\zeta_n = -\infty$ (resp. ∞) if $r(\gamma)>c$ (resp. $r(\gamma)<c$) and the desired result follows upon using (9.20). ■

9.4.2 A proof of Lemma 9.4.2

Fix positive integers $n = 3, 4, \ldots$ and K with $K < n$. With γ in $(0, 1)$, we again assume that $n - \lfloor \gamma n \rfloor \geq K$ and $\lfloor \gamma n \rfloor > 1$. It is a simple matter to check that

$$\mathbb{E}\left[\chi_{n,\gamma,1}(K)\chi_{n,\gamma,2}(K)\right] = \left(\frac{\binom{n-\lfloor \gamma n \rfloor}{K}}{\binom{n-1}{K}}\right)^2 \left(\frac{\binom{n-3}{K}}{\binom{n-1}{K}}\right)^{\lfloor \gamma n \rfloor - 2}$$

and invoking (9.13) we readily conclude that

$$\begin{aligned}
&\frac{\mathbb{E}\left[\chi_{n,\gamma,1}(K)\chi_{n,\gamma,2}(K)\right]}{\left(\mathbb{E}\left[\chi_{n,\gamma,1}(K)\right]\right)^2} \\
&= \left(\frac{\binom{n-3}{K}}{\binom{n-1}{K}}\right)^{\lfloor \gamma n \rfloor - 2} \cdot \left(\frac{\binom{n-1}{K}}{\binom{n-2}{K}}\right)^{2(\lfloor \gamma n \rfloor - 1)} \\
&= \left(\left(\frac{n-1-K}{n-1}\right)\left(\frac{n-2-K}{n-2}\right)\right)^{\lfloor \gamma n \rfloor - 2} \cdot \left(\frac{n-1}{n-1-K}\right)^{2(\lfloor \gamma n \rfloor - 1)} \\
&= \left(\frac{n-2-K}{n-2}\right)^{\lfloor \gamma n \rfloor - 2} \cdot \left(\frac{n-1}{n-1-K}\right)^{\lfloor \gamma n \rfloor} \\
&= \left(1 - \frac{K}{n-2}\right)^{\lfloor \gamma n \rfloor - 2} \cdot \left(1 + \frac{K}{n-1-K}\right)^{\lfloor \gamma n \rfloor} \\
&\leq e^{-K \cdot E(n;K)} \qquad (9.21)
\end{aligned}$$

where we have set

$$E(n;K) := \frac{\lfloor \gamma n \rfloor - 2}{n-2} - \frac{\lfloor \gamma n \rfloor}{n-1-K}.$$

Elementary calculations show that

$$-K \cdot E(n;K) = \frac{\lfloor \gamma n \rfloor}{n-2} \cdot \frac{K(K-1)}{n-1-K} + \frac{2K}{n-2}.$$

Now pick a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that (9.1) holds for some $c > 0$. It is plain that $\lim_{n\to\infty} K_n E(n; K_n) = 0$ and the conclusion (9.12) follows from (9.21). ■

9.5 A proof of Theorem 3.3.7

Pick $0 < \gamma_1 < \gamma_2 < \ldots < \gamma_\ell \leq 1$ and consider a scaling $K : \mathbb{N}_0 \to \mathbb{N}_0$ such that

$$K_n \sim c \, \frac{\log n}{\gamma_1}$$

for some $c > 1$. It is plain that (3.33) will hold provided

$$\lim_{n\to\infty} \mathbb{P}\left[C_{n,\gamma_k}(K_n)\right] = 1, \quad k = 1, \ldots, \ell. \tag{9.22}$$

For each $k = 1, 2, \ldots, \ell$, we note that

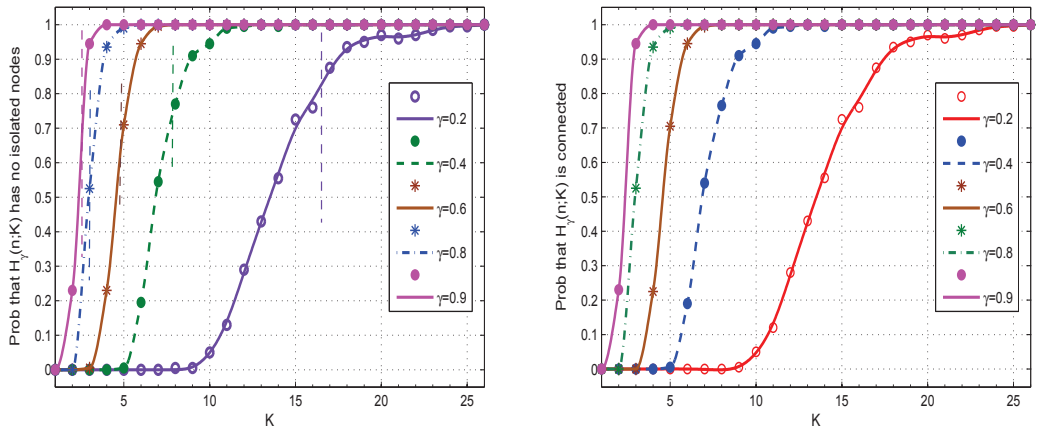
$$c \, \frac{\log n}{\gamma_1} = c_k \, \frac{\log n}{\gamma_k} \quad \text{with } c_k := c \, \frac{\gamma_k}{\gamma_1}$$

for all $n = 1, 2, \ldots$. But $c > 1$ implies $c_k > 1$ since $\gamma_1 < \ldots < \gamma_\ell$. As a result, $\mathbb{H}_{\gamma_k}(n; K_n)$ will be a.a.s. connected by virtue of Theorem 9.2.1 applied to $\mathbb{H}_{\gamma_k}(n; K)$, and (9.22) indeed holds. ■

9.6 Simulation study

We now present experimental results in support of theoretical findings. In each set of experiments, we fix n and γ. Then, we generate random graphs $\mathbb{H}_\gamma(n;K)$ for each $K = 1,\ldots,K_{\max}$ where the maximal value $K_{\max}$ is selected large enough. In each case, we check whether the generated random graph has isolated nodes and is connected. We repeat the process 200 times for each pair of values γ and K in order to estimate the probabilities of the events of interest. For various values of γ, Figure 9.2(a) depicts the estimated probability $P^\star_\gamma(n;K)$ that $\mathbb{H}_\gamma(n;K)$ has no isolated nodes as a function of K. Here, n is taken to be $1,000$. The plots in Figure 9.2(a) clearly confirm the claims of Theorem 9.2.2: In each case $P^\star_\gamma(n;K)$ exhibits a threshold behavior and the transitions from $P^\star_\gamma(n;K) = 0$ to $P^\star_\gamma(n;K) = 1$ take place around $K = r(\gamma)\frac{\log n}{\gamma}$ as dictated by Theorem 9.2.2; the critical value $K = r(\gamma)\frac{\log n}{\gamma}$ is shown by a vertical dashed line in each plot.

Similarly, Figure 9.2(b) shows the estimated probability $P_\gamma(n;K)$ as a function of K for various values of γ with $n = 1000$. For each specified γ, we see that the variation of $P_\gamma(n;K)$ with K is almost indistinguishable from that of $P^\star_\gamma(n;K)$ supporting the claim that $P_\gamma(n;K)$ exhibits a full zero-one law similar to that of Theorem 9.2.2 with a threshold behaving like $r(\gamma)$. We can also conclude by monotonicity that $P_\gamma(n;K) = 1$ whenever (3.29) holds with $c > 1$; this is in line with Theorem 9.2.1. Furthermore, it is evident from Figure 9.2(b) that for a given K and n, $P_\gamma(n;K)$ increases as γ increases supporting Theorem 3.3.7.



(a) (b)

Figure 9.2: *a) Probability that $\mathbb{H}_\gamma(n;K)$ contains no isolated for $n = 1000$; in each case, the empirical probability value is obtained through 200 experiments. Vertical dashed lines stand for the critical thresholds asserted by Theorem 9.2.2. The theoretical findings are in perfect agreement with the practical observations. b) Probability that $\mathbb{H}_\gamma(n;K)$ is connected for $n = 1,000$ obtained in the same way. The curves are almost indistinguishable from the corresponding ones of Part (a); this supports the claim that absence of isolated nodes and connectivity are asymptotically equivalent properties.*

Chapter 10

Connectivity of the pairwise scheme under an ON-OFF channel

10.1 Introduction

The previous two chapters were devoted to establishing connectivity results for the pairwise key distribution scheme of Chan et al. [7] under the assumption of full visibility. Here, we consider a more realistic setting that accounts for the possibility that communication links between nodes may not be available. Namely, we study the connectivity properties of the pairwise scheme under a simple communication model where channels are mutually independent, and are either on or off. As mentioned earlier, this amounts to an overall system model that is constructed by *intersecting* the random pairwise graph with an Erdős-Rényi (ER) graph [4]. For this new random graph structure, denoted $\mathbb{H}(n; K, p)$, we establish zero-one laws for graph connectivity and the absence of isolated nodes, as the model parameters are scaled with the number of users – We identify the critical thresholds and show that they coincide. To the

best of our knowledge, these full zero-one laws constitute the first *complete* analysis of a key distribution scheme under *non*-full visibility – Contrast this with the partial results by Yi et al. [48] for the absence of isolated nodes (under additional conditions) when the communication model is the disk model.

10.2 A proof of Theorem 3.3.9

The proof of Theorem 3.3.9 passes through the next two results which will be established in this Chapter. To lighten the notation we often group the parameters K and p into the ordered pair $\theta \equiv (K, p)$. Hence, a mapping $\theta : \mathbb{N}_0 \to \mathbb{N}_0 \times (0, 1)$ defines a scaling for the intersection graph $\mathbb{H} \cap \mathbb{G}(n; \theta)$ provided the condition (3.12) holds on the first component.

Recall the definition of the threshold function $\tau : [0, 1] \to [0, 1]$ given in (3.37). We first establish a zero-one law for the absence of isolated nodes in $\mathbb{H}(n; K, p)$.

Theorem 10.2.1 *Consider scalings $K : \mathbb{N}_0 \to \mathbb{N}_0$ and $p : \mathbb{N}_0 \to (0, 1)$ such that*

$$p_n \left(2K_n - \frac{K_n^2}{n-1} \right) \sim c \log n, \quad n = 1, 2, \ldots \tag{10.1}$$

for some $c > 0$. If $\lim_{n\to\infty} p_n = p^\star$ for some $p^\star$ in $[0,1]$, then we have

$$\lim_{n\to\infty} \mathbb{P}\left[\begin{array}{c} \mathbb{H} \cap \mathbb{G}(n;\theta_n) \text{ contains} \\ \text{no isolated nodes} \end{array}\right] = \begin{cases} 0 & \text{if } c < \tau(p^\star) \\ \\ 1 & \text{if } c > \tau(p^\star). \end{cases} \tag{10.2}$$

Next, we establish an analog of Theorem 10.2.1 for the property of graph connectivity.

Theorem 10.2.2 *Consider scalings $K : \mathbb{N}_0 \to \mathbb{N}_0$ and $p : \mathbb{N}_0 \to (0,1)$ such that (10.1) holds for some $c > 0$. If $\lim_{n\to\infty} p_n = p^\star$ for some $p^\star$ in $[0,1]$, then we have*

$$\lim_{n\to\infty} \mathbb{P}\left[\mathbb{H} \cap \mathbb{G}(n;\theta_n) \text{ is connected}\right] = \begin{cases} 0 & \text{if } c < \tau(p^\star) \\ \\ 1 & \text{if } c > \tau(p^\star) \end{cases} \tag{10.3}$$

where the threshold $\tau(p^\star)$ is given by (3.37).

The condition (10.1) on the scaling $\mathbb{N}_0 \to (0,1) \times \mathbb{N}_0$ will often be used in the equivalent form

$$p_n\left(2K_n - \frac{K_n^2}{n-1}\right) = c_n \log n, \quad n = 1, 2, \ldots \tag{10.4}$$

with the sequence $c : \mathbb{N}_0 \to \mathbb{R}_+$ satisfying $\lim_{n\to\infty} c_n = c$.

10.3 A proof of Theorem 10.2.1

We prove Theorem 10.2.1 by the method of first and second moments (see Chapter 5.1) applied to the total number of isolated nodes in $\mathbb{H} \cap \mathbb{G}(n;\theta)$. First some notation: Fix $n = 2, 3, \ldots$ and consider $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K such that $K < n$. With

$$\chi_{n,i}(\theta) := \mathbf{1}\,[\text{Node } i \text{ is isolated in } \mathbb{H} \cap \mathbb{G}(n;\theta)]$$

for each $i = 1, \ldots, n$, the number of isolated nodes in $\mathbb{H} \cap \mathbb{G}(n;\theta)$ is simply given by

$$I(n;\theta) := \sum_{i=1}^{n} \chi_{n,i}(\theta).$$

The random graph $\mathbb{H} \cap \mathbb{G}(n;\theta)$ has no isolated nodes if and only if $I(n;\theta) = 0$.

Recall the arguments given in Chapter 5.1 regarding the method of first and second moments. Observe that the rvs $\chi_{n,1}(\theta), \ldots, \chi_{n,n}(\theta)$ are exchangeable and therefore the special condition (5.6) holds. This yields (5.5) and (5.7) with Z_n replaced by the count variable $I(n;\theta)$, the index m_n replaced by n, and the indicator variables $\{\chi_{m_n,i}, i = 1, \ldots, m_n\}$ replaced by $\{\chi_{n,i}(\theta)$, i=1,..., n$\}$ as defined above. Thus, the proof of Theorem 10.2.1 passes through the next two technical propositions which establish (5.8), (5.9) and (5.10) under the appropriate conditions on the scaling $\theta : \mathbb{N}_0 \to \mathbb{N}_0 \times (0, 1)$.

Proposition 10.3.1 *Consider scalings $K : \mathbb{N}_0 \to \mathbb{N}_0$ and $p : \mathbb{N}_0 \to (0,1)$ such that (10.1) holds for some $c > 0$. Assume also that $\lim_{n\to\infty} p_n = p^\star$ exists. Then, we have*

$$\lim_{n\to\infty} n\mathbb{E}\left[\chi_{n,1}(\theta_n)\right] = \begin{cases} 0 & \text{if } \ c > \tau(p^\star) \\ \\ \infty & \text{if } c < \tau(p^\star) \end{cases} \tag{10.5}$$

where the threshold $\tau(p^\star)$ is given by (3.37).

A proof of Proposition 10.3.1 is given in Section 10.5.

Proposition 10.3.2 *Consider scalings $K : \mathbb{N}_0 \to \mathbb{N}_0$ and $p : \mathbb{N}_0 \to (0,1)$ such that (10.1) holds for some $c > 0$. Assume also that $\lim_{n\to\infty} p_n = p^\star$ exists. Then, we have*

$$\limsup_{n\to\infty} \left(\frac{\mathbb{E}\left[\chi_{n,1}(\theta_n)\chi_{n,2}(\theta_n)\right]}{\left(\mathbb{E}\left[\chi_{n,1}(\theta_n)\right]\right)^2} \right) \leq 1. \tag{10.6}$$

whenever $p^\star < 1$.

A proof of Proposition 10.3.2 can be found in Section 10.7. To complete the proof of Theorem 10.2.1, pick a scaling $\theta : \mathbb{N}_0 \to \mathbb{N}_0 \times (0,1)$ such that (10.1) holds for some $c > 0$ and $\lim_{n\to\infty} p_n = p^\star$ exists. Under the condition $c > \tau(p^\star)$ we get (5.8) from Proposition 10.3.1, and the one-law $\lim_{n\to\infty} \mathbb{P}\left[I(n;\theta_n) = 0\right] = 1$ follows. Next, assume that $c < \tau(p^\star)$ – This case is possible only if $p^\star < 1$ since $\tau(1) = 0$ as seen at (3.37). When $p^\star < 1$, we obtain (5.9) and (5.10) with the help of Propositions

10.3.1 and 10.3.2, respectively. The conclusion $\lim_{n\to\infty} \mathbb{P}\left[I(n;\theta_n) = 0\right] = 0$ is now immediate via the arguments provided in Chapter 5.1.

10.4 A preparatory result

Fix $n = 2, 3, \ldots$ and consider $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K such that $K < n$. Under the enforced assumptions, for all $i = 1, \ldots, n$, we easily see that

$$\mathbb{E}\left[\chi_{n,i}(\theta)\right] = \mathbb{E}\left[(1-p)^{D_{n,i}}\right] \tag{10.7}$$

where $D_{n,i}$ denotes the degree of node i in $\mathbb{H}(n; K)$. Note that

$$D_{n,i} \;=\; K + \sum_{j=1, j\notin\Gamma_{n,i}\cup\{i\}}^{n} \mathbf{1}\left[i \in \Gamma_{n,j}\right]. \tag{10.8}$$

By independence, since

$$\left|\{j = 1, \ldots, n:\ j \notin \Gamma_{n,i} \cup \{i\}\}\right| = n - K - 1,$$

the second term in (10.8) is a binomial rv with $n-K-1$ trials and success probability given by

$$\mathbb{P}\left[i \in \Gamma_{n,j}\right] = \frac{\binom{n-2}{K-1}}{\binom{n-1}{K}} = \frac{K}{n-1}, \tag{10.9}$$

whence

$$\mathbb{E}\left[\chi_{n,i}(\theta)\right] = (1-p)^K \cdot \left(1 - \frac{pK}{n-1}\right)^{n-K-1}. \tag{10.10}$$

The proof of Proposition 10.3.1 uses a somewhat simpler form of the expression (10.10) which we develop next.

Lemma 10.4.1 *Consider scalings $K : \mathbb{N}_0 \to \mathbb{N}_0$ and $p : \mathbb{N}_0 \to (0,1)$ such that (10.1) holds for some $c > 0$. We have*

$$n\mathbb{E}\left[\chi_{n,1}(\theta_n)\right] = e^{\beta_n + o(1)} \quad n = 1, 2, \ldots \tag{10.11}$$

with

$$\beta_n := (1 - c_n)\log n + K_n(p_n + \log(1 - p_n)) \tag{10.12}$$

where the sequence $c : \mathbb{N}_0 \to \mathbb{R}$ is the one appearing in the form (10.4) of the condition (10.1).

Proof. Consider a scaling $\theta : \mathbb{N}_0 \to \mathbb{N}_0 \times (0,1)$ such that (10.1) holds for some $c > 0$ and assume the existence of the limit $\lim_{n\to\infty} p_n = p^\star$. Replacing θ by θ_n in (10.10) for each $n = 2, 3, \ldots$ we get

$$n\mathbb{E}\left[\chi_{n,1}(\theta_n)\right] = e^{\gamma_n} \tag{10.13}$$

where γ_n is given by

$$\gamma_n = \log n + K_n \log(1 - p_n) - \eta_n$$

with

$$\eta_n := -(n - K_n - 1)\log\left(1 - \frac{p_n K_n}{n-1}\right.$$

The decomposition (5.16) now yields

$$\begin{aligned}
\eta_n &:= (n - K_n - 1)\left(\frac{p_n K_n}{n-1} + \Psi\left(\frac{p_n K_n}{n-1}\right)\right) \\
&= \left(1 - \frac{K_n}{n-1}\right) K_n p_n + (n - K_n - 1)\Psi\left(\frac{p_n K_n}{n-1}\right) \\
&= -K_n p_n + \left(2 - \frac{K_n}{n-1}\right) K_n p_n + (n - K_n - 1)\Psi\left(\frac{p_n K_n}{n-1}\right) \\
&= -K_n p_n + c_n \log n + (n - K_n - 1)\Psi\left(\frac{p_n K_n}{n-1}\right)
\end{aligned}$$

where the last step used the form (10.4) of the condition (10.1) on the scaling. Reporting this calculation into the expression for γ_n we find

$$\gamma_n = \beta_n - (n - K_n - 1)\Psi\left(\frac{p_n K_n}{n-1}\right).$$

Lemma 10.4.1 will be established if we show that

$$\lim_{n\to\infty} (n - K_n - 1)\Psi\left(\frac{p_n K_n}{n-1}\right) = 0. \tag{10.14}$$

To that end, for each $n = 2, 3, \ldots$ we note that

$$p_n K_n \leq p_n\left(2K_n - \frac{K_n^2}{n-1}\right) \leq 2p_n K_n$$

since $K_n < n$. The condition (10.4) implies

$$\frac{c_n}{2}\log n \leq p_n K_n \leq c_n \log n, \tag{10.15}$$

and it is now plain that

$$\lim_{n\to\infty} \frac{p_n K_n}{n-1} = 0 \quad \text{and} \quad \lim_{n\to\infty} (n - K_n - 1)\frac{p_n^2 K_n^2}{(n-1)^2} = 0.$$

Invoking the behavior (5.18) of $\Psi(x)$ at $x = 0$, we conclude from these facts that

$$\lim_{n\to\infty} \left((n - K_n - 1)\frac{p_n^2 K_n^2}{(n-1)^2} \right) \left(\frac{\Psi\left(\frac{p_n K_n}{n-1}\right)}{\left(\frac{p_n K_n}{n-1}\right)^2} \right) = 0. \tag{10.16}$$

This establishes (10.14) and the proof of Lemma 10.4.1 is completed. ■

10.5 A proof of Proposition 10.3.1

In view of Lemma 10.4.1, Proposition 10.3.1 will be established if we show

$$\lim_{n\to\infty} \beta_n = \begin{cases} -\infty & \text{if } \ c > \tau(p^\star) \\ \\ +\infty & \text{if } c < \tau(p^\star). \end{cases} \tag{10.17}$$

To see this, first note from (5.16) that for each $n = 1, 2, \ldots$, we have $p_n + \log(1 - p_n) \leq 0$ and the lower bound in (10.15) implies

$$\beta_n \ \leq \ (1 - c_n)\log n + c_n \left(\frac{\log n}{2p_n} \right) \cdot (p_n + \log(1 - p_n))$$

$$= \left(1 - \frac{c_n}{2}\left(1 - \frac{\log(1-p_n)}{p_n}\right)\right) \cdot \log n. \tag{10.18}$$

Letting n go to infinity in this last expression, we get $\lim_{n\to\infty} \beta_n = -\infty$ whenever

$$c > \lim_{n\to\infty} \frac{2}{1 - \frac{\log(1-p_n)}{p_n}} = \tau(p^\star) \tag{10.19}$$

since $\lim_{n\to\infty} c_n = c$.

Next, we show that if $c < \tau(p^\star)$, then $\lim_{n\to\infty} \beta_n = +\infty$. We only need to consider the case $0 \leq p^\star < 1$ since $\tau(1) = 0$ and the constraint $c < \tau(1)$ is vacuous. We begin by assuming $p^\star = 0$, in which case for each $n = 2, 3, \ldots$, we have

$$\begin{aligned}
\beta_n &= (1 - c_n)\log n + K_n(p_n + (-p_n - \Psi(p_n))) \\
&= (1 - c_n)\log n - K_n\Psi(p_n) \\
&= (1 - c_n)\log n - \left(\frac{\Psi(p_n)}{p_n^2}\right) \cdot K_n p_n^2 \\
&\geq (1 - c_n)\log n - c_n \log n \cdot \left(\frac{\Psi(p_n)}{p_n^2}\right) p_n \\
&= \log n \cdot \left(1 - c_n\left(1 + \left(\frac{\Psi(p_n)}{p_n^2}\right) p_n\right)\right)
\end{aligned} \tag{10.20}$$

with the inequality following from the upper bound in (10.15). Let n grow large in the last expression. Since we have assumed $\lim_{n\to\infty} p_n = 0$, we get

$$\lim_{n\to\infty} p_n \left(\frac{\Psi(p_n)}{p_n^2}\right) = 0,$$

and the desired conclusion $\lim_{n\to\infty} \beta_n = +\infty$ is obtained whenever $c < 1 = \tau(0)$ upon

using $\lim_{n\to\infty} c_n = c$.

Finally we assume $0 < p^\star < 1$. For each $\varepsilon > 0$, there exists a finite positive integer $n^\star(\varepsilon)$ such that $p_n \geq (1-\varepsilon)p^\star$ when $n \geq n^\star(\varepsilon)$. On that range the upper bound in (10.15) yields

$$K_n \leq \frac{c}{(1-\varepsilon)p^\star} \cdot \log n,$$

whence the conclusions $K_n^2 = o(n)$ and

$$p_n \left(2K_n - \frac{K_n^2}{n-1} \right) = 2K_n p_n + o(1)$$

follow. Comparing this last fact against the lefthand side of (10.4) yields

$$K_n p_n = \frac{c_n}{2} \log n + o(1),$$

so that

$$K_n p_n \sim \frac{c_n}{2} \log n. \tag{10.21}$$

From (10.12) it follows that

$$\frac{\beta_n}{\log n} = (1 - c_n) + \left(1 + \frac{\log(1-p_n)}{p_n} \right) \cdot \frac{K_n p_n}{\log n}$$

for all n sufficiently large. Letting n go to infinity in this last expression and using

(10.21) with the earlier remarks, we readily conclude

$$\lim_{n\to\infty} \frac{\beta_n}{\log n} = (1-c) + \frac{c}{2}\left(1 + \frac{\log(1-p^\star)}{p^\star}\right) = 1 - \frac{c}{\tau(p^\star)}$$

where the last step follows by direct inspection. It is now clear that $\lim_{n\to\infty} \beta_n = \infty$ when $c < \tau(p^\star)$ with $0 < p^\star < 1$. This establishes (10.17) and the proof of Proposition 10.3.1 is now completed. ■

10.6 Negative dependence and consequences

Fix positive integers $n = 2, 3, \ldots$ and K with $K < n$. Several properties of the $\{0,1\}$-valued rvs

$$\left\{ \mathbf{1}\left[j \in \Gamma_{n,i}\right], \quad \begin{array}{c} i \neq j \\ i,j = 1,\ldots,n \end{array} \right\} \tag{10.22}$$

and

$$\left\{ \mathbf{1}\left[j \in \Gamma_{n,i} \ \vee \ i \in \Gamma_{n,j}\right], \quad \begin{array}{c} i \neq j \\ i,j = 1,\ldots,n \end{array} \right\} \tag{10.23}$$

will play a key role in some of the forthcoming arguments.

10.6.1 Negative association

The properties of interest can be couched in terms of *negative association*, a form of negative correlation introduced to Joag-Dev and Proschan [23]. We first develop the needed definitions and properties: Let $\{X_\lambda,\ \lambda \in \Lambda\}$ be a collection of $\mathbb{R}$-valued rvs indexed by the finite set Λ. For any non-empty subset A of Λ, we write X_A to denote the $\mathbb{R}^{|A|}$-valued $X_A = (X_\lambda,\ \lambda \in A)$. The rvs $\{X_\lambda,\ \lambda \in \Lambda\}$ are then said to be *negatively associated* if for any non-overlapping subsets A and B of Λ and for any monotone increasing mappings $\varphi : \mathbb{R}^{|A|} \to \mathbb{R}$ and $\psi : \mathbb{R}^{|B|} \to \mathbb{R}$, the covariance inequality

$$\mathbb{E}\left[\varphi(X_A)\psi(X_B)\right] \leq \mathbb{E}\left[\varphi(X_A)\right]\mathbb{E}\left[\psi(X_B)\right] \tag{10.24}$$

holds whenever the expectations in (10.24) are well defined and finite. Note that φ and ψ need only be monotone increasing on the support of X_A and X_B, respectively.

This definition has some easy consequences to be used repeatedly in what follows: The negative association of $\{X_\lambda,\ \lambda \in \Lambda\}$ implies the negative association of the collection $\{X_\lambda,\ \lambda \in \Lambda'\}$ where Λ' is any subset of Λ. It is also well known [23, P2, p. 288] that the negative association of the rvs $\{X_\lambda,\ \lambda \in \Lambda\}$ implies the inequality

$$\mathbb{E}\left[\prod_{\lambda \in A} f_\lambda(X_\lambda)\right] \leq \prod_{\lambda \in A} \mathbb{E}\left[f_\lambda(X_\lambda)\right] \tag{10.25}$$

where A is a subset of Λ and the collection $\{f_\lambda,\ \lambda \in A\}$ of mappings $\mathbb{R} \to \mathbb{R}_+$ are all monotone increasing; by non-negativity all the expectations exist and finiteness is

moot.

We can apply these ideas to collections of indicator rvs, namely for each λ in Λ, $X_\lambda = \mathbf{1}[E_\lambda]$ for some event E_λ. From the definitions, it is easy to see that if the rvs $\{\mathbf{1}[E_\lambda],\ \lambda \in \Lambda\}$ are negatively associated, so are the rvs $\{\mathbf{1}[E_\lambda^c],\ \lambda \in \Lambda\}$. Moreover, for any subset A of Λ, we have

$$\mathbb{P}[E_\lambda,\ \lambda \in A] \leq \prod_{\lambda \in A} \mathbb{P}[E_\lambda]. \tag{10.26}$$

This follows from (10.25) by taking $f_\lambda(x) = x^+$ on $\mathbb{R}$ for each λ in Λ.

10.6.2 Useful consequences

A key observation for our purpose is as follows: For each $i = 1, \ldots, n$, the rvs

$$\{\mathbf{1}[j \in \Gamma_{n,i}],\ j \in \mathcal{N}_{-i}\} \tag{10.27}$$

form a collection of negatively associated rvs. This is a consequence of the fact that the random set $\Gamma_{n,i}$ represents a random sample (without replacement) of size K from $\mathcal{N}_{-i}$; see [23, Example 3.2(c)] for details.

The n collections (10.27) are mutually independent, so that by the "closure under products" property of negative association [23, P7, p. 288] [12, p. 35], the rvs (10.22) also form a collection of negatively associated rvs.

Hence, by taking complements, the rvs

$$\left\{ \mathbf{1}\left[j \notin \Gamma_{n,i}\right], \quad \begin{array}{c} i \neq j \\ i,j = 1,\ldots,n \end{array} \right\} \tag{10.28}$$

also form a collection of negatively associated rvs. With distinct $i, j = 1, \ldots, n$, we note that

$$\mathbf{1}\left[i \notin \Gamma_{n,j}, j \notin \Gamma_{n,i}\right] = f\left(\mathbf{1}\left[i \notin \Gamma_{n,j}\right], \mathbf{1}\left[j \notin \Gamma_{n,i}\right]\right) \tag{10.29}$$

with mapping $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x^+ y^+$ for all x, y in $\mathbb{R}$. This mapping being non-decreasing on $\mathbb{R}^2$, it follows [23, P6, p. 288] that the rvs

$$\left\{ \mathbf{1}\left[j \notin \Gamma_{n,i}, i \notin \Gamma_{n,j}\right], \quad \begin{array}{c} i \neq j \\ i,j = 1,\ldots,n \end{array} \right\} \tag{10.30}$$

are also negatively associated. Taking complements one more time, we see that the rvs (10.23) are also negatively associated.

For each $k = 1, 2$ and $j = 3, \ldots, n$, we shall find it useful to define

$$u_{n,j,k}(\theta) := \mathbb{E}\left[(1-p)^{\mathbf{1}[k \in \Gamma_{n,j}]}\right]$$

and

$$b_{n,j}(\theta) := \mathbb{E}\left[(1-p)^{\mathbf{1}[1 \in \Gamma_{n,j}] + \mathbf{1}[2 \in \Gamma_{n,j}]}\right].$$

Under the enforced assumptions, we have $b_{n,3}(\theta) = \ldots = b_{n,n}(\theta) \equiv b_n(\theta)$ and $u_{n,3,1}(\theta) = \ldots = u_{n,n,1}(\theta) = u_{n,3,2}(\theta) = \ldots = u_{n,n,2}(\theta) \equiv u_n(\theta)$.

Before computing either one of the quantities $u_n(\theta)$ and $b_n(\theta)$, we note that

$$b_n(\theta) \leq u_n(\theta)^2. \tag{10.31}$$

This is a straightforward consequence of the negative association of the rvs (10.22) – In (10.24), with A and B singletons, use the increasing functions $\varphi, \psi : \mathbb{R} \to \mathbb{R} : x \to -(1-p)^x$.

Using (10.9) we get

$$u_n(\theta) = (1-p)\frac{K}{n-1} + \left(1 - \frac{K}{n-1}\right) = 1 - p\frac{K}{n-1}. \tag{10.32}$$

An expression for $b_n(\theta)$ is available but will not be needed due to the availability of (10.31).

10.7 A proof of Proposition 10.3.2

As expected, the first step in proving Proposition 10.3.2 consists in evaluating the cross moment appearing in the numerator of (10.6). Fix $n = 2, 3, \ldots$ and consider $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K such that $K < n$. Define the

$\mathbb{N}_0$-valued rvs $B_n(\theta)$ and $U_n(\theta)$ by

$$B_n(\theta) := \sum_{j=3}^{n} \mathbf{1}\left[j \notin \Gamma_{n,1}\right] \mathbf{1}\left[j \notin \Gamma_{n,2}\right] \tag{10.33}$$

and

$$U_n(\theta) \;:=\; \sum_{j=3}^{n} \mathbf{1}\left[j \notin \Gamma_{n,1}\right] \mathbf{1}\left[j \in \Gamma_{n,2}\right] + \sum_{j=3}^{n} \mathbf{1}\left[j \notin \Gamma_{n,2}\right] \mathbf{1}\left[j \in \Gamma_{n,1}\right]. \tag{10.34}$$

Proposition 10.7.1 *Fix $n = 2, 3, \ldots$. For any p in $(0, 1)$ and positive integer K such that $K < n$, we have*

$$\mathbb{E}\left[\chi_{n,1}(\theta)\chi_{n,2}(\theta)\right] = (1-p)^{2K}\mathbb{E}\left[\frac{b_n(\theta)^{B_n(\theta)} \cdot u_n(\theta)^{U_n(\theta)}}{(1-p)^{\mathbf{1}[2\in\Gamma_{n,1}, 1\in\Gamma_{n,2}]}}\right] \tag{10.35}$$

where the rvs $B_n(\theta)$ and $U_n(\theta)$ given by (10.33) and (10.34), respectively.

A proof of Proposition 10.7.1 is available in Section 10.11. Still in the setting of Proposition 10.7.1, we can use (10.31) in conjunction with (10.35) to get

$$\mathbb{E}\left[\chi_{n,1}(\theta)\chi_{n,2}(\theta)\right] \leq (1-p)^{2K}\mathbb{E}\left[\frac{u_n(\theta)^{2B_n(\theta)+U_n(\theta)}}{(1-p)^{\mathbf{1}[2\in\Gamma_{n,1}, 1\in\Gamma_{n,2}]}}\right]. \tag{10.36}$$

It is plain that

$$2B_n(\theta) + U_n(\theta) = \sum_{j=3}^{n} \mathbf{1}\left[j \notin \Gamma_{n,1}\right] + \sum_{j=3}^{n} \mathbf{1}\left[j \notin \Gamma_{n,2}\right].$$

We note that

$$\begin{aligned}\sum_{j=3}^{n} \mathbf{1}\left[j \notin \Gamma_{n,1}\right] &= \sum_{j=2}^{n} \mathbf{1}\left[j \notin \Gamma_{n,1}\right] - \mathbf{1}\left[2 \notin \Gamma_{n,1}\right] \\ &= (n-1-K) - (1 - \mathbf{1}\left[2 \in \Gamma_{n,1}\right]) \\ &= (n-2-K) + \mathbf{1}\left[2 \in \Gamma_{n,1}\right]\end{aligned}$$

and

$$\sum_{j=3}^{n} \mathbf{1}\left[j \notin \Gamma_{n,2}\right] = (n-2-K) + \mathbf{1}\left[1 \in \Gamma_{n,2}\right]$$

by similar arguments. The expression

$$2B_n(\theta) + U_n(\theta) = 2(n-2-K) + \mathbf{1}\left[2 \in \Gamma_{n,1}\right] + \mathbf{1}\left[1 \in \Gamma_{n,2}\right]$$

now follows, and we find

$$\mathbb{E}\left[\chi_{n,1}(\theta)\chi_{n,2}(\theta)\right] \leq (1-p)^{2K} u_n(\theta)^{2(n-2-K)} \cdot R_n(\theta) \tag{10.37}$$

with

$$R_n(\theta) := \mathbb{E}\left[\frac{u_n(\theta)^{\mathbf{1}[2\in\Gamma_{n,1}]+\mathbf{1}[1\in\Gamma_{n,2}]}}{(1-p)^{\mathbf{1}[2\in\Gamma_{n,1},1\in\Gamma_{n,2}]}}\right].$$

Next, with the help of (10.10) and (10.32) we conclude that

$$\frac{\mathbb{E}\left[\chi_{n,1}(\theta)\chi_{n,2}(\theta)\right]}{\left(\mathbb{E}\left[\chi_{n,1}(\theta)\right]\right)^2} \leq \frac{(1-p)^{2K} \cdot u_n(\theta)^{2(n-2-K)}}{\left((1-p)^K \cdot u_n(\theta)^{n-1-K}\right)^2} \cdot R_n(\theta)$$

$$
\begin{aligned}
&= u_n(\theta)^{-2} R_n(\theta) \\
&= \mathbb{E}\left[\frac{u_n(\theta)^{\mathbf{1}[2\in\Gamma_{n,1}]+\mathbf{1}[1\in\Gamma_{n,2}]-2}}{(1-p)^{\mathbf{1}[2\in\Gamma_{n,1},1\in\Gamma_{n,2}]}}\right]. \qquad (10.38)
\end{aligned}
$$

Direct inspection readily yields

$$
\frac{u_n(\theta)^{\mathbf{1}[2\in\Gamma_{n,1}]+\mathbf{1}[1\in\Gamma_{n,2}]-2}}{(1-p)^{\mathbf{1}[2\in\Gamma_{n,1},1\in\Gamma_{n,2}]}} = \begin{cases} \frac{1}{1-p} & \text{if } \ 2\in\Gamma_{n,1}, 1\in\Gamma_{n,2} \\ \left(1-\frac{pK}{n-1}\right)^{-2} & \text{if } 2\notin\Gamma_{n,1}, 1\notin\Gamma_{n,2} \\ \left(1-\frac{pK}{n-1}\right)^{-1} & \text{otherwise.} \end{cases} \qquad (10.39)
$$

Taking expectation and reporting into (10.38) we then find

$$
\begin{aligned}
\frac{\mathbb{E}\left[\chi_{n,1}(\theta)\chi_{n,2}(\theta)\right]}{\left(\mathbb{E}\left[\chi_{n,1}(\theta)\right]\right)^2} &\leq \frac{1}{1-p}\mathbb{P}\left[2\in\Gamma_{n,1}, 1\in\Gamma_{n,2}\right] + \left(1-p\frac{K}{n-1}\right)^{-2} \\
&= \frac{1}{1-p}\left(\frac{K}{n-1}\right)^2 + \left(1-p\frac{K}{n-1}\right)^{-2} \qquad (10.40)
\end{aligned}
$$

by a crude bounding argument.

Now consider a scaling $\theta : \mathbb{N}_0 \to \mathbb{N}_0 \times (0,1)$ such that (10.1) holds for some $c > 0$ and $\lim_{n\to\infty} p_n = p^\star < 1$. Replace θ by θ_n in the bound (10.40) with respect to this scaling. It is immediate that (10.6) will be established if we show that

$$
\lim_{n\to\infty} \frac{1}{1-p_n}\left(\frac{K_n}{n-1}\right)^2 = 0
$$

and that

$$\lim_{n\to\infty}\left(1-p_n\frac{K_n}{n-1}\right)=1.$$

These limits are an easy consequence of the inequalities (10.15) by virtue of the fact that $\lim_{n\to\infty} p_n = p^\star < 1$. ■

We close with a proof of (3.45): Consider $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K. It follows from (10.10) that

$$\lim_{n\to\infty}\mathbb{E}\left[\chi_{n,1}(\theta)\right]=(1-p)^K\,e^{-pK},$$

whence $\lim_{n\to\infty}\mathbb{E}\left[I(n;\theta)\right]=\infty$. It also immediate from (10.40) that

$$\limsup_{n\to\infty}\frac{\mathbb{E}\left[\chi_{n,1}(\theta)\chi_{n,2}(\theta)\right]}{\left(\mathbb{E}\left[\chi_{n,1}(\theta)\right]\right)^2}\le 1.$$

The arguments outlined in Section 10.3 now yield

$$\lim_{n\to\infty}\mathbb{P}\left[I(n;\theta)=0\right]=0,$$

and this establishes (3.45). The conclusion (3.46) immediately follows; see discussion at (10.41).

10.8 A proof of Theorem 10.2.2 (Part I)

Fix $n = 2, 3, \ldots$ and consider $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K such that $K < n$. As expected, we define the events

$$C_n(\theta) := [\mathbb{H} \cap \mathbb{G}(n; \theta) \text{ is connected}]$$

and

$$I_n(\theta) := [\mathbb{H} \cap \mathbb{G}(n; \theta) \text{ contains no isolated nodes}] .$$

If the random graph $\mathbb{H} \cap \mathbb{G}(n; \theta)$ is connected, then it does not contain any isolated node, whence $C_n(\theta)$ is a subset of $I_n(\theta)$, and the conclusions

$$\mathbb{P}[C_n(\theta)] \leq \mathbb{P}[I_n(\theta)] \tag{10.41}$$

and

$$\mathbb{P}[C_n(\theta)^c] = \mathbb{P}[C_n(\theta)^c \cap I_n(\theta)] + \mathbb{P}[I_n(\theta)^c] \tag{10.42}$$

obtain.

Taken together with Theorem 10.2.1, the relations (10.41) and (10.42) pave the way to proving Theorem 10.2.2. Indeed, pick a scaling $\theta : \mathbb{N}_0 \to \mathbb{N}_0 \times (0, 1)$ such that (10.1) holds for some $c > 0$ and $\lim_{n\to\infty} p_n = p^\star$ exists. If $c < \tau(p^\star)$, then $\lim_{n\to\infty} \mathbb{P}[I_n(\theta_n)] = 0$ by the zero-law for the absence of isolated nodes in Theorem 10.2.1, whence $\lim_{n\to\infty} \mathbb{P}[C_n(\theta_n)] = 0$ with the help of (10.41). If $c > \tau(p^\star)$, then

$\lim_{n\to\infty} \mathbb{P}\left[I_n(\theta_n)\right] = 1$ by the one-law for the absence of isolated nodes, and the desired conclusion $\lim_{n\to\infty} \mathbb{P}\left[C_n(\theta_n)\right] = 1$ (or equivalently, $\lim_{n\to\infty} \mathbb{P}\left[C_n(\theta_n)^c\right] = 0$) will follow via (10.42) if we show the following:

Proposition 10.8.1 *For any scaling $\theta : \mathbb{N}_0 \to \mathbb{N}_0 \times (0,1)$ such that $\lim_{n\to\infty} p_n = p^\star$ exists and (10.1) holds for some $c > \tau(p^\star)$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left[C_n(\theta_n)^c \cap I_n(\theta_n)\right] = 0. \tag{10.43}$$

The proof of Proposition 10.8.1 starts below and runs through two more sections, namely Sections 10.9 and 10.10. The basic idea is to find a sufficiently tight upper bound on the probability in (10.43), and then to show that this bound goes to zero as n becomes large. This approach is similar to the one used for proving the one-law for connectivity in ER graphs [4, p. 164], and in random key graphs; see Chapter 7.

We begin by finding the needed upper bound. Indeed, as in Chapter 7.4.3, we will use an appropriate version of the union bound (5.14) (established in Chapter 5.2): Fix $n = 2, 3, \ldots$ and consider $\theta = (K, p)$ with p in $(0,1)$ and positive integer K such that $K < n$. Consider the definitions given in Chapter 5.2 with $\mathbb{G}(n;\upsilon)$ replaced by $\mathbb{H} \cap \mathbb{G}(n;\theta)$, m_n replaced by n, I_n replaced by $I_n(\theta)$ and for each $r = 1, \ldots, n$, $A_{n,r}$, $B_{n,r}$, $C_{n,r}$ replaced by $A_{n,r}(\theta)$, $B_{n,r}(\theta)$. $C_{n,r}(\theta)$, respectively. For $r = n$ we use (with a slight abuse of notation) $C_n(\theta)$ as defined above. The arguments of Chapter 5.2

now lead to the key bound:

$$\mathbb{P}\left[C_n(\theta)^c \cap I_n(\theta)\right] \leq \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta)\right] \tag{10.44}$$

upon using (5.14) with the aforementioned substitutions.

Now, consider a scaling $\theta : \mathbb{N}_0 \to \mathbb{N}_0 \times (0,1)$ as in the statement of Proposition 10.8.1. Substitute θ by θ_n by means of this scaling in the right hand side of (10.44). The proof of Proposition 10.8.1 will be completed once we show

$$\lim_{n\to\infty} \sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] = 0. \tag{10.45}$$

The means to do so are provided in the next section.

10.9 Bounding probabilities

Fix $n = 2, 3, \ldots$ and consider $\theta = (K, p)$ with p in $(0,1)$ and positive integer K such that $K < n$.

10.9.1 Bounding the probabilities $\mathbb{P}\left[B_{n,r}(\theta)\right]$

The following result will be used to efficiently bound the probability $\mathbb{P}\left[B_{n,r}(\theta)\right]$.

Lemma 10.9.1 *For each $r = 2, \ldots, n-1$, we have the inequality*

$$\mathbb{P}\left[B_{n,r}(\theta) \;\middle|\; \Gamma_{n,1}, \ldots, \Gamma_{n,r}\right] \leq (1-p)^{E^{\star}_{n,r}} \cdot u_n(\theta)^{r(n-r)-E^{\star}_{n,r}} \tag{10.46}$$

with $u_n(\theta)$ defined by (10.32) and the rv $E^{\star}_{n,r}$ given by

$$E^{\star}_{n,r} := \sum_{i=r+1}^{n} \sum_{\ell=1}^{r} \mathbf{1}\left[\ell \in \Gamma_{n,i}\right]. \tag{10.47}$$

A proof of Lemma 10.9.1 is available in Appendix 10.12. The rv $E^{\star}_{n,r}$, which appears prominently in (10.46), has a tail controlled through the following result.

Lemma 10.9.2 *Fix $r = 2, \ldots, n-1$. For any t in $(0,1)$ we have*

$$\mathbb{P}\left[E^{\star}_{n,r} \leq (1-t) rK \cdot \frac{n-r}{n-1}\right] \leq e^{-\frac{t^2}{2} rK \cdot \frac{n-r}{n-1}}. \tag{10.48}$$

Proof. Fix $n = 2, 3, \ldots$ and consider a positive integer K such that $K < n$. From the facts reported in Section 10.6, the negative association of the rvs (10.27) implies that of the rvs $\{\mathbf{1}\left[\ell \in \Gamma_{n,i}\right],\ i = r+1, \ldots, n;\ \ell = 1, \ldots, r\}$. We are now in a position to apply the Chernoff-Hoeffding bound to the sum (10.47). We use the bound in the

form

$$\mathbb{P}\left[E^{\star}_{n,r} \leq (1-t)\mathbb{E}\left[E^{\star}_{n,r}\right]\right] \leq e^{-\frac{t^2}{2}\mathbb{E}\left[E^{\star}_{n,r}\right]} \tag{10.49}$$

as given for negatively associated rvs in [12, Thm. 1.1, p. 6]. The conclusion (10.48) follows upon noting that

$$\mathbb{E}\left[E^{\star}_{n,r}\right] = \sum_{i=r+1}^{n}\sum_{\ell=1}^{r}\mathbb{P}\left[\ell \in \Gamma_{n,i}\right] = r(n-r)\frac{K}{n-1}$$

as we use (10.9). ■

10.9.2 Bounding the probabilities $\mathbb{P}\left[C_{n,r}(\theta)\right]$

For each $r = 2, \ldots, n$, let $\mathbb{H} \cap \mathbb{G}_r(n;\theta)$ stand for the subgraph $\mathbb{H} \cap \mathbb{G}(n;\theta)(S)$ when $S = \{1, \ldots, r\}$. Also let $\mathcal{T}_r$ denote the collection of all spanning trees on the vertex set $\{1, \ldots, r\}$.

Lemma 10.9.3 *Fix $r = 2, \ldots, n$. For each T in $\mathcal{T}_r$, we have*

$$\mathbb{P}\left[T \subset \mathbb{H} \cap \mathbb{G}_r(n;\theta)\right] \leq \left(p\lambda_n(K)\right)^{r-1} \tag{10.50}$$

where the notation $T \subset \mathbb{H} \cap \mathbb{G}_r(n;\theta)$ indicates that the tree T is a subgraph spanning $\mathbb{H} \cap \mathbb{G}_r(n;\theta)$.

Since $p\lambda_n(K)$ is the probability of link assignment, the situation is reminiscent to the one found in ER graphs [4] and random key graphs (see Lemma 7.4.6) where in each case the bound (10.50) holds with equality.

Proof. Fix $r = 2, 3, \ldots, n$ and pick a tree T in $\mathcal{T}_r$. Let $\mathcal{E}(T)$ be the set of edges that appear in T. It is plain that $T \subseteq \mathbb{H} \cap \mathbb{G}_r(n, ; \theta)$ occurs if and only if the set of conditions

$$\Sigma_{n,i} \cap \Sigma_{n,j} \neq \emptyset \quad \text{and} \quad B_{ij}(p) = 1, \qquad \{i, j\} \in \mathcal{E}(T)$$

holds. Therefore, under the enforced independence assumptions, since $|\mathcal{E}(T)| = r - 1$, we get

$$\begin{aligned}
&\mathbb{P}\left[T \subset \mathbb{H} \cap \mathbb{G}_r(n; \theta)\right] \\
&= \quad p^{r-1} \cdot \mathbb{E}\left[\prod_{i,j:\{i,j\}\in\mathcal{E}(T)} \mathbf{1}\left[\Sigma_{n,i} \cap \Sigma_{n,j} \neq \emptyset\right]\right] \\
&= \quad p^{r-1} \cdot \mathbb{E}\left[\prod_{i,j:\{i,j\}\in\mathcal{E}(T)} \mathbf{1}\left[i \in \Gamma_{n,j} \ \vee \ j \in \Gamma_{n,i}\right]\right] \\
&\leq \quad p^{r-1} \cdot \prod_{i,j:\{i,j\}\in\mathcal{E}(T)} \mathbb{P}\left[i \in \Gamma_{n,j} \ \vee \ j \in \Gamma_{n,i}\right] \qquad (10.51)
\end{aligned}$$

by making use of (10.26) with the negatively associated rvs (10.23). The desired result (10.50) is now immediate from (3.6) and the relation $|\mathcal{E}(T)| = r - 1$. ■

As in the case of ER graphs [4] and random key graphs [37] we have the following bound.

Lemma 10.9.4 *For each $r = 2, \ldots, n$, we have*

$$\mathbb{P}\left[C_{n,r}(\theta)\right] \leq r^{r-2} \left(p\lambda_n(K)\right)^{r-1}. \tag{10.52}$$

Proof. Fix $r = 2, \ldots, n$. If $\mathbb{H} \cap \mathbb{G}_r(n;\theta)$ is a connected graph, then it must contain a spanning tree on the vertex set $\{1, \ldots .r\}$, and a union bound argument yields

$$\mathbb{P}\left[C_{n,r}(\theta)\right] \leq \sum_{T \in \mathcal{T}_r} \mathbb{P}\left[T \subset \mathbb{H} \cap \mathbb{G}(n;\theta)(S)\right]. \tag{10.53}$$

By Cayley's formula [25] there are r^{r-2} trees on r vertices, i.e., $|\mathcal{T}_r| = r^{r-2}$, and (10.52) follows upon making use of (10.50). ■

10.10 A proof of Proposition 10.8.1 (Part II)

Consider a scaling $\theta : \mathbb{N}_0 \to \mathbb{N}_0 \times (0,1)$ as in the statement of Proposition 10.8.1. Pick integers $R \geq 2$ and $n^\star(R) \geq 2(R+1)$ (to be specified in Section 10.10.2). On

the range $n \geq n^\star(R)$ we consider the decomposition

$$\sum_{r=2}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] \;=\; \sum_{r=2}^{R} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] + \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right],$$

and let n go to infinity. The desired convergence (10.45) will be established if we show

$$\lim_{n\to\infty} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] = 0 \tag{10.54}$$

for *each* $r = 2, 3, \ldots$ and

$$\lim_{n\to\infty} \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] = 0. \tag{10.55}$$

We establish (10.54) and (10.55) in turn.

10.10.1 Establishing (10.54)

Fix $r = 2, 3, \ldots$ and consider $n = 2, 3, \ldots$ such that $r < n$. Also let $\theta = (K, p)$ with p in $(0, 1)$ and positive integer K such that $K < n$. With (10.47) in mind, for each $i = 1, \ldots, r$, we note that

$$\begin{aligned} \sum_{k=r+1}^{n} \mathbf{1}\left[k \in \Gamma_{n,i}\right] &= \sum_{k=1}^{n} \mathbf{1}\left[k \in \Gamma_{n,i}\right] - \sum_{k=1}^{r} \mathbf{1}\left[k \in \Gamma_{n,i}\right] \\ &= K - \sum_{k=1}^{r} \mathbf{1}\left[k \in \Gamma_{n,i}\right] \end{aligned} \tag{10.56}$$

since $|\Gamma_{n,i}| = K$. The bounds

$$(K-r)^+ \leq \sum_{k=r+1}^{n} \mathbf{1}\left[k \in \Gamma_{n,i}\right] \leq K$$

follow, whence

$$r(K-r)^+ \leq E^{\star}_{n,r} \leq rK.$$

It is also the case that

$$r(n-r-K)^+ \leq r(n-r) - E^{\star}_{n,r}.$$

Reporting these lower bounds into (10.46), we get

$$\begin{aligned} \mathbb{P}\left[B_{n,r}(\theta) \;\middle|\; \Gamma_{n,1}, \ldots, \Gamma_{n,r}\right] &\leq (1-p)^{r(K-r)^+} \cdot u_n(\theta)^{r(n-r-K)^+} \\ &\leq (1-p)^{r(K-r)} \cdot u_n(\theta)^{r(n-r-K)} \end{aligned} \tag{10.57}$$

since $0 < p, u_n(\theta) < 1$. If we set

$$F_{n,r}(\theta) := (1-p)^{(K-r)} \cdot u_n(\theta)^{(n-r-K)},$$

it is now plain that

$$\begin{aligned} \mathbb{P}\left[A_{n,r}(\theta)\right] &= \mathbb{E}\left[\mathbf{1}\left[C_{n,r}(\theta)\right] \mathbb{P}\left[B_{n,r}(\theta) \;\middle|\; \Gamma_{n,1}, \ldots, \Gamma_{n,r}\right]\right] \\ &\leq \mathbb{P}\left[C_{n,r}(\theta)\right] \cdot F_{n,r}(\theta)^r. \end{aligned} \tag{10.58}$$

Applying Lemma 10.9.4 we find

$$\begin{aligned}\binom{n}{r}\mathbb{P}\left[A_{n,r}(\theta)\right] &\leq \binom{n}{r}\mathbb{P}\left[C_{n,r}(\theta)\right]\cdot F_{n,r}(\theta)^r \\ &\leq \left(\frac{en}{r}\right)^r r^{r-2}\left(p\lambda_n(K)\right)^{r-1}F_{n,r}(\theta)^r \\ &= \frac{1}{r^2}\left(en\right)^r\left(p\lambda_n(K)\right)^{r-1}F_{n,r}(\theta)^r \qquad (10.59)\end{aligned}$$

as we make use of (5.25).

We also note that

$$F_{n,r}(\theta) \leq e^{F^\star_{n,r}(\theta)} \qquad (10.60)$$

with

$$\begin{aligned}F^\star_{n,r}(\theta) &:= (K-r)\log(1-p) - (n-r-K)p\frac{K}{n-1} \\ &= (K-r)\log(1-p) - \left(1-\frac{K}{n-1}-\frac{r-1}{n-1}\right)pK \\ &= (K-r)\log(1-p) - p\left(K-\frac{K^2}{n-1}\right)+\frac{r-1}{n-1}pK \\ &= K\left(p+\log(1-p)\right) - r\log(1-p) - p\left(2K-\frac{K^2}{n-1}\right) \\ &\quad +\frac{r-1}{n-1}pK. \qquad (10.61)\end{aligned}$$

Now, pick any given positive integer $r = 2, 3, \ldots$ and consider a scaling $\theta : \mathbb{N}_0 \to \mathbb{N}_0 \times (0,1)$ such that $\lim_{n\to\infty} p_n = p^\star$ exists and (10.1) holds for some $c > \tau(p^\star)$. Replace θ by θ_n in (10.59) according to this scaling. In order to establish (10.54) it

suffices to show that

$$\lim_{n\to\infty} (en)^r \left(p_n \lambda_n(K_n)\right)^{r-1} \cdot F_{n,r}(\theta_n)^r = 0. \tag{10.62}$$

For n sufficiently large, from (10.4) and (10.59) we first get

$$\begin{aligned}
\binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta)\right] &\leq (en)^r \left(p_n \lambda_n(K_n)\right)^{r-1} \cdot F_{n,r}(\theta_n)^r \\
&= (en)^r \left(c_n \frac{\log n}{n-1}\right)^{r-1} \cdot F_{n,r}(\theta_n)^r \\
&= en \left(ec_n \frac{n}{n-1} \log n\right)^{r-1} \cdot F_{n,r}(\theta_n)^r. \qquad (10.63)
\end{aligned}$$

On the other hand, upon making use of the bounds at (10.15), we find

$$\begin{aligned}
F^{\star}_{n,r}(\theta_n) &\leq K_n \left(p_n + \log(1-p_n)\right) - r\log(1-p_n) \\
&\quad - p_n \left(2K_n - \frac{K_n^2}{n-1}\right) + \frac{r}{n}\, p_n K_n \\
&= K_n \left(p_n + \log(1-p_n)\right) - r\log(1-p_n) \\
&\quad - c_n \log n + \frac{r}{n}\, p_n K_n \\
&\leq K_n \left(p_n + \log(1-p_n)\right) - c_n \log n \\
&\quad - r\log(1-p_n) + \frac{r}{n}\, c_n \log n \\
&= p_n K_n \left(1 + \frac{\log(1-p_n)}{p_n}\right) - c_n \log n \\
&\quad - r\log(1-p_n) + \frac{r}{n}\, c_n \log n \\
&\leq \frac{c_n}{2} \log n \cdot \left(1 + \frac{\log(1-p_n)}{p_n}\right) - c_n \log n
\end{aligned}$$

$$
\begin{aligned}
& - r\log(1-p_n) + \frac{r}{n}\, c_n \log n \\
= \;& -\frac{c_n}{2} \cdot \left(1 - \frac{\log(1-p_n)}{p_n}\right) \log n \\
& - r\log(1-p_n) + \frac{r}{n}\, c_n \log n. \\
= \;& \log n \left(-\frac{c_n - \frac{2rp_n}{\log n}}{2} \left(1 - \frac{\log(1-p_n)}{p_n}\right)\right) \\
& - rp_n + \frac{r}{n}\, c_n \log n \\
\leq \;& -\frac{\log n}{2} \left(c_n - \frac{2rp_n}{\log n}\right) \left(1 - \frac{\log(1-p_n)}{p_n}\right) \\
& + \frac{r}{n}\, c_n \log n. \qquad (10.64)
\end{aligned}
$$

As a result, (10.61) implies

$$
nF_{n,r}(\theta_n)^r \;\leq\; n^{1-\frac{r}{2}\left(c_n - \frac{2rp_n}{\log n}\right)\cdot\left(1-\frac{\log(1-p_n)}{p_n}\right)} e^{o(1)}. \qquad (10.65)
$$

Under the enforced assumptions of Theorem 10.2.2 we get

$$
\begin{aligned}
& \lim_{n\to\infty} \left(1 - \frac{r}{2}\left(c_n - \frac{2rp_n}{\log n}\right) \cdot \left(1 - \frac{\log(1-p_n)}{p_n}\right)\right) \\
= \;& 1 - r\frac{c}{2} \cdot \left(1 - \frac{\log(1-p^\star)}{p^\star}\right) \\
= \;& 1 - r\frac{c}{\tau(p^\star)} < 0, \qquad (10.66)
\end{aligned}
$$

and the desired conclusion (10.62) follows upon making use of the inequalities (10.63) and (10.65).

10.10.2 Establishing (10.55)

Fix $n = 2, 3, \ldots$ and consider $\theta = (K, p)$ with p in $(0, 1)$, and positive integer K such that $K < n$.

Pick $r = 1, 2, \ldots, n-1$. By Lemma 10.9.1 we conclude that

$$\mathbb{P}\left[B_{n,r}(\theta) \;\middle|\; \Gamma_{n,1}, \ldots, \Gamma_{n,r}\right] \leq (1-p)^{E^{\star}_{n,r}} \tag{10.67}$$

since $0 < u_n(\theta) < 1$, and preconditioning arguments similar to the ones leading to (10.58) yield

$$\mathbb{P}\left[A_{n,r}(\theta)\right] \leq \mathbb{E}\left[\mathbf{1}\left[C_{n,r}(\theta)\right](1-p)^{E^{\star}_{n,r}}\right].$$

The event $C_{n,r}(\theta)$ depends only on $\Gamma_{n,1}, \ldots, \Gamma_{n,r}$ whereas $E^{\star}_{n,r}$ is determined solely by $\Gamma_{n,r+1}, \ldots, \Gamma_{n,n}$. Thus, the event $C_{n,r}(\theta)$ is independent of the rv $(1-p)^{E^{\star}_{n,r}}$ under the enforced assumptions, whence

$$\mathbb{P}\left[A_{n,r}(\theta)\right] \leq \mathbb{P}\left[C_{n,r}(\theta)\right]\mathbb{E}\left[(1-p)^{E^{\star}_{n,r}}\right]. \tag{10.68}$$

Pick t arbitrary in $(0, 1)$ and recall Lemma 10.9.2. A simple decomposition argument shows that

$$\begin{aligned}\mathbb{E}\left[(1-p)^{E^{\star}_{n,r}}\right] \;\leq\; & \mathbb{E}\left[(1-p)^{E^{\star}_{n,r}}\,\mathbf{1}\left[E^{\star}_{n,r} > (1-t)rK\cdot\frac{n-r}{n-1}\right]\right] \\ & + \mathbb{P}\left[E^{\star}_{n,r} \leq (1-t)rK\cdot\frac{n-r}{n-1}\right]\end{aligned}$$

$$
\begin{aligned}
&\leq (1-p)^{(1-t)rK\cdot\frac{n-r}{n-1}} + e^{-\frac{t^2}{2}rK\cdot\frac{n-r}{n-1}} \\
&\leq e^{-(1-t)rpK\cdot\frac{n-r}{n-1}} + e^{-\frac{t^2}{2}rK\cdot\frac{n-r}{n-1}} \\
&\leq e^{-(1-t)rpK\cdot\frac{n-r}{n-1}} + e^{-\frac{t^2}{2}rpK\cdot\frac{n-r}{n-1}}.
\end{aligned}
$$

Therefore, whenever $r = 2, 3, \ldots, \lfloor \frac{n}{2} \rfloor$, we have

$$
\mathbb{E}\left[(1-p)^{E^{\star}_{n,r}}\right] \leq e^{-\frac{1-t}{2}\cdot rpK} + e^{-\frac{t^2}{4}\cdot rpK} \tag{10.69}
$$

since on that range we have

$$
\frac{n-r}{n-1} \geq \frac{n/2}{n-1} \geq \frac{1}{2}.
$$

Now consider a scaling $\theta : \mathbb{N}_0 \to \mathbb{N}_0 \times (0,1)$ such that $\lim_{n\to\infty} p_n = p^{\star}$ exists and (10.1) holds for some $c > \tau(p^{\star})$. Replace θ by θ_n in both (10.68) and (10.69) according to this scaling and use the bound of Lemma 10.9.4 in the resulting inequalities. Pick an integer $R \geq 2$ (to be further specified shortly) and for $n \geq 2(R+1)$ note that

$$
\begin{aligned}
\sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] &\leq \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} r^{r-2} \left(p_n \lambda_n(K_n)\right)^{r-1} e^{-\frac{1-t}{2}\cdot rp_nK_n} \\
&\quad + \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} r^{r-2} \left(p_n \lambda_n(K_n)\right)^{r-1} e^{-\frac{t^2}{4}\cdot rp_nK_n} \\
&\leq \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} en \left(ec_n \frac{n}{n-1}\log n\right)^{r-1} e^{-\frac{1-t}{2}\cdot rp_nK_n} \\
&\quad + \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} en \left(ec_n \frac{n}{n-1}\log n\right)^{r-1} e^{-\frac{t^2}{4}\cdot rp_nK_n}
\end{aligned}
$$

by the same arguments as the ones leading to (10.63). Upon invoking the lower bound in (10.15) we now conclude for all sufficiently large $n > 2(R+1)$ that

$$\begin{aligned}\sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] &\leq \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} en \left(ec_n \frac{n}{n-1} \log n \right)^r e^{-\frac{1-t}{4} \cdot r c_n \log n} \\ &\quad + \sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} en \left(ec_n \frac{n}{n-1} \log n \right)^r e^{-\frac{t^2}{8} \cdot r c_n \log n}. \\ &\leq \sum_{r=R+1}^{\infty} en \left(ec_n \frac{n}{n-1} \log n \cdot n^{-\frac{1-t}{4} \cdot c_n} \right)^r \\ &\quad + \sum_{r=R+1}^{\infty} en \left(ec_n \frac{n}{n-1} \log n \cdot n^{-\frac{t^2}{8} \cdot c_n} \right)^r.\end{aligned}$$

Furthermore, for all sufficiently large $n \geq 2(R+1)$ it also the case that

$$ec_n \frac{n}{n-1} \log n \cdot \max\left(n^{-\frac{1-t}{4} c_n}, n^{-\frac{t^2}{8} c_n} \right) < 1 \tag{10.70}$$

and the two infinite series converge. Let $n^\star(R)$ denote any integer larger than $2(R+1)$ such that (10.70) holds for all $n \geq n^\star(R)$. On that range, by our earlier discussion we get

$$\sum_{r=R+1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{r} \mathbb{P}\left[A_{n,r}(\theta_n)\right] \leq e \left(ec_n \frac{n}{n-1} \log n \right)^{R+1} (\ldots)$$

with

$$\ldots := \frac{n^{1-\frac{1-t}{4} c_n (R+1)}}{1 - ec_n \frac{n}{n-1} \log n \cdot n^{-\frac{1-t}{4} c_n}} + \frac{n^{1-\frac{t^2}{8} c_n (R+1)}}{1 - ec_n \frac{n}{n-1} \log n \cdot n^{-\frac{t^2}{8} c_n}}.$$

Finally, let n go to infinity in this last expression: The desired conclusion (10.55)

follows whenever the conditions $(1-t)c(R+1) > 4$ and $c(R+1)t^2 > 8$ are satisfied. This can be achieved by taking R so that

$$R + 1 > \max\left(\frac{4}{c(1-t)}, \frac{8}{ct^2}\right).$$

This is always feasible for any given t in $(0,1)$ by taking R sufficiently large. ■

10.11 A proof of Proposition 10.7.1

The basis for deriving (10.35) lies in the observation that nodes 1 and 2 are both isolated in $\mathbb{H} \cap \mathbb{G}(n;\theta)$ if and only if each edge in $\mathbb{H}(n;K)$ incident to one of these nodes is *not* present in $\mathbb{G}(n;p)$. Thus, $\chi_{n,1}(\theta) = \chi_{n,2}(\theta) = 1$ if and only if both sets of conditions

$$B_{1j}(p) = 0 \quad \text{if} \quad \Sigma_{n,1} \cap \Sigma_{n,j} \neq \emptyset,\ j \in \mathcal{N}_{-1}$$

and

$$B_{2k}(p) = 0 \quad \text{if} \quad \Sigma_{n,2} \cap \Sigma_{n,k} \neq \emptyset,\ k \in \mathcal{N}_{-2}$$

hold.

To formalize this observation, we introduce the random sets $N_{n,1}(\theta)$ and $N_{n,2}(\theta)$ defined by

$$N_{n,1}(\theta) := \{j = 3, \ldots, n :\ j \in \Gamma_{n,1}\ \vee 1 \in \Gamma_{n,j}\} \tag{10.71}$$

and

$$N_{n,2}(\theta) := \{k = 3, \ldots, n : \ k \in \Gamma_{n,2} \ \vee 2 \in \Gamma_{n,k}\}. \tag{10.72}$$

Thus, node j in $N_{n,1}(\theta)$ is neither node 1 nor node 2, and is K-adjacent to node 1. Similarly, node k in $N_{n,2}(\theta)$ is neither node 1 nor node 2, and is K-adjacent to node 2. Let $Z_n(\theta)$ denote the total number of edges in $\mathbb{H}(n; K)$ which are incident to either node 1 or node 2. It is plain that

$$Z_n(\theta) \ = \ |N_{n,1}(\theta)| + |N_{n,2}(\theta)| + \mathbf{1}\left[2 \in \Gamma_{n,1} \ \vee 1 \in \Gamma_{n,2}\right] \tag{10.73}$$

with the last term accounting for the possibility that nodes 1 and 2 are K-adjacent. By conditioning on the rvs $\Gamma_{n,1}, \ldots, \Gamma_{n,n}$, we readily conclude that

$$\mathbb{E}\left[\chi_{n,1}(\theta)\chi_{n,2}(\theta)\right] = \mathbb{E}\left[(1-p)^{Z_n(\theta)}\right] \tag{10.74}$$

under the enforced independence of the collections of rvs $\{\Gamma_{n,1}, \ldots, \Gamma_{n,n}\}$ and $\{B_{ij}(p), 1 \leq i < j \leq n\}$.

To proceed we need to assess the various contributions to $Z_n(\theta)$: Using the basic identity

$$\mathbf{1}\left[E \cup F\right] = \mathbf{1}\left[E\right] + \mathbf{1}\left[F\right] - \mathbf{1}\left[E \cap F\right]. \tag{10.75}$$

valid for any pair of events E and F, we find

$$|N_{n,1}(\theta)| \ = \ \sum_{j=3}^{n} \mathbf{1}\left[j \in \Gamma_{n,1} \ \vee 1 \in \Gamma_{n,j}\right]$$

$$
\begin{aligned}
&= \sum_{j=3}^{n} \mathbf{1}\left[j \in \Gamma_{n,1}\right] + \sum_{j=3}^{n} \mathbf{1}\left[1 \in \Gamma_{n,j}\right] \\
&\quad - \sum_{j=3}^{n} \mathbf{1}\left[j \in \Gamma_{n,1},\ 1 \in \Gamma_{n,j}\right] \\
&= \sum_{j=3}^{n} \mathbf{1}\left[j \in \Gamma_{n,1}\right] + \sum_{j=3}^{n} \mathbf{1}\left[j \notin \Gamma_{n,1},\ 1 \in \Gamma_{n,j}\right] \\
&= K - \mathbf{1}\left[2 \in \Gamma_{n,1}\right] + \sum_{j=3}^{n} \mathbf{1}\left[j \notin \Gamma_{n,1},\ 1 \in \Gamma_{n,j}\right] \qquad (10.76)
\end{aligned}
$$

where the last step used the fact $|\Gamma_{n,1}| = K$. Similar arguments show that

$$
\begin{aligned}
|N_{n,2}(\theta)| &= \sum_{k=3}^{n} \mathbf{1}\left[k \in \Gamma_{n,2} \ \vee 2 \in \Gamma_{n,k}\right] \\
&= K - \mathbf{1}\left[1 \in \Gamma_{n,2}\right] + \sum_{k=3}^{n} \mathbf{1}\left[k \notin \Gamma_{n,2},\ 2 \in \Gamma_{n,k}\right]. \qquad (10.77)
\end{aligned}
$$

As a result, from the definition of $Z_n(\theta)$ we get

$$
Z_n(\theta) = 2K - \mathbf{1}\left[2 \in \Gamma_{n,1},\ 1 \in \Gamma_{n,2}\right] + Z_n^{\star}(\theta) \qquad (10.78)
$$

upon using (10.75) one more time, where

$$
Z_n^{\star}(\theta) := \sum_{j=3}^{n} \mathbf{1}\left[j \notin \Gamma_{n,1},\ 1 \in \Gamma_{n,j}\right] + \sum_{j=3}^{n} \mathbf{1}\left[j \notin \Gamma_{n,2},\ 2 \in \Gamma_{n,j}\right]. \qquad (10.79)
$$

In order to evaluate the expression (10.74), we first compute the conditional expectation

$$
\mathbb{E}\left[(1-p)^{Z_n(\theta)} \middle| \Gamma_{n,1},\ \Gamma_{n,2}\right]. \qquad (10.80)
$$

From (10.78) we see that this quantity can be evaluated as the product of the two terms

$$(1-p)^{2K-(\mathbf{1}[2\in\Gamma_{n,1},1\in\Gamma_{n,2}])} \tag{10.81}$$

and

$$\mathbb{E}\left[(1-p)^{Z_n^\star(\theta)}\Big|\Gamma_{n,1},\ \Gamma_{n,2}\right]. \tag{10.82}$$

To evaluate this last conditional expectation, for each $j=3,\ldots,n$, we set

$$V_{n,j}(\theta;S,T) \ := \ \mathbb{E}\left[(1-p)^{\mathbf{1}[j\notin S,\ 1\in\Gamma_{n,j}]+\mathbf{1}[j\notin T,\ 2\in\Gamma_{n,j}]}\right]$$

with S and T subsets of $\mathcal{N}$, each being of size K. It is straightforward to check that

$$\begin{aligned} V_{n,j}(\theta;S,T) \ = \ & \mathbf{1}\left[j\notin S\right]\mathbf{1}\left[j\notin T\right]\mathbb{E}\left[(1-p)^{\mathbf{1}[1\in\Gamma_{n,j}]+\mathbf{1}[2\in\Gamma_{n,j}]}\right] \\ & +\mathbf{1}\left[j\notin S\right]\mathbf{1}\left[j\in T\right]\mathbb{E}\left[(1-p)^{\mathbf{1}[1\in\Gamma_{n,j}]}\right] \\ & +\mathbf{1}\left[j\notin T\right]\mathbf{1}\left[j\in S\right]\mathbb{E}\left[(1-p)^{\mathbf{1}[2\in\Gamma_{n,j}]}\right] \\ & +\mathbf{1}\left[j\in S\right]\mathbf{1}\left[j\in T\right]. \end{aligned}$$

Then, with the notation introduced earlier in Section 10.6, we can write

$$\begin{aligned} V_{n,j}(\theta;S,T) \ = \ & \mathbf{1}\left[j\notin S\right]\mathbf{1}\left[j\notin T\right]b_n(\theta) \\ & +\left(\mathbf{1}\left[j\notin S\right]\mathbf{1}\left[j\in T\right]+\mathbf{1}\left[j\notin T\right]\mathbf{1}\left[j\in S\right]\right)u_n(\theta) \\ & +\mathbf{1}\left[j\in S\right]\mathbf{1}\left[j\in T\right]. \end{aligned}$$

Next, the two rvs $\Gamma_{n,1}$ and $\Gamma_{n,2}$ being jointly independent of the rvs $\Gamma_{n,3}, \ldots, \Gamma_{n,n}$, we find

$$\begin{aligned} \mathbb{E}\left[(1-p)^{Z_n^\star(\theta)}\Big|\Gamma_{n,1},\ \Gamma_{n,2}\right] &= \prod_{j=3}^{n} V_{n,j}(\theta;\Gamma_{n,1},\Gamma_{n,2}) \\ &= b_n(\theta)^{B_n(\theta)} \cdot u_n(\theta)^{U_n(\theta)} \end{aligned} \tag{10.83}$$

where the rvs $B_n(\theta)$ and $U_n(\theta)$ are given by (10.33) and (10.34), respectively. Therefore, since

$$\mathbb{E}\left[(1-p)^{Z_n(\theta)}\right] = \mathbb{E}\left[\mathbb{E}\left[(1-p)^{Z_n(\theta)}\Big|\Gamma_{n,1},\Gamma_{n,2}\right]\right]$$

by a standard preconditioning argument, we get the expression (10.35) upon writing (10.80) as the product of the quantities (10.81) and (10.82), and using (10.83). ■

10.12 A proof of Lemma 10.9.1

The defining conditions for $B_{n,r}(\theta)$ lead to the representation

$$B_{n,r}(\theta) = \cap_{i=1}^{r} \cap_{k=r+1}^{n} E_{n,ik}(\theta)$$

where we have set

$$E_{n,ik}(\theta) := ([k \notin \Gamma_{n,i}] \cap [i \notin \Gamma_{n,k}]) \cup [B_{ik}(p) = 0]$$

with $i = 1, \ldots, r$ and $k = r+1, \ldots, n$. In terms of indicator functions, with the help of (10.75) this definition reads

$$\begin{aligned}
\mathbf{1}\left[E_{n,ik}(\theta)\right] &= \mathbf{1}\left[k \notin \Gamma_{n,i}\right]\mathbf{1}\left[i \notin \Gamma_{n,k}\right] + \left(1 - B_{ik}(p)\right) \\
&\quad -\mathbf{1}\left[k \notin \Gamma_{n,i}\right]\mathbf{1}\left[i \notin \Gamma_{n,k}\right]\left(1 - B_{ik}(p)\right) \\
&= \left(1 - B_{ik}(p)\right) + \mathbf{1}\left[k \notin \Gamma_{n,i}\right]\mathbf{1}\left[i \notin \Gamma_{n,k}\right] B_{ik}(p).
\end{aligned}$$

Therefore, under the enforced independence assumptions,

$$\begin{aligned}
&\mathbb{P}\left[B_{n,r}(\theta) \;\middle|\; \Gamma_{n,1}, \ldots, \Gamma_{n,n}\right] \\
&\quad = \mathbb{E}\left[\prod_{i=1}^{r}\prod_{k=r+1}^{n} W\left(\mathbf{1}\left[k \notin \Gamma_{n,i}\right]\mathbf{1}\left[i \notin \Gamma_{n,k}\right]; p\right)\right]
\end{aligned}$$

where

$$W(x; p) = 1 - p + px, \quad x \in \mathbb{R}.$$

Since $W(x, p) = (1-p)^{1-x}$ for $x = 0, 1$, we obtain

$$\mathbb{P}\left[B_{n,r}(\theta) \;\middle|\; \Gamma_{n,1}, \ldots, \Gamma_{n,n}\right] = \mathbb{E}\left[\prod_{i=1}^{r}\prod_{k=r+1}^{n} (1-p)^{1-\mathbf{1}[k\notin\Gamma_{n,i}]\mathbf{1}\left[i\notin\Gamma_{n,k}\right]}\right],$$

and it is now plain that

$$\mathbb{P}\left[B_{n,r}(\theta) \;\middle|\; \Gamma_{n,1}, \ldots, \Gamma_{n,r}\right] = (1-p)^{r(n-r)} G_{n,r}(\Gamma_{n,1}, \ldots, \Gamma_{n,r}; \theta)$$

where we have set

$$G_{n,r}(S_1,\ldots,S_r;\theta) \;=\; \mathbb{E}\left[\prod_{i=1}^{r}\prod_{k=r+1}^{n}(1-p)^{-\mathbf{1}[k\notin S_i]\mathbf{1}\left[i\notin\Gamma_{n,k}\right]}\right]$$

with $S_1,\ldots,S_r$ subsets of $\mathcal{N}$, each of size K.

Next, we find

$$\begin{aligned} G_{n,r}(S_1,\ldots,S_r;\theta) &= \mathbb{E}\left[\prod_{k=r+1}^{n}\prod_{i=1}^{r}(1-p)^{-\mathbf{1}[k\notin S_i]\mathbf{1}\left[i\notin\Gamma_{n,k}\right]}\right] \\ &= \mathbb{E}\left[\prod_{k=r+1}^{n}(1-p)^{-\sum_{i=1}^{r}\mathbf{1}[k\notin S_i]\mathbf{1}\left[i\notin\Gamma_{n,k}\right]}\right] \\ &= \prod_{k=r+1}^{n}\mathbb{E}\left[(1-p)^{-\sum_{i=1}^{r}\mathbf{1}[k\notin S_i]\mathbf{1}\left[i\notin\Gamma_{n,k}\right]}\right] \end{aligned}$$

as we again use the enforced independence assumptions. Fix $k = r+1,\ldots,n$ and note that

$$\begin{aligned} \mathbb{E}\left[(1-p)^{-\sum_{i=1}^{r}\mathbf{1}[k\notin S_i]\mathbf{1}\left[i\notin\Gamma_{n,k}\right]}\right] &= \mathbb{E}\left[\prod_{i=1}^{r}\left((1-p)^{-\mathbf{1}[k\notin S_i]}\right)^{\mathbf{1}\left[i\notin\Gamma_{n,k}\right]}\right] \\ &\leq \prod_{i=1}^{r}\mathbb{E}\left[\left((1-p)^{-\mathbf{1}[k\notin S_i]}\right)^{\mathbf{1}\left[i\notin\Gamma_{n,k}\right]}\right] \quad (10.84) \\ &= \prod_{i=1}^{r}\mathbb{E}\left[(1-p)^{-\mathbf{1}\left[i\notin\Gamma_{n,k}\right]}\right]^{\mathbf{1}[k\notin S_i]} \end{aligned}$$

where (10.84) follows from the negative association of the rvs (10.22) – Use (10.25) and note that

$$(1-p)^{-\mathbf{1}[k\notin S_i]} \geq 1, \quad i = 1,\ldots,r.$$

Next we observe that for each $i = 1, \ldots, r$, we have

$$\begin{aligned}
\mathbb{E}\left[(1-p)^{-\mathbf{1}\left[i \notin \Gamma_{n,k}\right]}\right] &= (1-p)^{-1}\,\mathbb{P}\left[i \notin \Gamma_{n,k}\right] + \mathbb{P}\left[i \in \Gamma_{n,k}\right] \\
&= (1-p)^{-1}\left(1 - \frac{K}{n-1}\right) + \frac{K}{n-1} \\
&= \frac{u_n(\theta)}{1-p}
\end{aligned}$$

whence

$$\prod_{i=1}^{r} \mathbb{E}\left[(1-p)^{-\mathbf{1}\left[i \notin \Gamma_{n,k}\right]}\right]^{\mathbf{1}[k \notin S_i]} = \left(\frac{u_n(\theta)}{1-p}\right)^{\sum_{i=1}^{r} \mathbf{1}[k \notin S_i]}.$$

Combining these observations readily yields

$$\begin{aligned}
G_{n,r}(S_1, \ldots, S_r; \theta) &\leq \prod_{k=r+1}^{n} \left(\frac{u_n(\theta)}{1-p}\right)^{\sum_{i=1}^{r} \mathbf{1}[k \notin S_i]} \\
&= \left(\frac{u_n(\theta)}{1-p}\right)^{\sum_{i=1}^{r} \sum_{k=r+1}^{n} \mathbf{1}[k \notin S_i]}.
\end{aligned}$$

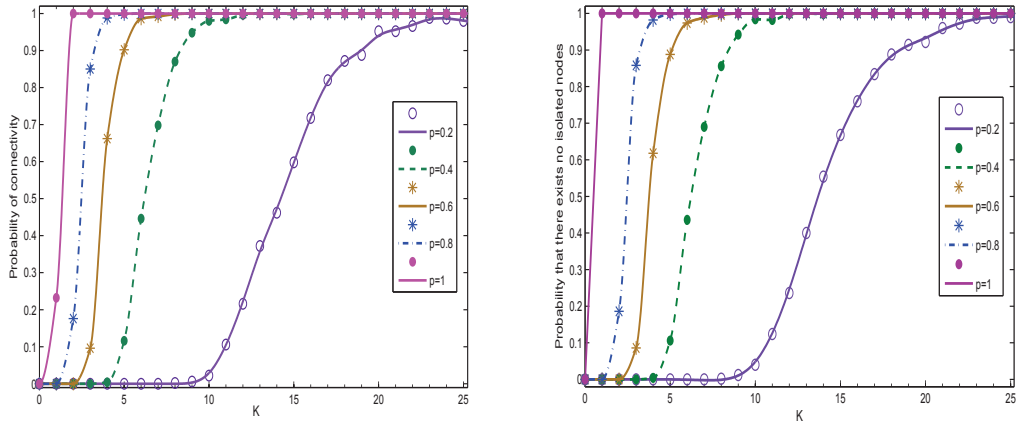
We finally obtain

$$\begin{aligned}
&\mathbb{P}\left[B_{n,r}(\theta) \;\middle|\; \Gamma_{n,1}, \ldots, \Gamma_{n,r}\right] \\
&\quad\leq (1-p)^{r(n-r)} \left(\frac{u_n(\theta)}{1-p}\right)^{\sum_{i=1}^{r} \sum_{k=r+1}^{n} \mathbf{1}[k \notin \Gamma_{n,i}]}
\end{aligned}$$

and the desired conclusion (10.46) follows. ■

10.13 Simulation study

We now present numerical results that verify (3.42). In all the simulations, we fix the number of nodes at $n = 200$. We consider the channel parameters $p = 0.2$, $p = 0.4$, $p = 0.6$, $p = 0.8$, and $p = 1$ (the full visibility case), while varying the parameter K from 1 to 25. For each parameter pair (K, p), we generate 500 independent samples of the graph $\mathbb{H} \cap \mathbb{G}(n; K, p)$ and count the number of times (out of a possible 500) that the obtained graphs i) have no isolated nodes and ii) are connected. Dividing the counts by 500, we obtain the (empirical) probabilities for the events of interest. The results for connectivity are depicted in Figure 10.1(a), where the curve fitting tool of MATLAB is used. It is easy to check that for each value of $p \neq 1$, the connectivity threshold matches the prescription (3.42), namely $K = \widehat{\tau}(p) \log n$. It is also seen that, if the channel is poor, i.e., if p is close to zero, then the required value for K to ensure connectivity can be much larger than the one in the full visibility case $p = 1$. The results regarding the absence of node isolation are depicted in Figure 10.1(b). For each value of $p \neq 1$, Figure 10.1(b) is indistinguishable from Figure 10.1(a), with the difference between the estimated probabilities of graph connectivity and absence of isolated nodes being quite small, in agreement with (3.42).



(a) (b)

Figure 10.1: *a) Probability that* $\mathbb{H} \cap \mathbb{G}(n; K, p)$ *is connected as a function of* K *for* $p = 0.2$, $p = 0.4$, $p = 0.6$, $p = 0.8$ *and* $p = 1$ *with* $n = 200$. *b) Probability that* $\mathbb{H} \cap \mathbb{G}(n; K, p)$ *has no isolated nodes as a function of* K *for* $p = 0.2$, $p = 0.4$, $p = 0.6$, $p = 0.8$ *and* $p = 1$ *with* $n = 200$. *This figure clearly resembles Figure 10.1(a) for all* $p \neq 1$.

BIBLIOGRAPHY

[1] I. F. Akyildiz, Y. Sankarsubramaniam, W. Su, E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks* **38:4**, pp. 393-422, 2002.

[2] N. P. Anthapadmanabhan and A. M. Makowski, "On the Absence of Isolated Nodes in Wireless Ad-Hoc Networks with Unreliable Links - a Curious Gap," Proceedings of the 29th Conference on Computer Communications (INFOCOM 2010), San Diego (CA), March 2010.

[3] S.R. Blackburn and S. Gerke, "Connectivity of the uniform random intersection graph," *Discrete Mathematics* **309:16**, pp. 5130-5140, 2009.

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

[5] S. A. Çamtepe and B. Yener, "Key Distribution Mechanisms for Wireless Sensor Networks: a Survey," Technical Report TR-05-07, Rensselaer Polytechnic Institute, Troy (NY), Computer Science Department, March 2005.

[6] A. Cayley, "A theorem on trees," *Quarterly Journal of Mathematics* **23**, pp. 376-378, 1889.

[7] H. Chan, A. Perrig and D. Song, "Random key predistribution schemes for sensor networks," Proceedings of the 2003 IEEE Symposium on Security and Privacy (S&P 2003), Oakland (CA), May 2003, pp. 197-213.

[8] R. Di Pietro, L.V. Mancini, A. Mei, A. Panconesi and J. Radhakrishnan, "Connectivity properties of secure wireless sensor networks," Proceedings of the 2nd ACM Workshop on Security of Ad Hoc And Sensor Networks (SASN 2004), Washington (DC), October 2004.

[9] R. Di Pietro, L.V. Mancini, A. Mei, A. Panconesi and J. Radhakrishnan, "Sensor networks that are provably secure," Proceedings of the 2nd IEEE/CreateNet International Conference on Security and Privacy in Communication Networks (SecureComm 2006), Baltimore (MD), August 2006.

[10] R. Di Pietro, L.V. Mancini, A. Mei, A. Panconesi and J. Radhakrishnan, "Redoubtable sensor networks," *ACM Transactions on Information Systems Security* **TISSEC 11:3** pp. 1-22, 2008.

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

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

[13] P. Erdös and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960), pp. 17-61.

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

[15] J. Fill, E.R. Scheinerman and K.B. Cohen-Singer, "Random intersection graphs when $m = \omega(n)$: An equivalence theorem relating the evolution of the $G(n, m, p)$ and $G(n, p)$ models," *Random Structures and Algorithms* **16** (2000), pp. 249-258.

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

[17] E. Godehardt and J. Jaworski "Two models of random intersection graphs for classification," in *Studies in Classification, Data Analysis and Knowledge Organization* **22**, Eds. O. Optiz and M. Schwaiger, Springer, Berlin (2003), pp. 67-82.

[18] E. Godehardt, J. Jaworski and K. Rybarczyk, "Random intersection graphs and classification," in *Studies in Classification, Data Analysis and Knowledge Organization* **33**, Eds. H.J. Lens and R,. Decker, Springer, Berlin (2007), pp. 67-74.

[19] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks, Chapter in *Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming*, Edited by W.M. McEneany, G. Yin and Q. Zhang, Birkhauser, Boston (MA), 1998.

[20] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, **46:2** (2000), pp. 388-404.

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

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

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

[24] M.K. Karoński, E.R. Scheinerman and K.B. Singer-Cohen, "On random intersection graphs: The subgraph problem," *Combinatorics, Probability and Computing* **8** (1999), pp. 131-159.

[25] G.E. Martin, *Counting: The Art of Enumerative Combinatorics*, Springer Verlag New York, 2001.

[26] A. Mei, A. Panconesi and J. Radhakrishnan, "Unassailable sensor networks," Proceedings of the 4th International Conference on Security and Privacy in Communication Networks (SecureComm 2008), Istanbul (Turkey), September 2008.

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

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

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

[30] K. Rybarczyk, Ph.D. Dissertation, Department of Discrete Mathematics, Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poznan (Poland), 2009.

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

[32] J. Spencer, "Nine lectures on random graphs," in Ecole d'Eté de Probabilités de Saint Flour XXI - 1991, Editor P.L. Hennequin, Springer Lecture Notes in Mathematics **1541**, Springer-Verlag Berlin Heidelberg 1993. pp. 293-347.

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

[34] D. West, *Introduction to Graph Theory*, Prentice Hall, New Jersey (NJ), 1996.

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

[36] O. Yağan and A.M. Makowski, "On the random graph induced by a random key predistribution scheme under full visibility (Extended version)," Available online at `http://hdl.handle.net/1903/7498`, January 2008.

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

[38] O. Yağan and A. M. Makowski, "Connectivity in random graphs induced by a key predistribution scheme – Small key pools," Proceedings of the 44th Annual Conference on Information Sciences and Systems (CISS 2010), March 2010. Earlier draft available online at `http://hdl.handle.net/1903/9055`, January 2009.

[39] O. Yağan and A.M. Makowski, "Zero-one laws for connectivity in random key graphs," Submitted to *IEEE Transactions on Information Theory*, October 2010. Available online at arXiv:0908.3644v1 [math.CO], August 2009. Earlier draft available online at `http://hdl.handle.net/1903/8716`, January 2009.

[40] O. Yağan and A. M. Makowski, "On the gradual deployment of random pairwise key distribution schemes," Proceedings of the Ninth International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt 2011), Princeton (NJ), May 2011.

[41] O. Yağan and A.M. Makowski, "Designing securely connected wireless sensor

networks in the presence of unreliable links," Proceedings of IEEE International Conference on Communications (ICC 2011), Kyoto (Japan), June 2011.

[42] O. Yağan and A.M. Makowski, "Modeling the pairwise key distribution scheme in the presence of unreliable links," Submitted to *IEEE Transactions on Information Theory*, October 2010. Available online at `arXiv:1102.2250v1[cs.IT]`.

[43] O. Yağan and A.M. Makowski, "On random pairwise graphs," Submitted to *Discrete Mathematics*, May 2011. Available online at `http://hdl.handle.net/1903/10601`.

[44] O. Yağan and A.M. Makowski, "Unassailability of sensor networks under a pairwise key distribution scheme," Submitted to the Twenty Second Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2011), Toronto (Canada), September 2011.

[45] O. Yağan and A.M. Makowski, "Key ring sizes in the random pairwise key distribution scheme," Submitted to the Twenty Second Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2011), Toronto (Canada), September 2011.

[46] O. Yağan and A. M. Makowski, "Secure and scalable wireless sensor networks: A comparison of random key distribution schemes," To be submitted to *ACM Transactions on Information and System Security (***TISSEC***)*.

[47] O. Yağan, "Performance of the Eschenauer-Gligor key distribution scheme under an ON-OFF channel model," In preperation.

[48] C.W. Yi, P.J. Wan, K.W. Lin and C.H. Huang, "Asymptotic distribution of the number of isolated nodes in wireless ad hoc networks with unreliable nodes and links," Proceedings of IEEE GLOBECOM 2006, San Francisco (CA), November 2006.