Robustness in Large-Scale Random Networks

by

Minkyu Kim

B.S., Electrical Engineering (1998)
Seoul National University

Submitted to the
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology

June 2003

Author . . . .
Department of Electrical Engineering and Computer Science
May 15, 2003

Certified by . . . .
Muriel Médard
Esther and Harold E. Edgerton Associate Professor
of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by . . . .
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Robustness in Large-Scale Random Networks

by

Minkyu Kim

Submitted to the Department of Electrical Engineering and Computer Science
on May 15, 2003, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

We consider the issue of protection in very large networks displaying randomness in topology. We employ random graph models to describe such networks, and obtain probabilistic bounds on several parameters related to various protection schemes. In particular, we take the case of random regular networks for simplicity, where the degree of each node is the same, and consider the length of primary and backup paths in terms of the number of hops. First, for a randomly picked pair of nodes, we derive a lower bound on the average distance between the pair and discuss the tightness of the bound. In addition, noting that primary and protection paths form cycles, we obtain a lower bound on the average length of the shortest cycle around the pair. Finally, we show that the protected connections of a given maximum finite length are rare. We then generalize our network model so that different degrees are allowed according to some arbitrary distribution. Notably, we derive an upper bound on the mean number of non-finite length cycles in generalized random networks. More importantly, we show that most of the results in regular networks carry over with minor modifications, which significantly broadens the scope of networks to which our approach applies. Our main contributions are the following. First, we take an analytical approach by bringing the concept of randomness into network topologies that can provide concise rules to relate basic network parameters to robustness. Second, we establish analytical results for the length of backup paths for path and link-based protection schemes rather than for the efficiency of backup capacity, upon which most studies concentrate. Finally, we develop a unified framework for studying the issue of robustness in very general random networks with arbitrary degree distributions.

Thesis Supervisor: Muriel Médard
Title: Esther and Harold E. Edgerton Associate Professor of Electrical Engineering and Computer Science

Acknowledgments

I wish to express my deep and sincere gratitude to Professor Muriel Médard for her guidance and patience throughout my past two years at MIT. She has always been truly open to her students and eager to provide insight, warm support and invaluable encouragement. I am greatly indebted to her for my development as a graduate student.

Contents

List of Figures

Chapter 1

Introduction

Providing resilient service against failures is an important issue for high-speed networks. As they operate at very high data rates, a single failure may cause a severe loss of data. For networks of ring topologies, the self-healing ring architecture has been successful as a means of providing simple and fast recovery. However, today's high-speed networks are becoming increasingly complex and also dynamic in topology. Furthermore, in response to growing and shifting communication demands, they expand rapidly. As networks evolve into large mesh topologies, a more complicated recovery mechanism is required.

Restoration has been extensively researched for general mesh topologies, but very few analytical results have been given so far. The typical approach is to give linear programming formulations or heuristic algorithms and to rely on simulations on some standard networks for evaluating their performance [26], [17]. However, this type of method may in a sense lack scalability. To be more precise, it can provide numerical results for each network with a specific topology, but it may fail to give an analytical view of how the algorithm works or how the parameters scale as the size of the network grows.

Every network evolves over time, that is, nodes are added and deleted, or different networks can be interconnected. Furthermore, as a network grows rapidly and becomes very complex in topology, it tends to no longer remain a single unit under the control of a single entity. In such networks, the topological changes may be well de-

scribed as random events rather than controlled occurrences. Hence, those networks need an appropriate tool to describe the randomness they display in topology.

In this research, we investigate the robustness of general mesh networks when they grow very large in a random manner described by probabilistic models. To this end, we employ random graph models for describing such networks. Assuming the network is very large, we consider the length of paths used for primary and backup traffic in a probabilistic sense and the probability that such paths exist. To be more precise, we obtain bounds on those parameters in terms of the size of the network in order to know how they scale as the network grows.

By studying these parameters, we can obtain an analytical sense of how networks will measure if they grow in the way described by such random graph models, which may be an interesting problem in its own. Though we do not develop a specific scheme or protocol for protection in this research, we can use the knowledge of those parameters to choose or evaluate which protection schemes are more appropriate in such large-scale networks. This study can further contribute to designing protection mechanisms that take advantage of the topological properties in such networks [14].

This chapter gives an overview of relevant work in the area of network protection and random graph theory. In Chapter 2, we consider random regular networks, where the degree of each node is the same, and obtain bounds on several parameters related to protection. Chapter 3 extends those results to more general cases of non-regular networks. Finally in Chapter 4, we present the conclusions of our research and discuss some directions for further work.

1.1 Background

1.1.1 Protection in Networks

For networks of ring topologies, SONET self-healing ring (SHR) architecture is a very successful protection technique [37], [20]. Protection can be done in two different ways in SHR. First, in a unidirectional path-switched ring (UPSR), each connection

maintains two node-disjoint paths, and signals are transmitted on both fibers in different directions from source to destination. When a failure occurs and affects one of the signals, the destination node detects the failure and chooses to receive the other signal which is still valid. A UPSR is the fastest SHR scheme because no switching of signals is needed. However, it doubles bandwidth requirements and is often referred to as 1+1 protection.

In the second case, a bidirectional line-switched ring (BLSR) provides a more capacity-efficient protection method. In a BLSR, two nodes adjacent to a failure loop the signal back to another fiber in the opposite direction. Since this switching is done dynamically when a failure occurs, the fibers required for the second route can be shared between many connections. However, this sharing may cause some parts of a network to lose their protection because of protection elsewhere in the network. Since both end nodes of the failed link are required to detect the failure and reroute the traffic between them, protection in a BLSR is slower than in a UPSR.

As traffic demands grow and become dynamic, nodes are added in a network and sometimes different networks are interconnected, which may make ring-based structures difficult to maintain. Moreover, ring-based architectures may be more expensive than meshes [6] and do not scale well as the network grows [37]. Hence, mesh-based architectures become more promising candidates for future networks than ring-based architectures such as simple rings or interconnected rings.

Protection methods employed in mesh networks can be classified as either path or link protection, each of which can be considered a generalization of the case of a UPSR or BLSR, respectively. Path protection refers to recovery applied to connections following a particular path across a network [22]. For each primary traffic, an end-to-end protection path from source to destination is established and reserved to carry the backup traffic. The primary and backup path should be link-disjoint or node-disjoint to protect against link or node failure, respectively. Path protection itself may be done in two different ways. In the first case, just as in a UPSR, signals are transmitted on both primary and backup paths, and upon the failure of a link or node on the primary path, the destination node switches to receiving the signal

13

on the backup path. Again, protection is very fast, but the capacity requirement is doubled. In the second case, when a failure occurs, the source and destination nodes of each connection affected by the failure switch to the corresponding backup path. That is, the backup path is activated only when a failure occurs, and thus, backup capacity can be shared by many different primary connections. However, for each connection affected by the failure, the corresponding source-destination pair needs to be notified of the failure and reroute the traffic between them. This coordination may yield delays and management overhead.

On the other hand, link protection refers to recovery of all the traffic across a failed link. When a link failure occurs, as in a BLSR, the traffic is looped back along a backup path around the failed link. The same procedure may be done around the failed node in case of node failure, which is called node protection. Only the two nodes adjacent to the failure are concerned in protection, regardless of the end nodes of the connections affected by the failure. Hence, link protection offers a further advantage over path protection in that, since it is not dependent on specific traffic patterns, the preplanned protection paths can be used without complete knowledge of the traffic pattern in the network [22].

In any case, establishing primary and backup paths involves a cycle (or a ring). More precisely, if we use the path protection scheme, we have to establish a backup path which is link(node)-disjoint from source to destination. We see that the primary and the backup paths form a cycle along the source and the destination. Also in the link protection scheme, the backup path around the failed link, together with the failed link itself, form a cycle. In light of these observations, cycles are crucial for protection in networks.

Extensive research has been done in the area of covering mesh topologies with rings. Every link to carry protected traffic is to be covered by at least one ring [16], or exactly two rings [12], or cycles known as p-cycles [17]. Then, in the same manner as in traditional SHRs, protection is done along the ring(s) to which the failed link belongs. However, minimizing the amount of fiber covering the network [34] or finding such cycle covers for non-planar topologies may be difficult for a large-scale

14

network [18]. Moreover, these structures may need a significant reconfiguration as a link or node is added. There is a more recent approach that does not rely on the ring structures, called generalized loopback [22]. This approach utilizes the redundancy embedded in mesh structure itself by selecting a digraph whose conjugate is used to carry backup traffic.

Most work on the robustness of networks is concerned with the bandwidth efficiency of protection schemes in terms of the capacity devoted solely to backup purposes. Since any protection mechanism comes from allowing some amount of redundant capacity for backup, which leads to additional cost, it is quite natural to try to reduce the redundancy. Hence, the objective functions of linear programming formulation, the most common approach in the field, are often related to the total usage of capacity for primary and backup traffic (e.g., in [26]). The speed of protection is also considered [27], sometimes jointly with capacity [17]. Some other considerations are transparency, flexibility, and vulnerability [22], [20].

In this research, we are concerned with the length of paths in terms of the number of hops. While the length of paths is less widely considered than the bandwidth efficiency, it is important in several contexts. For instance, in optical networks, backup paths must remain within a moderate range for optical signal quality [28]. To be more specific, optical signals are attenuated as they propagate and, after some distance, they may not be able to remain in optical form owing to the cumulative loss. The most common solution for this is to convert the optical signal into an electrical signal, process it, and convert it back into optical form. This regeneration process is subject to the bit rate and specific modulation format used by the system, which limits an upgrade to a higher rate or use for multiple wavelengths. For this reason, in high-performance optical networks, optical amplifiers which have several advantages over regenerators are becoming preferable, yet this still requires costly devices. Hence, in any case, a shorter length either for a primary or backup path is desirable.

Furthermore, shorter path lengths are desirable simply in terms of economy of resource use, i.e., if a backup path becomes longer, a larger amount of resource is required along the path. Hence, length of backup paths, in fact, indirectly affects

efficiency. In addition, a longer path tends to entail a larger amount of delay and management overhead.

In [22], the length of backup path is considered in selecting directions in the backup digraph. The tradeoff between path length and capacity is investigated in [10]. In particular, the authors show that, by including terms for path lengths in the objective function, the lengths of backup paths can be significantly shortened with very small or no spare capacity penalty.

1.1.2 Random Graph Theory

A random graph is a set of vertices with edges connecting pairs of vertices at random [23]. The theory of random graphs was founded in the 1950s by Erdős and Rényi, who had discovered that probabilistic methods were often useful in dealing with extremal problems in graph theory [5]. The random graph model introduced by Erdős [13] is very natural. Given a vertex set $[n]=\{1,2,...,n\}$, we choose a graph at random with equal probabilities, from the set of all $2^{\binom{n}{2}}$ graphs. This is equivalent to tossing a fair coin $\binom{n}{2}$ times independently, where each toss corresponds to a pair of vertices among a total of $\binom{n}{2}$ pairs, and drawing an edge between the pair if the corresponding toss turns up heads.

Since then, random graph theory has further developed and now has become an important topic of modern discrete mathematics. Among several current models of random graphs, two basic models are the binomial model and the uniform model [19]. First, given a real number p, $0 \leq p \leq 1$, the *binomial random graph* $G(n,p)$ is defined by taking the set of all graphs on the vertex set $[n]$ as the sample space Ω, and setting

$$\Pr(G) = p^{e_G}(1-p)^{\binom{n}{2}-e_G}, \qquad G \in \Omega$$

where e_G denotes the number of edges in G. Also, $G(n,p)$ can be viewed as the result of $\binom{n}{2}$ independent coin tosses with the probability of heads being equal to p, where an edge is placed between the pair of vertices corresponding to heads. Hence, the previous model by Erdős is a special case of this, where $p = 1/2$. However, in most

work on random graphs, it is typically assumed that n, the number of vertices, tends to infinity and also, in binomial random graphs, $p = p(n) \to 0$ as $n \to \infty$.

The most crucial assumption in the binomial model is that the presence or absence of an edge between two vertices is independent of the presence or absence of any other edges [23]. This assumption greatly simplifies the analysis, which is the reason why this model has been studied extensively. The probability that a vertex has degree k, which we denote by p_k, is

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \sim \frac{z^k e^{-z}}{k!},$$

where the approximation holds when $n \to \infty$ and $z = np$. We recognize this distribution as the Poisson distribution, and hence in some literature, this model is called the Poisson random graph model [25]. A drawback of this model is that the number of edges is not fixed; it varies according to a binomial distribution with expectation $\binom{n}{2}p$. On the other hand, if we condition that e_G, the number of edges in G, is fixed to M, then we obtain another standard random graph model, which we will discuss next.

Given an integer M, $0 \le M \le \binom{n}{2}$, the *uniform random graph* $G(n, M)$ is defined by taking the set of all graphs on vertex set $[n]$ with exactly M edges as the sample space Ω, and setting the probability of a certain graph $G \in \Omega$ to be uniform, i.e.,

$$\Pr(G) = \binom{\binom{n}{2}}{M}^{-1}, \qquad G \in \Omega.$$

However, the two basic models turn out to be asymptotically equivalent in many cases, provided $\binom{n}{2}p$ is close to M [19].

Similarly, we can further consider a broader family of uniform random graphs defined by taking the uniform distribution over a certain set $\mathcal{F}$ of graphs [19]. The earliest model by Erdős can also be considered to belong here, with $\mathcal{F}$ corresponding to the set of all graphs on a given vertex set. Another popular model of this type is that of random regular graphs, where $\mathcal{F}$ is the set of all graphs on n vertices

with the degree of each vertex set to d, provided dn is even. This model yields a uniform random regular graph, denoted by $G(n, d)$. Random regular graphs, which we will discuss in depth in the next chapter, turn out to have many useful asymptotic properties.

Random graphs are not merely a mathematical toy; they have been widely used as models of real-world networks of various types [23]. Also, extensive research has been done on the structures and properties of many kinds of real-world networks. Examples include social networks of acquaintance, networks of citations between papers, networks of business relations between companies, neural networks, metabolic networks, the World Wide Web, and many others [25]. However, in many of these studies, the distribution of vertex degrees in real networks is notably different from the Poisson distribution. In other words, the binomial random graph model, which is a standard model in the random graph theory, often is not well suited for modelling real-world networks. Hence, we need to employ a revised graph model to describe real-world networks more closely such that the model allows general degree distributions, which will be our main topic in Chapter 3.

Note that, if we define p_k to be the fraction of vertices in the network that have degree k, then equivalently, p_k is the probability that a vertex chosen uniformly at random has degree k. Therefore, the set of values of p_k represents the probability distribution of degrees in the network, which we call the degree distribution. A very common result is that, in most real-world networks, the degree distribution is highly right-skewed, i.e., it has a long right tail of values that are far above the mean [25]. More specifically, it is often of the form $p_k \sim k^{-\alpha}$ for some positive constant α, which is called the power law distribution. Note that networks with this property are often called scale-free networks. A considerable number of studies have been carried out on this topic, e.g., [3], [33], [1].

Despite the extensive research on various types of complex networks, there is a very limited amount of literature focusing mainly on communication networks, where the issue of protection is of critical importance. In contrast, for example, in social networks such as networks of citations between scientific papers, protection is of no

interest at all; rather, in such networks, the notion of failure is not well defined.

The telephone network or networks of wireless service providers, which may be the most typical examples of communication networks, are presumably studied within the relevant corporations but not yet by academic researchers [25]. A more widely considered network is the Internet[1]. The actual and complete structure of physical connections on the Internet, however, is inherently difficult to capture because the number of computers is very large and rapidly changing, and moreover, the infrastructure is maintained by many separate organizations [25]. Therefore, a typical approach is to reconstruct the network by reasoning from large samples of point-to-point data routes. Using the data collected by this method, Faloutos *et al.* [14] discover that several parameters such as outdegrees of a node or eigenvalues of the Internet graph display power law distributions. However, more recently Chen *et al.* [7] observe that the previous method used by Faloutos *et al.* to construct the Internet topologies may miss a significant number of connections in the maps constructed by a novel method using additional data sources. Moreover, Chen *et al.* show that the degree distributions in the new maps are also heavy-tailed, but deviate significantly from a strict power law. This indicates that, in considering communication networks, we do not need to limit ourselves to networks with the degree distribution of a power law.

As we mentioned, studies on large-scale communication networks are few, but those on the robustness in such networks are even fewer. Albert *et al.* [2] study the error tolerance of complex networks by measuring the change in diameter when a small fraction of the nodes is removed. They address the fact that scale-free networks, i.e., networks with power-law degree distributions, display a very high degree of robustness. That is, the ability of their nodes to communicate remain unaffected even by unrealistically high failure rates. Also, they show that these networks are extremely vulnerable to the selection and removal of a few nodes that play a vital role in maintaining the network's connectivity. Note, however, that the authors of [2]

---

[1]The Internet should not be confused with the World Wide Web. The former refers to a physical network of computers linked together by optical fiber or other communication media, while the latter refers to a network of Web pages linked together by citing other pages with hyperlinks. The World Wide Web may more appropriately be considered a kind of information network.

mainly focus on the generic property of scale-free networks regarding their connectivity against failures – which is solely due to their scale-free property – but not on the issue of protection.

Chapter 2

Random Regular Networks

In this chapter, we consider the networks described by random regular graph models, where the degree of each node is the same. In a sense, this condition may seem too restrictive for random regular graph models to represent real networks in practice. However, our intention is to first obtain some intuition by looking at the simplest case, and then to generalize it for more complicated and realistic cases. Indeed, many simple and useful properties that hold asymptotically for random regular graphs yield very intuitive results. In the next chapter, we will find that many of these results can carry over with minor modifications to more general networks.

2.1   Models of Random Regular Graphs

We begin by investigating the two models of random regular graphs by which we describe our random networks. Later, we will discuss which model we choose to use for further analysis. Note that, especially in this section, we will use terms from the mathematical graph theory, that is, each vertex in the graph corresponds to a node in the network and each edge to a link.

Before proceeding, for clarification, we give detailed descriptions of the notation that we will often use in the sequel. If we let a_n and b_n be sequences of numbers depending on n and assume $b_n > 0$ for all sufficiently large n,

- $a_n = O(b_n)$ if there exist constants C and n_0 such that $|a_n| \leq C b_n$ for $n \geq n_0$.

- $a_n = \Omega(b_n)$ if there exist constants $c > 0$ and n_0 such that $a_n \geq cb_n$ for $n \geq n_0$.

- $a_n = \Theta(b_n)$ if there exist constants $C, c > 0$ and n_0 such that $cb_n \leq a_n \leq Cb_n$ for $n \geq n_0$.

- $a_n \sim b_n$ if $\frac{a_n}{b_n} \to 1$ as $n \to \infty$.

- $a_n = o(b_n)$ if $\frac{a_n}{b_n} \to 0$ as $n \to \infty$, i.e., if for every $\varepsilon > 0$ there exists n_ε such that $|a_n| < \varepsilon b_n$ for $n \leq n_\varepsilon$.

Also, for an event $\mathcal{E}_n$, which describes a property of a random structure depending on a parameter n,

- We say that $\mathcal{E}_n$ holds *asymptotically almost surely* (*a.a.s.*) if $\Pr(\mathcal{E}_n) \to 1$ as $n \to 1$.

We use the standard notation of double factorial, defined as

$$
n!! = \begin{cases}
n \cdot (n-2)...5 \cdot 3 \cdot 1 & n > 0 \text{ odd} \\
n \cdot (n-2)...6 \cdot 4 \cdot 2 & n > 0 \text{ even} \\
1 & n = -1, 0.
\end{cases}
$$

2.1.1   Configuration Model

Most work on random regular graphs is based on the following construction model introduced by Bender and Canfield [4]. Here we will use Janson *et al.*'s interpretation [19] of the original model.

Suppose we have natural numbers n and d denoting the number of vertices and the common degree of every vertex, respectively. We let $3 \leq d \leq n-1$ and assume dn is even. Then we can think of the set of all possible d-regular graphs on those n vertices. We turn this set into a probability space by assigning the same probability to each element of the set. In other words, we get a d-random graph $G(n, d)$ by picking an element uniformly at random among all possible d-regular graphs. This is an intuitive description of the *configuration model*, also called *uniform model*, which is a standard

method for constructing random regular graphs. Below is another description that is more detailed and enables further analysis of the model in a structured way.

Let V be the set of vertices $[n]$ corresponding to n places along the horizontal axis. For each place in V, we introduce d vertices and call this two-dimensional set of dn vertices $W = [n] \times [d]$. A *configuration* is a partition of W into $(dn/2)$ pairs, and a *random configuration* is a configuration chosen uniformly at random from all possible partitions. If we project the set W onto $V = [n]$ by simply ignoring the second coordinate, we obtain a multigraph $\pi(F)$ where each pair in the configuration is considered an edge. However, this is not an ordinary graph because it allows loops around the same vertex and multiple edges between two vertices, which, in other words, are cycles of length 1 and 2, respectively. In particular, if $\pi(F)$ lacks those loops and multiple edges, it is a simple graph which is d-regular. Note that each simple d-regular graph corresponds to precisely $(d!)^n$ configurations. Hence, if we choose a configuration uniformly at random, conditioned on it being a simple graph, we get $G(n, d)$ as desired.

For simplicity, it is often advantageous to first allow loops and multiple edges and to work with d-regular multigraphs. Then, if necessary, we condition on the graphs being simple. The configuration model has an important feature in that the probability that we obtain a simple graph is bounded below by some positive number for all $n > r$. Therefore, the probability of an event in (simple) random regular graph $G(n, r)$ is bounded by a constant times the probability of the same event in random regular multigraph $G^*(n, r)$, where the latter is the probability of the corresponding event in a random configuration. Hence, if we obtain an asymptotic property on $G^*(n, r)$ by working with the random configurations, then it also applies to $G(n, r)$. This can be summarized by the following theorem [19]:

Theorem 2.1.1 *Any property that holds* a.a.s. *for $G^*(n, d)$ also holds* a.a.s. *for $G(n, d)$.*

Connectivity of graphs is a critical issue. If a graph model representing networks is not connected initially, then it breaks into several subgraphs, each of which is

disconnected from the other parts and can be dealt with as a separate problem. Moreover, if removal of a single edge or vertex would cause a certain set of source-destination pairs to be disconnected, then we have no viable option to restore the connections but to recover the failed edge or vertex itself. Therefore, in this case, there is no need to find a recovery path for them. However, in random regular graphs constructed by the configuration model, it is known that such disconnectivity happens rarely as n tends to infinity. More precisely, we have the following result regarding connectivity [35]:

Theorem 2.1.2 *If $d \geq 3$ and fixed, then $G(n,d)$ is* a.a.s. *d-connected.*

Note that we say a graph is *d-connected* if, for any pair of vertices i and j, there is a path connecting i and j in every subgraph obtained by deleting $(d-1)$ vertices other than i and j together with their adjacent edges from the graph. Therefore, for sufficiently large n, we still get a connected graph even if we remove a single edge or vertex from $G(n,d)$ for $d \geq 3$.

Now, let us consider the distribution of cycles in a graph. Define a random variable Z_k to be the number of cycles of length k in $G(n,d)$. It is known that, for any set of k's that are fixed and $k \geq 3$, Z_k's are asymptotically distributed according to independently joint Poisson random variables [5].

Theorem 2.1.3 *For each fixed j, a sequence of random variables $(Z_3, Z_4, ..., Z_j)$ converges in distribution to $(Z_{3\infty}, Z_{4\infty}, ..., Z_{j\infty})$, where $\{Z_{k\infty}\}_{k=3}^{j}$ is a sequence of independent Poisson distributed random variables with $E(Z_{k\infty}) = \frac{(d-1)^k}{2k}$.*

Note, however, that the previous theorem applies only for the cycles of fixed length, that is, where the length of cycle does not grow with n. On the other hand, if the length is equal to n, the cycle is called a Hamiltonian cycle. A result on the existence of a Hamiltonian cycle is given by Robinson and Wormald [29].

Theorem 2.1.4 *If $d \geq 3$ is fixed, then $G(n,d)$* a.a.s. *has a Hamiltonian cycle.*

The case in the middle of these two extremes has been considered more recently by Garmo [15]. If the length of cycles k is defined as a function of n, i.e., $k = k(n)$,

the limiting distribution turns out to depend on whether $k(n)/n \to 0$ or $k(n)/n \to q$, $0 < q < 1$, as $n \to \infty$. We address here only the results of the former case, which is of our interest. It is shown in [15] that, if $k(n)/n \to 0$ as $n \to \infty$, the number of k-cycles, Z_k, converges in probability to its mean as given below.

Theorem 2.1.5 *In random d-regular graph, where d is fixed and $d \geq 3$, let $k(n) \geq (\frac{2}{\ln(d-1)}) \ln n$ and assume that $\mu_n = k(n)/n \to 0$ as $n \to \infty$. Then,*

$$\frac{Z_k}{E(Z_k)} \to 1 \quad as \ n \to \infty$$

in probability.

In [15], Garmo also obtains the asymptotic distribution of Z_k, i.e., how Z_k is distributed closely around $E(Z_k)$. But what is more relevant here is the value of $E(Z_k)$ when k grows with n. Recall that, by projecting a random configuration onto the horizontal axis, we obtain a random regular multigraph $G^*(n,d)$ unless we condition that there are no simple loops or multiple edges. Let Z_k^* be the number of k-cycles in $G^*(n,d)$. By counting the number of cycles on the two-dimensional set $W = [n] \times [d]$ and using Stirling's formula, Garmo calculates $E(Z_k^*)$, $k = 1, 2, ..., n$, as in the lemma below.

Lemma 2.1.6 *Let k be an integer, $1 \leq k \leq n$, and $\lambda = k/n$. Then,*

$$\begin{aligned}
E(Z_k^*) &= \frac{(d(d-1))^k}{2k} \frac{n!}{(n-k)!} \frac{(dn-2k-1)!!}{(dn-1)!!} \\
&= \frac{(d-1)^k}{2k} \frac{1}{\exp\{\frac{1}{2}(\frac{d-2}{d}k-1)\lambda + O(k\lambda^2)\} + O(\frac{1}{n})}.
\end{aligned} \quad (2.1)$$

Also, Garmo shows the following relation between $E(Z_k)$ and $E(Z_k^*)$:

Lemma 2.1.7

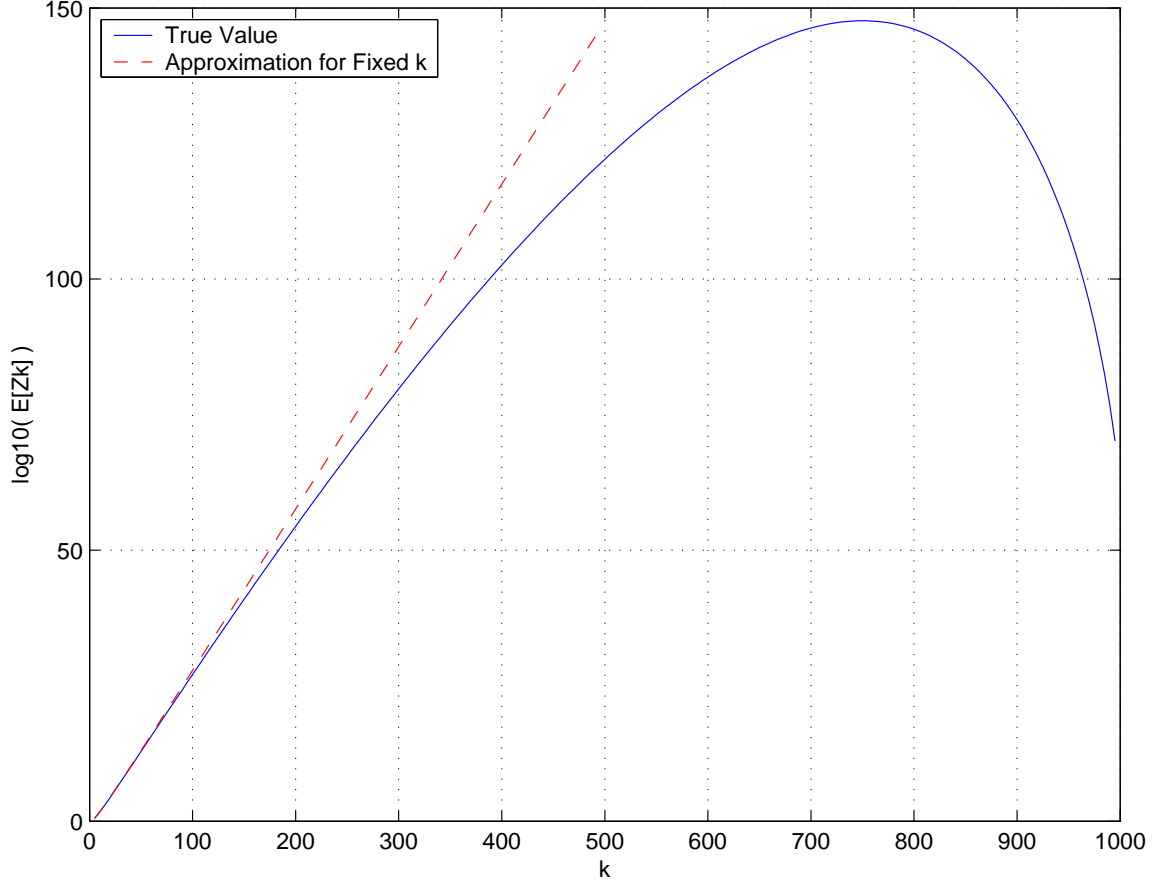$$\frac{E(Z_k)}{E(Z_k^*)} \to 1 \quad as \ n \to \infty,$$

Figure 2-1: $\log \mathrm{E}(Z_k)$ with respect to k in $G(n,d)$ where $n = 1,000$ and $d = 3$

which agrees with Theorem 2.1.1. Therefore, we can conclude that $\mathrm{E}(Z_k)$ is also represented as $\mathrm{E}(Z_k^*)$ in Lemma 2.1.6 as n tends to infinity, which applies to any k[1], $3 \leq k \leq n$, i.e., either k remains fixed or grows with n. We recall that, in Theorem 2.1.3, $\mathrm{E}(Z_k)$ is shown to be $a.a.s.$ $\frac{(d-1)^k}{2k}$ for fixed k. Now, we see that this is congruent to Lemma 2.1.6 since if k is fixed, then $\lambda = k/n$ tends to zero as $n \to \infty$. On the other hand, if k grows with n, in (2.1), the exponent terms in the denominator, $\frac{1}{2}(\frac{d-2}{d}k - 1)\lambda + O(k\lambda^2)$, may no longer be negligible, which leads $\mathrm{E}(Z_k)$ to become smaller than $\frac{(d-1)^k}{2k}$ (see Figure 2-1).

In subsequent sections, we will use these asymptotic results to quantify the relia-

---

[1]In a (simple) random regular graph $G(n,d)$, the smallest possible length of a cycle is 3 because cycles of length 1 or 2 are not allowed.

bility issues of networks represented by the configuration model.

2.1.2 Sequential Model

Though the configuration model is the most general method of describing random regular graphs and possesses many useful properties, its construction process is less intuitive than the *sequential model*, which we briefly discuss in this section. Furthermore, in a practical sense, it is difficult to efficiently generate an instance of random regular graph, $G(n, d)$. Hence, the desire for models which are more algorithmic, i.e., practically implementable for generating random regular graphs. These models can often be described in simpler and more intuitive ways. But the price for this simplicity is that most of these models have nonuniform distributions, some of which are currently not well understood [36].

The model we discuss in this section is a process that generates random regular graphs sequentially [30]. To build a d-regular graph, start with n isolated vertices and repeatedly add edges joining vertices of degree strictly less than d. Each time, we choose uniformly at random the edge to add from all possible positions and stop when no more edges can be added. This process is called the degree restricted graph process with parameter d, or d-process. In [30], the authors show that *a.a.s.* the final graph is regular if dn is even, and otherwise, almost regular, i.e., one vertex of degree $(d - 1)$ and the rest of degree d.

We see that the above process is conceptually easier to understand than the configuration model and with slight modifications, it can be easily extended to model the processes that occur in real networks. However, for this process, we lack further knowledge on the asymptotic properties that are known for the configuration model. Regarding connectivity, it has recently been shown that the resulting graph is *a.a.s.* connected for $d \geq 3$ [32]. But it is yet to be shown whether this graph is also d-connected as $G(n, d)$ from the configuration model. For the distribution of cycles, it is shown that the number of cycles of fixed length is asymptotically Poisson distributed, but only when $d = 2$ [31].

Theorem 2.1.8 *Let G_n denote a graph generated by a random 2-process. If $k \geq 3$ is fixed, then the number of cycles of length k is asymptotically Poisson. For $k = 3$, the mean converges to*

$$\frac{1}{2} \int_0^\infty \frac{(\log(1+x))^2 dx}{xe^x} \approx 0.188735349357788830.$$

However, for general k, in contrast to the case of $G(n, d)$, the expected number of cycles of length k involves a k-dimensional integral. Moreover, the previous result holds only where $d = 2$; nothing is yet known about the distribution of cycles for $d \geq 3$.

In the absence of established results on these asymptotic properties, the sequential model is not best suited for our further quantified analysis. Hence, in subsequent sections we will use only the configuration model for describing our random networks.

2.2 Shortest Path

Throughout the remaining sections in this chapter, we assume that our network is a large random d-regular graph generated by the configuration model. Suppose n, the number of nodes, is large enough so that all the asymptotic properties in Section 2.1.1 are assumed to hold, i.e., the deviation from the asymptotic behavior is assumed to be negligible.

Before proceeding, we present an important property of the model that we will use frequently in further analysis. If we pick a pair of nodes randomly and define a random variable X representing some parameter related to the pair, e.g., the distance between the pair, then there are two sources of randomness: one is from the selection of a graph and the other from the selection a pair of nodes. However, note that by the symmetric structure of the configuration model, the value of X has no dependence on a specific pair. Hence, calculation of the expectation of X which is over the probability space of the selection of a graph is not affected by averaging X again

over the selection of a pair. Furthermore, by interchanging the order of calculation, we obtain a more convenient way to compute the expectation of X – that is, first conditioning on some graphs to get the expected value of X over the pair selection and then averaging the expectation over all graphs.

Let us fix a randomly chosen pair of nodes, s and t, and define a random variable L to be the length of the shortest path between s and t. Then, as argued above, assume that we have a certain d-regular graph and consider the value of L over the possible selections of a pair.

2.2.1 Lower Bound

It is clear that there are d nodes adjacent to s. If we consider the nodes two hops away from s, there can be at most $d(d-1)$ such nodes, but some of them may overlap and therefore $d(d-1)$ is an upper bound on the number of such nodes. Now if we count the total number of nodes within two hops of s, some nodes adjacent to s and some nodes two hops away from s may again overlap, but still there can be at most $d+d(d-1) = d^2$ such nodes if all of them are distinct. If we continue this counting, the number of nodes within k hops of s is at most $d+d(d-1)+d(d-1)^2+...+d(d-1)^{k-1} = d\{(d-1)^k - 1\}/(d-2)$ (see Figure 2-2). Note that, in the probability space of the pair selection, $\Pr(L \le k)$ is the probability that we pick another node t among those nodes within k hops of s. Hence,

$$\Pr(L \le k) \le \min[1, \left(\frac{d\{(d-1)^k - 1\}}{d-2} \cdot \frac{1}{n-1} \right)].$$
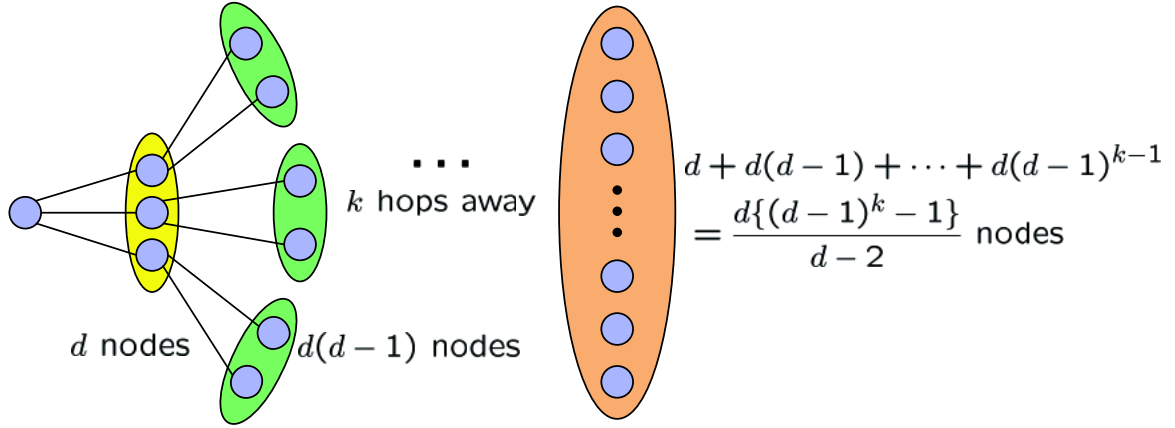
Figure 2-2: Maximum Number of Nodes Within k Hops of s

Note this argument is independent of the selection of a graph and thus the above inequality holds for every d-regular graph. Therefore,

$$
\begin{aligned}
\mathrm{E}(L) &= \sum_{k=1}^{n-1}(1 - \Pr(L \le k)) \\
&\ge \sum_{k=1}^{\lceil \log_{d-1}\{(\frac{d-2}{d})(n-1)+1\}\rceil}(1 - \frac{d\{(d-1)^k - 1\}}{d-2} \cdot \frac{1}{n-1}) \\
&\ge \sum_{k=1}^{\lceil \log_{d-1} n \rceil}(1 - \frac{d\{(d-1)^k - 1\}}{d-2} \cdot \frac{1}{n-1}) \\
&= \lceil \log_{d-1} n \rceil - (\frac{d}{d-2})(\frac{1}{n-1})\left(\frac{d-1\{(d-1)^{\lceil \log_{d-1} n \rceil} - 1\}}{d-2} - \lceil \log_{d-1} n \rceil\right) \\
&\sim \log_{d-1} n, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2.2)
\end{aligned}
$$

where we assume that n is large.

2.2.2 Approximate Average Distance

In [23], Newman *et al.* calculate z_2, the asymptotic average number of neighbors at distance two from a randomly chosen vertex, in a large random graph with an arbitrary degree distribution[2]. Further, they generalize the argument to compute the

---

[2]We will discuss the details of their method in the next chapter, where we generalize our model so that different degrees are allowed. Because in this chapter we assume that the degree of each node is the same, detailed description of the degree distribution may not be relevant.

asymptotic average number of mth-nearest neighbors and obtain the relation

$$z_m = \left(\frac{z_2}{z_1}\right)^{m-1} z_1,$$

where z_1 is the average number of the adjacent neighbors of a randomly chosen vertex. Then they give an asymptotic estimate of the typical length $\hat{L}$ of the shortest path between two randomly chosen vertices as follows:

$$1 + \sum_{m=1}^{\hat{L}} z_m = n,$$

where n denotes the total number of vertices in the graph. Solving these equations, they present an asymptotic estimate of the average distance $\hat{L}$ such that

$$\hat{L} = \frac{\log[(n-1)(z_2 - z_1) + z_1^2] - \log z_1^2}{\log(z_2/z_1)}. \tag{2.3}$$

In the case of d-regular graphs, z_1 and z_2 are simply d and d^2, respectively. Hence, from (2.3), we obtain

$$\hat{L} = \frac{\log[(n-1)(d^2 - d) + d^2] - \log d^2}{\log d}$$

$$\sim \log_d n.$$

The authors of [23] note that this approximation may not be correct if all the vertices are not reachable from a randomly chosen vertex. However, if $d \geq 3$, we know that $G(n,d)$ is *a.a.s.* d-connected, and hence, we can expect that the above approximation becomes tight as n tends to infinity.

If we compare this to the lower bound (2.2) that we derived in the previous section, we find that the difference between the two is only in the base of the logarithm, which is a constant factor and becomes negligible as n tends to infinity. Hence, we can conclude that our lower bound (2.2) is, in fact, close to existing estimates. This can be viewed as an indication of its tightness.

2.3   Shortest Cycle

Recall that cycles are of our interest because primary and backup paths together form a cycle in a graph. In this section, we also consider a randomly picked pair of nodes, and now we define the random variable X as the length of the shortest cycle including the pair. Note that, since $G(n, d)$ is simple and hence, does not allow cycles of length 1 or 2, $X \geq 3$. Moreover, because $G(n, d)$ *a.a.s.* has a Hamiltonian cycle (Theorem 2.1.4), based on our assumption that $n \to \infty$, X is well-defined and $X \leq n$.

We let Y_k be the event that the pair is on a k-cycle (cycle of length k), i.e., there exists a k-cycle around the pair. Also, based on the assumption that n tends to infinity, we can easily see that $\Pr(Y_3 \cup Y_4 \cup ... \cup Y_n) = 1$.

Note that, by the definition of X, $X \leq k$ implies the pair is on a certain cycle no longer than k and we find that

$$\Pr(Y_k) \leq \Pr(X \leq k) \leq \sum_{i=3}^{k} \Pr(Y_i), \qquad (2.4)$$

where we used the union bound for an upper bound. Therefore, we can lowerbound $\mathrm{E}(X)$ as follows:

$$
\begin{aligned}
\mathrm{E}(X) &= \sum_{k=3}^{n} k \Pr(X = k) \\
&= \sum_{k=3}^{m-1} k \{\Pr(X \leq k) - \Pr(X \leq k-1)\} + \sum_{k=m}^{n} k \Pr(X = k) \\
&\geq \sum_{k=3}^{m-1} k \{\Pr(X \leq k) - \Pr(X \leq k-1)\} + m(1 - \Pr(X \leq m-1)) \qquad (2.5) \\
&\geq \sum_{k=3}^{m-1} k \{\Pr(Y_k) - \sum_{j=3}^{k-1} \Pr(Y_j)\} + m(1 - \sum_{j=3}^{m-1} \Pr(Y_j)) \qquad (2.6) \\
&= m - \sum_{k=3}^{m-1} \Pr(Y_k)\{\sum_{j=k+1}^{m} j - k\}, \qquad (2.7)
\end{aligned}
$$

where m is an integer, $4 \leq m \leq n$. Note in (2.5) that, for each k larger than m, we replaced $k \Pr(X = k)$ by $m \Pr(X = k)$ to get a lower bound, and that (2.5) follows

from (2.4). We find that, in (2.7), the term multiplied by $\Pr(Y_k)$, $-(\sum_{j=k+1}^{m} j - k)$, is negative for $k = 3, 4, ..., m - 1$. Therefore, if we obtain a lower bound on $\Pr(Y_k)$, we can further bound $\mathrm{E}(X)$ from below.

Now define an indicator random variable I_k taking 1 if the pair is on a k-cycle, and 0, otherwise. To calculate $\mathrm{E}(I_k)$, as mentioned above, we first condition on a certain graph and consider a pair selection on the graph, and then average the result over all graphs. More specifically, if we define Z_k be the number of k-cycles in a graph, conditioned on $Z_k = j$, we calculate conditional expectation of I_k by considering a random selection of a pair of nodes, which we average over all possible values of Z_k. Identifying $\mathrm{E}(I_k)$ as equivalent to $\Pr(Y_k)$, we can write this calculation as follows:

$$
\begin{aligned}
\Pr(Y_k) &= \mathrm{E}(I_k) \\
&= \sum_j \mathrm{E}(I_k | Z_k = j) \Pr(Z_k = j) \\
&= \sum_j \Pr(Y_k | Z_k = j) \Pr(Z_k = j),
\end{aligned}
\tag{2.8}
$$

where the expectation and probability conditioned on Z_k are over the probability space of the pair selection.

Let us consider how we can maximize the conditional probability $\Pr(Y_k | Z_k = j)$, i.e., the probability that the pair is on a k-cycle given that the graph has a certain number of k-cycles. If we assume there is a total of n nodes,

$$
\Pr(Y_k | Z_k = j) = \frac{\text{(total number of pair selections on k-cycle)}}{\binom{n}{2}}.
\tag{2.9}
$$

In order to calculate the maximum number of pair selections on a k-cycle, we first take the case of two cycles. If the two cycles are disjoint, i.e. they share no vertex, the number of such selections is $2\binom{k}{2}$. We obtain the same result when there is only one vertex shared by the two cycles. However, if the two cycles share j vertices, where $2 \leq j \leq k - 1$, then the number of pair selections on a k-cycle is $2\binom{k}{2} - \binom{j}{2}$, which is strictly less than that of the previous case. Hence, we get the maximum number of pair selections when the two cycles share no or only one vertex. Note that

33

by repeating this argument, the result easily extends to the case of more than two cycles. That is, if we have j cycles of length k, by assuming all the cycles are disjoint, we can maximize the number of pair selections on a k-cycle, which is given by $j\binom{k}{2}$. Hence, it follows from (2.8) and (2.9) that

$$\Pr(Y_k) \le \sum_j \frac{j\binom{k}{2}}{\binom{n}{2}} \Pr(Z_k = j)$$
$$= \frac{k(k-1)}{n(n-1)} \mathrm{E}(Z_k).$$

Now, recall that, as discussed in Section 2.1.1, we have an upper bound on $E(Z_k)$ for any k, $3 \le k \le n$, such that

$$\mathrm{E}(Z_k) \le \frac{(d-1)^k}{2k}. \tag{2.10}$$

Therefore,

$$\Pr(Y_k) \le \frac{(k-1)}{2n(n-1)}(d-1)^k. \tag{2.11}$$

Combining (2.7) and (2.11), we obtain

$$\mathrm{E}(X) \ge m - \sum_{k=3}^{m-1} \frac{(k-1)(d-1)^k}{2n(n-1)} \left\{ \sum_{j=k+1}^{m} j - k \right\}, \tag{2.12}$$

which is valid for any m, $4 \le m \le n$. We can calculate this lower bound numerically for various m. In Figure 2-3, we notice that the bound grows until some value of m, where we obtain the tightest lower bound, and then it starts to decrease as m grows further.

We can collect these lower bounds for each n, which Figure 2-4 plots with respect to $\log n$, for n up to 10^{30}. Interestingly, those bounds are shown to grow almost linearly with $\log n$, which is congruent to the lower bound (2.2) on the path length in the previous section.

Also, we can find a more analytical explanation of this. For fixed d and $m = m(n)$
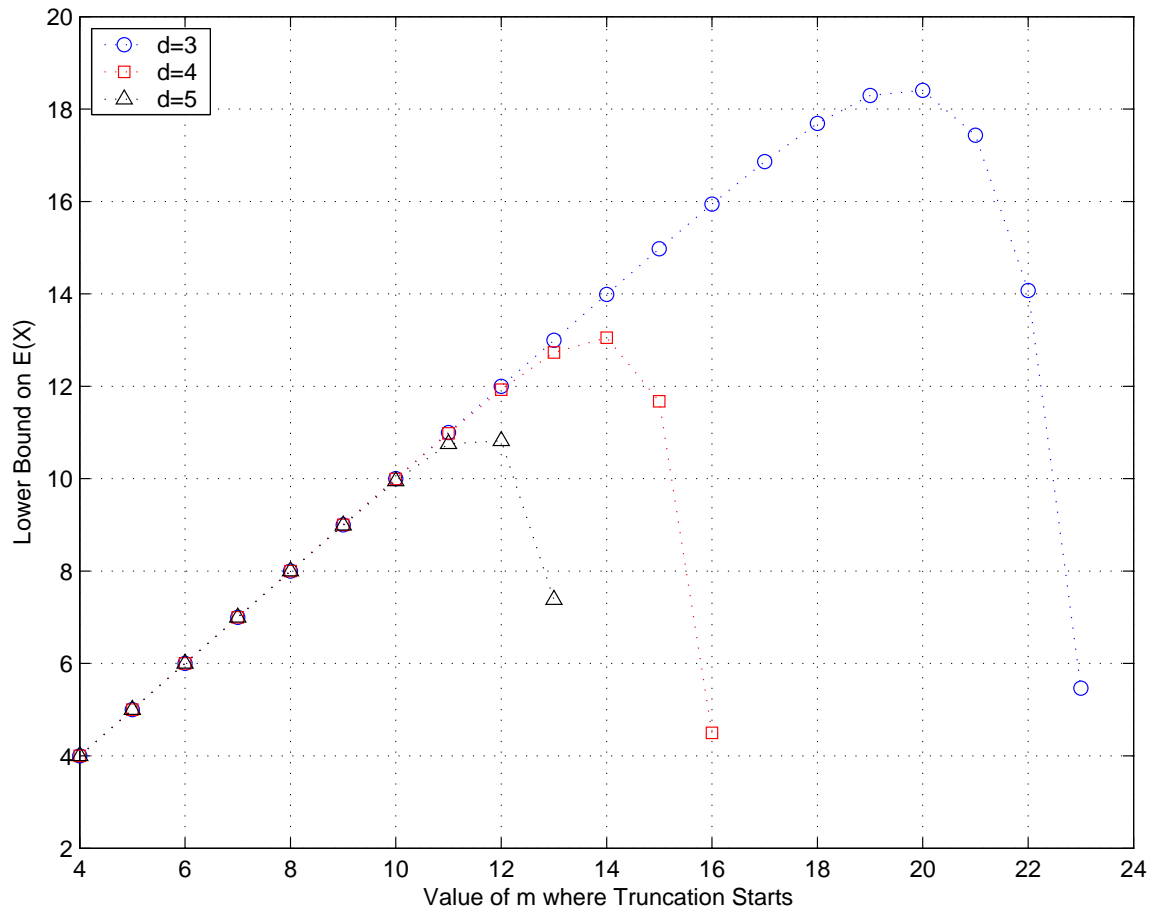
34

Figure 2-3: Lower Bound on $\mathrm{E}(X)$ with respect to m for $n = 10,000$

that grows with n, let B be the terms on the right-hand side in (2.12). Then,

$$
\begin{aligned}
B &= m + \sum_{k=3}^{m-1} \frac{(k-1)(d-1)^k}{2n(n-1)} \{k - \sum_{j=k+1}^{m} j\} \\
&= m - \frac{A}{n^2} \cdot \Theta\left(\sum_{k=3}^{m-1} k^3(d-1)^k + m^2 k(d-1)^k \right) \\
&= m - \frac{A}{n^2} \cdot \Theta(m^3(d-1)^m),
\end{aligned}
\tag{2.13}
$$

where A is a constant, $A > 0$. Note that, by examining the value of m that gives the tightest lower bound for each n, we can infer that the maximum may occur when m is approximately order of $\log n$. If we first let $m = c \log n$ for a constant $c > 0$, then $\Theta(m^3(d-1)^m) = \Theta((c \log n)^3 \cdot n^{c \log(d-1)})$. Hence, if $c < \frac{2}{\log(d-1)}$, then $B \sim c \log n$ since $\Theta(\frac{m^3(d-1)^m}{n^2}) \to 0$. Otherwise, if $c \geq \frac{2}{\log(d-1)}$, then B tends to below zero since the term $\Theta(\frac{m^3(d-1)^m}{n^2})$ dominates in B. Also, we can show that if $\frac{m}{\log n} \to 0$ or $\frac{m}{\log n} \to \infty$, then $B = \Theta(m)$ or $B \to -\infty$, respectively. Hence, we conclude that the best case is when $B = \Theta(\log n)$, which is the tightest lower bound on $\mathrm{E}(X)$.

2.4 Finite Length Cycle

Suppose we want to maintain the path lengths below a certain level in terms of the number of hops for the reasons mentioned in Section 1.1.1. Let a finite number l_{max} denote the maximum length of the paths allowed, and let us compute the probability that we can protect the network using only such paths.

2.4.1 Path Protection

Let us consider a path from s to t and keep it recoverable by the path protection scheme. To this end, there must exist a primary and a backup path, each of which does not exceed l_{max} but which together form a cycle (see Figure 2-5). Let us call a cycle with this property a *protection cycle*.

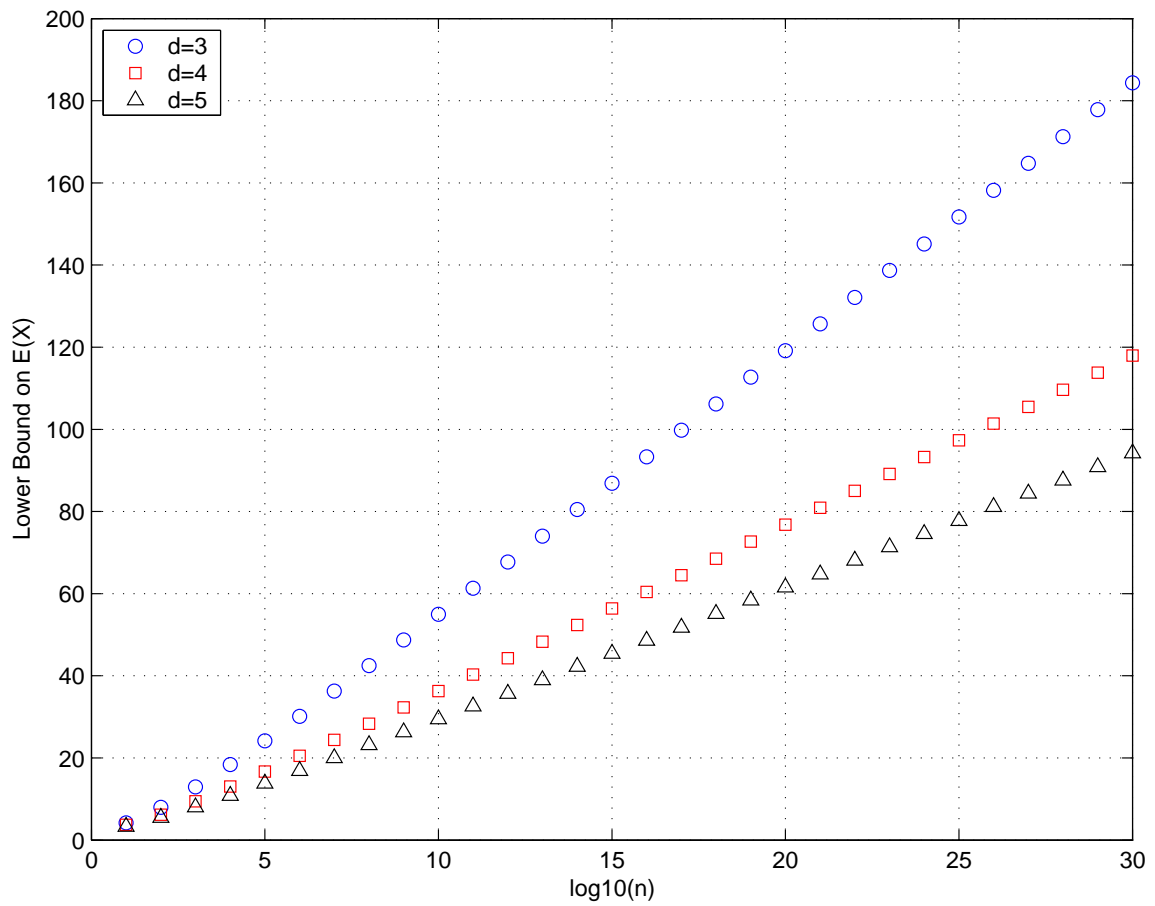Let C denote the set of all possible protection cycles including the pair and con-

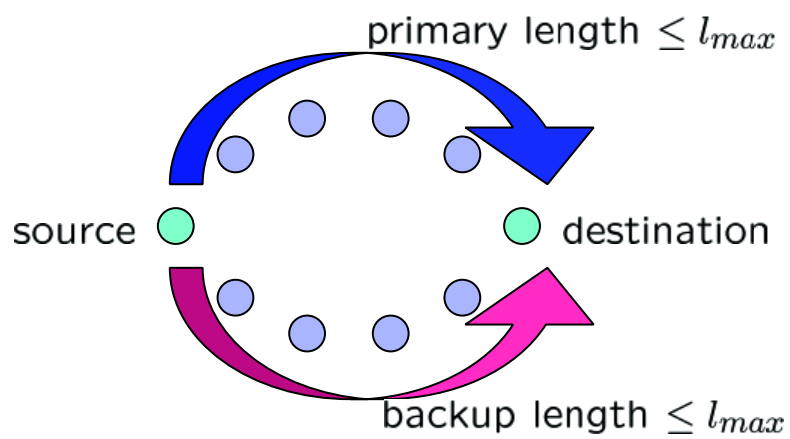Figure 2-4: Lower Bound on E(X) with respect to $\log n$

Figure 2-5: Protection Cycle for Path Protection

sider $E(|C|)$, i.e., the expected number of protection cycles. If, for any cycle c, we define an indicator random variable I_c taking 1 if c exists, and 0, otherwise, then

$$E(|C|) = E[\sum_{c \in C} I_c] = \sum_{c \in C} \Pr[\exists c]. \tag{2.14}$$

Note that any cycle of length k arises from a set of k edges in the corresponding configuration. Then we call such a set of k edges a k-cycle on the two-dimensional set $W = [n] \times [d]$. It is easy to see from the construction procedures of $G(n, d)$ that, for any k-cycle on W, the probability that it is contained in a random configuration is given the same, which we denote by p_k, i.e., it depends only on the number of edges. If a protection cycle c has length k, then $\Pr[\exists c]$ is given by the number of k-cycles on W corresponding to c multiplied by p_k. Therefore, we can calculate $E(|C|)$ by calculating p_k and the number of protection cycles of length k, and summing their product over all possible length k's.

Let us proceed to computing the probability p_k that any given set of k disjoint edges on the two-dimensional set W is contained in a random configuration. The total number of configurations is

$$\frac{\binom{dn}{2}\binom{dn-2}{2}...\binom{4}{2}\binom{2}{2}}{(dn/2)!} = \frac{(dn)!}{2^{(dn/2)}(dn/2)!} = (dn-1)!!,$$

where dn is even. Let us select k edges, or equivalently k pairs, on W and consider the configurations containing those edges. Since we have fixed $2k$ vertices, which are removed from the set of vertices to be paired, the number of configurations containing those k edges is $(rn - 2k - 1)!!$. Hence,

$$p_k = \frac{(dn - 2k - 1)!!}{(dn - 1)!!} = \frac{1}{(dn-1)(dn-3)...(dn-2k+1)},$$

and if k is fixed and $n \to \infty$, then

$$p_k \sim (dn)^{-k}. \tag{2.15}$$

38

Now, consider the number of protection cycles of length k, $3 \leq k \leq (l_{max} + 1)$ on W. Since we need $(k-2)$ intermediate nodes and allow any possible ordering of k nodes on the cycle, the number of possible protection cycles on W is

$$a_k = \binom{n-2}{k-2} \frac{(k-1)!}{2} (d(d-1))^k$$
$$\sim n^{k-2} \frac{(k-1)}{2} (d(d-1))^k, \qquad (2.16)$$

where $k = 3, ..., (l_{max} + 1)$. However, if $k \geq (l_{max} + 2)$, there exist some orderings on the cycle where s and t are located farther than l_{max} from each other, which we don't count. Hence,

$$a_k = \binom{n-2}{k-2} \frac{(k-2)!(2l_{max} - k + 1)}{2} (d(d-1))^k$$
$$\sim n^{k-2} \frac{(2l_{max} - k + 1)}{2} (d(d-1))^k, \qquad (2.17)$$

where $k = (l_{max} + 2), ..., 2l_{max}$. Therefore, using (2.16) and (2.17), we obtain

$$\mathrm{E}(|C|) = \sum_{k=3}^{2l_{max}} a_k p_k$$
$$\sim \sum_{k=3}^{l_{max}+1} \frac{(k-1)(d-1)^k}{2n^2} + \sum_{k=l_{max}+2}^{2l_{max}} \frac{(2l_{max} - k + 1)(d-1)^k}{2n^2}$$
$$= \sum_{k=3}^{2l_{max}} \frac{(d-1)^k}{2n^2} \min[k-1, 2l_{max} - k + 1].$$

If we consider the probability that there exists at least one protection cycle along the pair of nodes, it is bounded from above by $\mathrm{E}(|C|)$, which is a union bound including all possible protection cycles, and from below by the probability that there exists a cycle of length 3 on W. Hence,

$$\frac{1}{(dn)^3} \leq \Pr(\exists \text{protection cycle}) \leq \sum_{k=3}^{2l_{max}} \frac{(d-1)^k}{2n^2} \min[k-1, 2l_{max} - k + 1]. \qquad (2.18)$$

2.4.2 Link Protection

Now assume there is a link between s and t. To ensure that traffic between the pair is recoverable by the link protection scheme, there must exist a cycle not exceeding $(l_{max} + 1)$ around the pair. In this section, we let a protection cycle refer to a cycle of this property and again let C be the set of all possible protection cycles.

To get the expectation of the cardinality of C, again, we need to sum $a_k p_k$ over all possible k, the length of cycle. Note in this case that s and t should be adjacent to each other and thus the longest cycle possible is of length $(l_{max} + 1)$. Since

$$
\begin{aligned}
a_k &= \binom{n-2}{k-2} \frac{(k-2)!}{2} (d(d-1))^k \\
&\simeq n^{k-2} \frac{(d(d-1))^k}{2}, \qquad k = 3, \cdots, (l_{max}+1),
\end{aligned}
\tag{2.19}
$$

and p_k is given the same as before,

$$
\begin{aligned}
\mathrm{E}(|C|) &= \sum_{k=3}^{l_{max}+1} a_k p_k \\
&\simeq \sum_{k=3}^{l_{max}+1} \frac{(d-1)^k}{2n^2}.
\end{aligned}
\tag{2.20}
$$

Again, this is an upper bound on the probability that there exists at least one protection cycle along the pair of nodes. Therefore,

$$
\frac{1}{(dn)^3} \leq \mathrm{Pr}(\exists \text{protection cycle}) \leq \sum_{k=3}^{l_{max}+1} \frac{(d-1)^k}{2n^2}.
\tag{2.21}
$$

Note from the results above that, for both path and link protection schemes, the probability that we find a backup path of finite length decays in the order of $\frac{1}{n^2}$. In other words, in random networks described by the configuration model, it is highly unlikely to find a backup path within a finite range as the size of the network grows very large.

Chapter 3

Extension to General Networks

In the previous chapter, we used the configuration model as a method for constructing random networks. Despite its useful asymptotic properties, however, the model has a limitation in that the degree must be the same over all nodes. Hence, in this section, we develop an extended version of the configuration model, by which we can overcome the limitation. Our main goal in this extension is to relax the conditions on the degree so that the new model can describe more general networks, while also maintaining the overall framework of the original model so that major results carry over to the new model.

3.1   Extended Configuration Model

We recall that, in the configuration model for constructing a random d-regular graph with n vertices, we consider a two-dimensional set of $[d] \times [n]$ and partition the set into $(dn/2)$ pairs, then project onto the horizontal axis. A natural extension is that now we allow the degree to vary over a finite range and keep the remaining procedures the same as before. We describe in detail below the whole procedure of the extended model.

Suppose first that we are given a degree distribution for the graph. That is, if we let D_i be the degree of the ith node for $1, 2, ..., n$, we have a probability mass function that specifies the probability $\Pr(D_i = d)$ for each $d = 3, 4, ..., d_{max}$, where
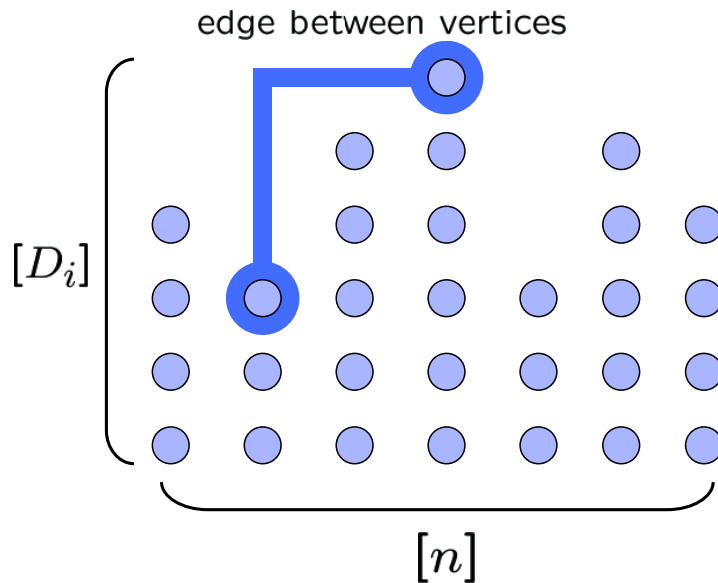
Figure 3-1: 2-Dimensional Set W

d_{max} is finite. Our goal is to construct a random graph whose degree follows the given distribution. Then, we proceed as follows.

- Determine a priori the degree of each node, D_i for $i = 1, 2, ..., n$, according to the given degree distribution. More specifically, we generate a random variable n times so that each D_i is identically independently distributed (i.i.d.) with the given probability mass function. If $m = \sum_{i=1}^{n} D_i$ is not even, we regenerate D_n until the sum becomes even.

- Consider a two-dimensional set $W = [D_i] \times [n]$ consisting of $m = \sum_{i=1}^{n} D_i$ vertices (see Figure 3-1).

- Choose two vertices randomly from W to make a pair. Continue this until we exhaust all the vertices, which is guaranteed because m is even. Hence, we obtain a random perfect matching, which we again name a random *configuration*.

- Project the two-dimensional set onto the horizontal axis by simply ignoring the vertical coordinate.

Again, the resulting graph may have self-loops around the same vertex or multiple edges between two vertices. Hence, we say the graph we constructed is a random

42

multigraph with the given degree distribution, and if we condition that there are no self-loops or multiple edges, then we obtain a random (simple) graph as desired.

3.1.1   Range of Degree

In the description of the extended model, we restricted ourselves to considering vertices of degree no less than three by setting the range of degree to be $d = 3, 4, ..., d_{max}$. In this section, we give a rationale for excluding vertices of degree less than three. Our claim is that, in considering the issue of protection, there is no effective change even if we ignore such vertices.

First, it is obvious that vertices of degree zero contribute nothing to a graph, and thus, we can simply delete such vertices with no changes to the topology of the graph. Or in a more practical sense, nodes of degree zero in a communication network, if any, are just isolated nodes unable to communicate with others. Therefore, there is no need to be concerned with such nodes in considering the operation of the network.

Proceeding to the case of degree one, again we take the interpretation of a graph as a communication network. If a vertex has degree one, then the corresponding node has only one adjacent node, through which further communication with other nodes must be done. To be precise, in Figure 3-2, node A has the only link l connected to node B, where we first assume that node B has degree at least three. We can easily see that any node of degree one such as node A cannot be located in the middle of a path, i.e., it must be an end node of a path.

If we consider the case of link protection, the only possible failure involved with node A is the failure of link l. However, it is obvious that there is no way for any other nodes to resume communicating with A until the failed link l is restored. In other words, if a link failure occurs to a node of degree one, there is no intelligent method to provide the node with resilience but to simply fix the failed link. Hence, we can say that, for link protection, we do not need to think of node A explicitly. On the other hand, in the case of path protection, we observe that any end-to-end path starting or ending at node A must pass through node B and include link l. Thus, it is easy to see that we cannot establish any two link-disjoint paths starting or ending
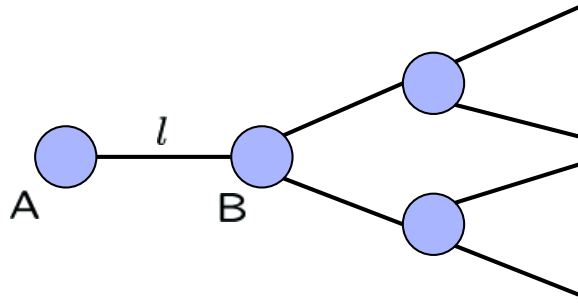
Figure 3-2: Node A of Degree One

at node A. Hence, there is also no need to consider node A for the case of path protection.

Now we think of the case where node B, which is adjacent to node A, has degree less than three. Obviously, it is impossible for node B to have degree zero because it already has A as an adjacent node. If node B has degree one, then nodes A and B form a subgraph isolated from other nodes, which is of no interest for further analysis and happens rarely enough if the overall graph is $a.a.s.$ connected. The only remaining case is where node B has degree two, which we will discuss next.

Also in the case of degree two, we assume that our graph represents a communication network. Suppose node B has degree two and let nodes A and C be the two adjacent nodes, where we assume both nodes A and C have degrees no less than three. Denote by l_1 the link between nodes A and B, and by l_2 the link between nodes B and C (see Figure 3-3). In the link protection case, if we assume that link l_1 fails, then the only remaining link that a protection path can use to reach node B is l_2 and thus, any such path should pass through node C. Hence, in order to establish a protection path for link l_1, we can equivalently establish one between nodes A and C, to which we later attach link l_2.

For the case of path protection, any end-to-end path that has node B as an endpoint must pass through nodes A or C and include either link l_1 or l_2, respectively. Moreover, if we consider a primary path including link l_1, then any link-disjoint protection path must include link l_2. Let us assume that we have an end-to-end primary path p_1 between node B and another node D, which passes through node A and thus includes link l_1. In order to find a protection path for p_1, we can consider
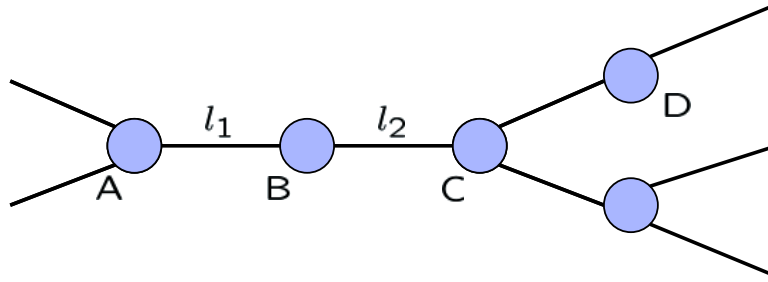
Figure 3-3: Node B of Degree Two

the primary path as being between nodes C and D, which includes both links l_2 and l_1, and then find a link-disjoint protection path between nodes C and D. Then the desired protection path is the above path extended by link l_2. Therefore, in either link or path protection, we can simply ignore node B of degree two at first, and after finding paths including node A and/or C, we can just add link l_1 and/or l_2.

Again let us consider the case where node A or C has degree less than three. It is clear that neither node A nor C can have degree zero. If at least one of nodes A and C has degree one, then without loss of generality, we can assume that node A has degree one. If C also has degree one, nodes A, B, and C form an isolated component, which we can ignore by the same reason we mentioned above. If C has degree two, then we can take the other node D adjacent to node C into account, and we can apply the same argument that we applied to the set of nodes A, B, C, now to the set of nodes A, B, C, D. We repeat this until we meet a node of degree one, or at least three, both of which fall into a case we mentioned above. The only remaining case is where both nodes A and C have degree two, and even if we extend on both sides, we never meet any node whose degree is not two. However, this happens only in a cycle that is isolated from other nodes and again we can ignore this for the same reason as above.

Hence, in considering the asymptotic behavior of the length of paths, we can ignore nodes of degree two and later, if needed, we can such nodes according to appropriate distributions, but this can be handled separately from our problem at hand.

We can find a more intuitive explanation for the exclusion of vertices of degree two [24]. As noticed in Figure 3-3, any vertex of degree two should fall in the middle of

edges between a different pair of vertices. Hence, even if we ignore a vertex of degree two and merge its two edges into one, there will be no changes in the topological structure of a graph.

Therefore, in remaining sections we set the range of degree as $d = 3, 4, ..., d_{max}$ in any graph of general degree distribution in the following.

3.1.2 Connectivity

Recall, in Section 2.1.1, we noted that connectivity of the graph is a critical issue and a random d-regular graph generated by the configuration model is *a.a.s.* d-connected for fixed $d \geq 3$. Hence, with the extended graph model of general degree distribution, we need to address the issue of connectivity as well.

The generalization of the case of regular graphs to that of graphs of general degree distribution can be obtained rather easily. In [35], Wormald takes a different approach to the issue of connectivity. To be precise, graphs with a given collection of degrees lying between r and R, $3 \leq r \leq R$, are considered first. Then it is shown that, for each possible set D of degrees within the region $[r, R]$, the probability that a graph with the set of degree D is r-connected tends to one as the number of vertices tends to infinity. In other words, such graphs are *a.a.s.* r-connected. By putting $r = R$, the theorem of connectivity of random regular graphs is obtained. In fact, the connectivity of graphs with an arbitrary set of degrees is shown as a preliminary step toward proving the connectivity of regular graphs.

Note that, in our model, we consider the case where degrees are random variables rather than fixed with a given set of values. However, since we assume that degrees are determined a priori, we can apply Wormald's result only if degrees lie between r and R, where $3 \leq r \leq R$. Hence, we conclude that graphs constructed by the extended configuration model are *a.a.s.* d_{min}-connected, where $d_{min} \geq 3$ is the minimum degree that each node can take.

3.1.3 Distribution of the Number of Cycles

Cycles of Fixed Length

We recall that, in random regular graphs described by the configuration model, the number of k-cycles is asymptotically Poisson distributed. (In fact, Theorem 2.1.3 provides a stronger argument that, for any set of different k_i's for $k_i \geq 3$, the corresponding numbers of k_i-cycles *jointly* converge in distribution to *independent* Poisson random variables.) In the proof of Theorem 2.1.3 in [19], we notice that the condition that each vertex has the same degree is not crucial and hence, we may easily extend the proof to the case where we allow different degrees. In the following, we present this extended proof by which we can generalize the argument that the distribution of the number of cycles is Poisson even in the case of networks that are not regular. It turns out this is a slight generalization of a work by Bollobás [5], which is used differently as part of the proof of another theorem. Though our proof shares the same idea with Bollobás's, we decide to present ours because it can give a comprehensive view of the extension of our whole argument in the previous chapter. Furthermore, we also consider long cycles whose lengths grow with the size of the network, while Bollobás [5] considers only cycles of fixed length.

For simplicity, we first let the degree take one of the two values according to independent Bernoulli trials. Then we will argue the same approach holds for more general cases, where the degree is allowed to vary over a larger range. Let V be the set of vertices $[n]$ corresponding to n places along the horizontal axis. For each place in V, we introduce D_i vertices on it, where D_i is i.i.d. random variable D' defined as

$$
D' = \begin{cases} d_1 & \text{with probability } \theta, \\ d_2 & \text{with probability } (1-\theta), \end{cases}
$$

for finite $d_1, d_2 \geq 3$, and fixed θ, $0 \leq \theta \leq 1$. Let n_1 denote the number of vertices with degree d_1 and n_2 with d_2. If we follow the steps presented in Section 3.1, we have a two-dimensional set of vertices $W = [n] \times [D_i]$, and $(d_1 n_1 + d_2 n_2)/2$ pairs on W. Then we project the set W onto $V = [n]$ and obtain a multigraph $\pi(F)$

as before. Let us denote this multigraph by $G_e^*(n, D')$ and the number of cycles of length k in $G_e^*(n, D')$ by Z_k^*. Then we show that, for fixed k, the random variable Z_k^* is asymptotically Poisson. In particular, we show the following theorem:

Theorem 3.1.1 *Let Z_k^* be the number of cycles of length k in $G_e^*(n, D')$. For each fixed j, a sequence of random variables $(Z_1^*, Z_2^*, ..., Z_j^*)$ converges in distribution to $(Z_{1\infty}, Z_{2\infty}, ..., Z_{j\infty})$, where $\{Z_{k\infty}\}_{k=1}^j$ is a sequence of independent Poisson distributed random variables with $E(Z_{k\infty}) = \frac{1}{2k}\left(\frac{d_1(d_1-1)\theta+d_2(d_2-1)(1-\theta)}{d_1\theta+d_2(1-\theta)}\right)^k$.*

Proof   Note that we use as a basis the proof of Theorem 9.5 in [19] (pp. 237-238), which deals with the multigraphs based on the original configuration model. Essentially, we show that the basic property of Poisson convergence does not change if we allow degrees to vary. If θ is 0 or 1, then it is the same as in the original configuration model, and, therefore, we assume $0 < \theta < 1$ and is fixed.

We start with calculating the first moment of Z_k^*. Let n_1 be the number of vertices with degree d_1 and n_2 with d_2. Conditioned on such n_1 and n_2, we let $p_{k|n_1,n_2}$ be the conditional probability that a given set of k disjoint edges on W is contained in a random configuration and $a_{k|n_1,n_2}$ be the number of k-cycles on the two-dimensional set W. Then $\mathrm{E}(Z_k^*|\, n_1, n_2)$, the expected number of k-cycles, is $a_{k|n_1,n_2}p_{k|n_1,n_2}$.

To compute $a_{k|n_1,n_2}$, we consider oriented cycles, with a specified initial vertex and a direction. Since each unoriented k-cycle corresponds to $2k$ oriented ones, the number of oriented k-cycles on W is $2ka_{k|n_1,n_2}$. On the other hand, an oriented k-cycle on W corresponds to a sequence of k distinct vertices, each of which has degree either d_1 or d_2, and for each vertex in the sequence, any two distinct indices can be selected among d_1 or d_2 indices, respectively. Equating these two expressions, we obtain

$$2ka_{k|n_1,n_2} = \sum_{j=0}^k \binom{n_1}{j}\binom{n_2}{k-j} k! \cdot (d_1(d_1-1))^j (d_2(d_2-1))^{k-j}$$

$$= \sum_{j=0}^k \binom{k}{j} \frac{n_1!}{(n_1-j)!} \frac{n_2!}{(n_2-(k-j))!} \cdot (d_1(d_1-1))^j (d_2(d_2-1))^{k-j}. \quad (3.1)$$

If $n_1 \to \infty$ while k is fixed, then for any j, $0 \le j \le k$,

$$\frac{n_1!}{(n_1 - j)!} = n_1(n_1 - 1)(n_1 - 2) \cdots (n_1 - j + 1)$$

$$\sim n_1^j,$$

and if $n_2 \to \infty$,

$$\frac{n_2!}{(n_2 - (k - j))!} = n_2(n_2 - 1)(n_2 - 2) \cdots (n_2 - (k - j) + 1)$$

$$\sim n_2^{k-j}$$

Hence, if $n_1, n_2 \to \infty$ and k is fixed, the following approximation holds:

$$a_{k|n_1,n_2} \sim \frac{1}{2k} \sum_{j=0}^{k} \binom{k}{j} n_1^j n_2^{k-j} \cdot (d_1(d_1 - 1))^j (d_2(d_2 - 1))^{k-j}$$

$$= \frac{1}{2k} (d_1(d_1 - 1)n_1 + d_2(d_2 - 1)n_2)^k. \tag{3.2}$$

Now, if we let $n = n_1 + n_2 \to \infty$, one of either n_1 or n_2, but not both, may remain finite, while the sum of them tends to infinity. If we assume, without loss of generality, that n_1 remains finite, then we can simply ignore the vertices of degree d_1 in the limit and thus $a_{k|n_1,n_2}$ is given by

$$a_{k|n_1,n_2} \sim \frac{1}{2k} (d_2(d_2 - 1)n_2)^k.$$

Note also in (3.2), if n_1 remains finite while n_2 tends to infinity, then

$$a_{k|n_1,n_2} \sim \frac{1}{2k} (d_1(d_1 - 1)n_1 + d_2(d_2 - 1)n_2)^k$$

$$\sim \frac{1}{2k} (d_2(d_2 - 1)n_2)^k. \tag{3.3}$$

Hence, if $n \to \infty$ and k is fixed, we can say that the approximation (3.2) holds regardless of whether n_1 or n_2 remains finite. Note, however, that this case is negligible since, by the strong law of large numbers, n_1 and n_2 tend to $n\theta$ and $n(1 - \theta)$

49

respectively with probability 1, both of which tend to infinity.

Similarly as in (2.15), the probability $p_{k|n_1,n_2}$ that a set of k distinct edges is contained in a random configuration is given by

$$
\begin{aligned}
p_{k|n_1,n_2} &= \frac{((d_1 n_1 + d_2 n_2) - 2k - 1)!!}{((d_1 n_1 + d_2 n_2) - 1)!!} \\
&= \frac{1}{((d_1 n_1 + d_2 n_2) - 1)((d_1 n_1 + d_2 n_2) - 3) \cdots ((d_1 n_1 + d_2 n_2) - 2k + 1)} \\
&\sim (d_1 n_1 + d_2 n_2)^{-k}.
\end{aligned} \tag{3.4}
$$

Consequently, we obtain the following conditional expectation

$$
\begin{aligned}
\mathrm{E}(Z_k^* \mid n_1, n_2) &= a_{k|n_1,n_2} p_{k|n_1,n_2} \\
&\sim \frac{1}{2k} \left(\frac{d_1(d_1 - 1)n_1 + d_2(d_2 - 1)n_2}{d_1 n_1 + d_2 n_2} \right)^k.
\end{aligned}
$$

Next, in order to get the unconditional expectation of Z_k^*, take the expectation of the equation above over n_1 and n_2. Note that, by definition, $\mathrm{E}(n_1) = n\theta$ and $\mathrm{E}(n_2) = n(1 - \theta)$ and therefore, we obtain the unconditional expectation

$$
\mathrm{E}(Z_k^*) \sim \frac{1}{2k} \left(\frac{d_1(d_1 - 1)\theta + d_2(d_2 - 1)(1 - \theta)}{d_1 \theta + d_2(1 - \theta)} \right)^k = \lambda_k.
$$

Note that, by a similar argument, if H is any multigraph with v vertices and e edges, then $a_k = O(n^v)$ and $p_k = O(n^{-e})$. Hence, the expected number of copies of H in $G_e^*(n, D')$ is $O(n^v) \cdot O(n^{-e}) = O(n^{v-e})$. Moreover, if H has more edges than vertices, i.e., $v < e$, then the expected number of copies of H is $O(n^{-1})$.

We proceed to calculate factorial moments. We start with $\mathrm{E}(Z_k^*)_2$, the number of ordered pairs of two distinct k-cycles in $G_e^*(n, D')$, and indicate later the small modifications needed in the general case. Since the two cycles may or may not intersect, we can write $(Z_k^*)_2 = Y' + Y''$, where Y' is the number of ordered pairs of (vertex) disjoint k-cycles, while Y'' is that of distinct k-cycles having at least one common vertex.

The number Y'' can be further decomposed according to the number of common

vertices and edges, and their relative positions. This expresses Y'' as the sum of terms Y_j'', where each term counts the number of copies of a multigraph H_j in $G_e^*(n, D')$. Note that the number of terms depends on k but not on n, and that each H_j is the union of two distinct cycles with at least one common vertex and, therefore, it is connected and has more than one cycle. Each such H_j has more edges than vertices and thus, by the result just shown, $\mathrm{E}(Y_j'') = O(n^{-1})$ for each j. Summing over all j, we obtain as in (3.1) and (3.2)

$$\mathrm{E}(Y'') = O(n^{-1}).$$

Therefore the main term is $\mathrm{E}(Y')$, which we compute in the same way as $\mathrm{E}(Z_k^*)$. If we denote by $a_{kk|n_1,n_2}$ the number of ordered pairs of possible k-cycles on W that project to disjoint cycles on V, by orienting both cycles, we obtain

$$
\begin{aligned}
(2k)^2 a_{kk|n_1,n_2} &= \sum_{j=0}^{2k} \binom{2k}{j} \frac{n_1!}{(n_1 - j)!} \frac{n_2!}{(n_2 - (2k - j))!} \cdot (d_1(d_1 - 1))^j (d_2(d_2 - 1))^{2k-j} \\
&\sim (d_1(d_1 - 1)n_1 + d_2(d_2 - 1)n_2)^{2k} \\
&\sim (2k a_{k|n_1,n_2})^2.
\end{aligned}
$$

The probability that a given pair of disjoint k-cycles on W is contained in a random configuration is $p_{2k|n_1,n_2} \sim (d_1 n_1 + d_2 n_2)^{-2k}$, and consequently,

$$\mathrm{E}(Y') = a_{kk|n_1,n_2} p_{2k|n_1,n_2} \sim (a_{k|n_1,n_2} p_{k|n_1,n_2})^2 \sim \lambda_k^2.$$

Therefore, we obtain

$$\mathrm{E}(Z_k^*)_2 = \mathrm{E}(Y') + \mathrm{E}(Y'') \sim \lambda_k^2 + o(1).$$

The same argument applies to any factorial moment $\mathrm{E}(Z_k^*)_m$, and more generally to any joint factorial moment $\mathrm{E}[(Z_1)_{m_1}(Z_2)_{m_2} \cdots (Z_l)_{m_l}]$, where we consider the number of sequences of $(m_1 + m_2 + \cdots + m_l)$ distinct cycles such that the first m_1 have

length 1, the next m_2 have length 2, and so on. More specifically, as before, we write this number as $Y' + Y''$, and we can again show $E(Y'') = O(n^{-1})$. For Y', the same argument as above yields $E(Y') \sim \lambda_1^{m_1} \lambda_2^{m_2} \cdots \lambda_l^{m_l}$, and hence

$$E[(Z_1)_{m_1}(Z_2)_{m_2} \cdots (Z_l)_{m_l}] \sim \lambda_1^{m_1} \lambda_2^{m_2} \cdots \lambda_l^{m_l}.$$

By the theorem of Poisson limits (Theorem 6.10 in [19]), this implies that the joint distribution of $Z_1, \cdots, Z_l$ converges to the joint distribution of $Z_{1\infty}, \cdots, Z_{l\infty}$, and this completes the proof. $\square$

Recall that $G_e^*(n, D')$ is simple if and only if $Z_1^* = Z_2^* = 0$. Hence, $G_e^*(n, D')$ conditioned on $Z_1^* = Z_2^* = 0$ yields $G_e(n, D')$ which is defined as a random simple graph following the given degree distribution. Note also that, from the above theorem, each Z_k^* converges to an independent Poisson random variable. Therefore, by conditioning on $Z_1^* = Z_2^* = 0$, we obtain the following theorem of the distribution of the number of cycles in the case of the extended model:

Theorem 3.1.2 *Let Z_k be the number of cycles of length k in $G_e(n, D')$. For each fixed j, a sequence of random variables $(Z_3, Z_4, ..., Z_j)$ converges in distribution to $(Z_{3\infty}, Z_{4\infty}, ..., Z_{j\infty})$, where $\{Z_{k\infty}\}_{k=3}^{j}$ is a sequence of independent Poisson distributed random variables with $E(Z_{k\infty}) = \frac{1}{2k}\left(\frac{d_1(d_1-1)\theta + d_2(d_2-1)(1-\theta)}{d_1\theta + d_2(1-\theta)}\right)^k$.*

However, we find that, in the proof of Theorem 3.1.1, as long as the degree takes finite values, the variation of the degree does not affect the convergence property except through the mean value. Therefore, we can further extend this to the case where the degree D_i has an arbitrary distribution over a finite range.

In particular, suppose ith vertex has degree D_i, $i = 1, 2, ..., n$, each of which is defined to be an i.i.d. random variable D such that

$$\Pr(D = d_j) = p_j \qquad j = 1, 2, ..., j_{max}, \qquad (3.5)$$

where we assume $3 \le d_1 < d_2 < \cdots < d_{j_{max}} < \infty$. Based on this degree distribution, the extended configuration model yields a random multigraph $G_e^*(n, D)$. Let Z_k^* be

the number of k-cycles in $G_e^*(n, D)$, and let n_j be the number of vertices of degree d_j for $j = 1, 2, ..., j_{max}$. Now, conditioned on such $\{n_j\}_{j=1}^{j_{max}}$, let us calculate the number of k-cycles on the two-dimensional set W for fixed k, which we denote by $a_{k|\{n_j\}}$. By a similar argument as in (3.1), an oriented k-cycle on W corresponds to a sequence of k distinct vertices, and for each vertex in the sequence, any two distinct indices can be selected among D_i indices corresponding to it. Assume that we select k vertices among n vertices such that k_j vertices are from the set of n_j vertices of degree d_j, $j = 1, 2, ..., j_{max}$, for $k_j \geq 0$ and $k_1 + k_2 + \cdots + k_{j_{max}} = k$. Then we have

$$a_{k|\{n_j\}} = \frac{1}{2k} \sum_{k_1 + \cdots + k_{j_{max}} = k} \frac{k!}{k_1! k_2! \cdots k_{j_{max}}!} \prod_{j=1}^{j_{max}} \frac{n_j!}{(n_j - k_j)!} (d_j(d_j - 1))^{k_j}. \qquad (3.6)$$

If $n_j \to \infty$, for $j = 1, 2, ..., j_{max}$, and k is fixed, then

$$\frac{n_j!}{(n_j - k_j)!} \sim n_j^{k_j}.$$

Hence, if each $n_j \to \infty$, for $j = 1, 2, ..., j_{max}$, we can approximate $a_{k|\{n_j\}}$ by

$$a_{k|\{n_j\}} \sim \frac{1}{2k} \sum_{k_1 + \cdots + k_{j_{max}} = k} \frac{k!}{k_1! k_2! \cdots k_{j_{max}}!} \prod_{j=1}^{j_{max}} n_j^{k_j} (d_j(d_j - 1))^{k_j}$$

$$= \left(\sum_{j=1}^{j_{max}} d_j(d_j - 1) n_j \right)^k. \qquad (3.7)$$

Also, as argued in the derivation of (3.3), even if some n_j's remain finite while $n = (n_1 + n_2 + \cdots + n_{j_{max}}) \to \infty$, the above approximation still holds since the corresponding terms can be ignored in the limit. However, note again that this case is negligible since, by the strong law of large numbers, each $n_j \to np_j$ with probability 1 and this tends to infinity.

Similarly as in (3.4), $p_{k|\{n_j\}}$, the conditional probability that a set of k distinct

edges is contained in a random configuration, is given by

$$p_{k|\{n_j\}} = \frac{(\sum_{j=1}^{j_{max}} d_j n_j - 2k - 1)!!}{(\sum_{j=1}^{j_{max}} d_j n_j - 1)!!} \tag{3.8}$$

$$= \frac{1}{\prod_{i=1}^{k}\{\sum_{j=1}^{j_{max}} d_j n_j - (2i - 1)\}}$$

$$\sim \left(\sum_{j=1}^{j_{max}} d_j n_j\right)^{-k}. \tag{3.9}$$

Hence, conditioned on $n_1, n_2, ..., n_{j_{max}}$, the numbers of k-cycles in $G_e^*(n, D)$ are *a.a.s.* independent and jointly Poisson distributed with mean

$$\mathrm{E}(Z_k^* | \, n_1, n_2, ..., n_{j_{max}}) = a_{k|\{n_j\}} p_{k|\{n_j\}} \sim \frac{1}{2k}\left(\frac{\sum_{j=1}^{j_{max}} d_j(d_j - 1)n_j}{\sum_{j=1}^{j_{max}} d_j n_j}\right)^k,$$

for any set of fixed k's, $k \geq 3$. If we average this over $n_1, n_2, ..., n_{j_{max}}$ according to the distribution of D, we find that the unconditional expectation is

$$\mathrm{E}(Z_k^*) = \mathrm{E}[\mathrm{E}(Z_k^* | \, n_1, n_2, ..., n_{j_{max}})]$$

$$= \frac{1}{2k}\left(\frac{\sum_{j=1}^{j_{max}} d_j(d_j - 1)p_j}{\sum_{j=1}^{j_{max}} d_j p_j}\right)^k$$

$$= \frac{1}{2k}\left(\frac{\mathrm{E}(D^2) - \mathrm{E}(D)}{\mathrm{E}(D)}\right)^k. \tag{3.10}$$

The argument for higher order moments in the proof of Theorem 3.1.1 can be extended in the same manner, and this establishes the following theorem:

Theorem 3.1.3 *Let Z_k^* be the number of cycles of length k in $G_e^*(n, D)$. For each fixed j, a sequence of random variables $(Z_1^*, Z_2^*, ..., Z_j^*)$ converges in distribution to $(Z_{1\infty}, Z_{2\infty}, ..., Z_{j\infty})$, where $\{Z_{k\infty}\}_{k=1}^{j}$ is a sequence of independent Poisson distributed random variables with $E(Z_{k\infty}) = \frac{1}{2k}(\frac{E(D^2) - E(D)}{E(D)})^k$.*

Again, since $G_e^*(n, D)$ conditioned on $Z_1^* = Z_2^* = 0$ yields $G_e(n, D)$ and each Z_k^* converges to independent Poisson random variable, from the above theorem by conditioning on $Z_1^* = Z_2^* = 0$, we obtain the following theorem:

Theorem 3.1.4 *Let Z_k be the number of cycles of length k in $G_e(n, D)$. For each fixed j, a sequence of random variables $(Z_3, Z_4, ..., Z_j)$ converges in distribution to $(Z_{3\infty}, Z_{4\infty}, ..., Z_{j\infty})$, where $\{Z_{k\infty}\}_{k=3}^{j}$ is a sequence of independent Poisson distributed random variables with $E(Z_{k\infty}) = \frac{1}{2k}\left(\frac{E(D^2)-E(D)}{E(D)}\right)^k$.*

We find that the crucial characteristic is the second moment of degree over the first moment, $\frac{\mathrm{E}(D^2)-\mathrm{E}(D)}{\mathrm{E}(D)}$. Hence, while maintaining the mean degree the same, if we increase the variance of degrees, then the mean number of cycles will increase. On the other hand, two random graphs with completely different distributions of degrees, but the same values of the second moment of degree over the first moment, will have asymptotically the same distribution of cycles.

Long Cycles

Note that Theorem 3.1.3 and 3.1.4 hold only where the lengths of cycles are fixed. Now, let us consider the case where $k = k(n)$ is a function of n, i.e., the length of cycle grows with n. Recall that, in Section 2.1.1, for a random regular graph $G(n, d)$, we argued that the expression of $\mathrm{E}(Z_k)$ for fixed k is an upper bound on $\mathrm{E}(Z_k)$ for general k (see Lemma 2.1.6 and Figure 2-1). Now, we will show that the same holds for the extended model. That is, in $G_e(n, D)$, (3.10) is an upper bound on $\mathrm{E}(Z_k)$ such that

$$\mathrm{E}(Z_k) \leq \frac{1}{2k}\left(\frac{\mathrm{E}(D^2)-\mathrm{E}(D)}{\mathrm{E}(D)}\right)^k, \tag{3.11}$$

where $k = k(n)$ grows with n.

Since it is easy to show by conditioning that the mean number of cycles is the same either in a (simple) graph $G_e(n, D)$ or in a multigraph $G_e^*(n, D)$, we will assume that the number of cycles in the analysis below represents $\mathrm{E}(Z_k)$ in $G_e(n, D)$.

From (3.6) and (3.9),

$$\mathrm{E}(Z_k|\,n_1, n_2, ..., n_{j_{max}})$$

$$= a_{k|\{n_j\}} p_{k|\{n_j\}}$$

$$= \frac{1}{2k} \sum_{k_1 + \cdots + k_{j_{max}} = k} \frac{k!}{k_1! k_2! \cdots k_{j_{max}}!} \prod_{j=1}^{j_{max}} \frac{n_j!}{(n_j - k_j)!} (d_j(d_j - 1))^{k_j}$$

$$\times \frac{(\sum_{j=1}^{j_{max}} d_j n_j - 2k - 1)!!}{(\sum_{j=1}^{j_{max}} d_j n_j - 1)!!}. \tag{3.12}$$

Note that the above expression is exact.

Our objective is to show that

$$\left(\prod_{j=1}^{j_{max}} \frac{n_j!}{(n_j - k_j)!} \right) \frac{(\sum_{j=1}^{j_{max}} d_j n_j - 2k - 1)!!}{(\sum_{j=1}^{j_{max}} d_j n_j - 1)!!} \leq \frac{\prod_{j=1}^{j_{max}} n_j^{k_j}}{(\sum_{j=1}^{j_{max}} d_j n_j)^k}, \tag{3.13}$$

because once we show this, it is easy to see from (3.12) that

$$\mathrm{E}(Z_k|\,n_1, n_2, ..., n_{j_{max}}) \leq \frac{1}{2k} \left(\frac{\sum_{j=1}^{j_{max}} d_j(d_j - 1) n_j}{\sum_{j=1}^{j_{max}} d_j n_j} \right)^k,$$

where the expression on the right-hand side is the conditional expectation $\mathrm{E}(Z_k|n_1, n_2, ..., n_{j_{max}})$ for fixed k, and by averaging both sides over $\{n_j\}_{j=1}^{j_{max}}$ we obtain (3.11).

Assume again that we select k vertices among n vertices such that k_j vertices are from the set of n_j vertices of degree d_j, $j = 1, 2, ..., j_{max}$, for $k_j \geq 0$ and $k_1 + k_2 + \cdots + k_{j_{max}} = k$. By applying Stirling's formula,

$$\frac{n_j!}{(n_j - k_j)!} \approx \frac{\sqrt{2\pi n_j}(n_j/e)^{n_j}}{\sqrt{2\pi(n_j - k_j)}((n_j - k_j)/e)^{n_j - k_j}} \tag{3.14}$$

$$= \frac{1}{(1 - k_j/n_j)^{n_j - k_j + 1/2}} \cdot \frac{n_j^{k_j}}{e^{k_j}}$$

$$\approx \frac{1}{(1 - k_j/n_j)^{n_j - k_j}} \cdot \frac{n_j^{k_j}}{e^{k_j}} \tag{3.15}$$

$$= \frac{1}{\exp\{(n_j - k_j) \ln(1 - \frac{k_j}{n_j})\}} \cdot \frac{n_j^{k_j}}{e^{k_j}}. \tag{3.16}$$

Note that the approximation in (3.14) comes from the fact that we ignored $O(1/n)$ terms in Stirling's formula. However, this can be negligible since, by strong law of large numbers, each n_j tends to infinity with probability 1, and as long as k_j is a sublinear function of n_j, $(n_j - k_j)$ tends to infinity as well, which also justifies the approximation in (3.15). For double factorials, we use the following approximation used by Garmo [15]:

Lemma 3.1.5 *Let n be an integer. Then,*

$$(n-1)!! = \sqrt{2}n^{n/2}e^{-n/2}\left[1 - \frac{1}{12n} + \frac{1}{288n^2} + O\left(\frac{1}{n^3}\right)\right].$$

For notational simplicity, we let $S = \sum_{j=1}^{j_{max}} d_j n_j$. Then, by Lemma 3.1.5,

$$\frac{(\sum_{j=1}^{j_{max}} d_j n_j - 2k - 1)!!}{(\sum_{j=1}^{j_{max}} d_j n_j - 1)!!} \approx \frac{\sqrt{2}\{S - 2k\}^{(S/2-k)} \cdot e^{(-S/2+k)}}{\sqrt{2}S^{(S/2)} \cdot e^{(-S/2)}}$$

$$= \left(1 - \frac{2k}{S}\right)^{(S/2-k)} \cdot \frac{e^k}{S^k}$$

$$= \exp\{(S/2 - k)\ln(1 - 2k/S)\} \cdot \frac{e^k}{S^k}, \qquad (3.17)$$

where we ignored $O(\frac{1}{n})$ or smaller terms by the same reason as before. Now, from (3.16) and (3.17), and using the identity $k_1 + k_2 + \cdots + k_{j_{max}} = k$, we find that, by strong law of large numbers, with probability 1 the left-hand side of (3.13) converges to

$$\frac{\exp\{(S/2 - k)\ln(1 - 2k/S)\}}{\exp\{\sum_{j=1}^{j_{max}}(n_j - k_j)\ln(1 - \frac{k_j}{n_j})\}} \cdot \frac{\prod_{j=1}^{j_{max}} n_j^{k_j}}{S^k}. \qquad (3.18)$$

Substituting (3.18) into (3.13), we obtain the following inequality which is equivalent to our objective:

$$\frac{\exp\{(S/2 - k)\ln(1 - 2k/S)\}}{\exp\{\sum_{j=1}^{j_{max}}(n_j - k_j)\ln(1 - \frac{k_j}{n_j})\}} \cdot \frac{\prod_{j=1}^{j_{max}} n_j^{k_j}}{S^k} \leq \frac{\prod_{j=1}^{j_{max}} n_j^{k_j}}{(\sum_{j=1}^{j_{max}} d_j n_j)^k}. \qquad (3.19)$$

57

Then, since the term $\frac{\prod_{j=1}^{jmax} n_j^{k_j}}{S^k} = \frac{\prod_{j=1}^{jmax} n_j^{k_j}}{(\sum_{j=1}^{jmax} d_j n_j)^k}$ is common on both sides of (3.19), our objective is now reduced to showing that

$$\exp\left\{ (S/2 - k)\ln(1 - 2k/S) - \sum_{j=1}^{jmax} (n_j - k_j)\ln(1 - \frac{k_j}{n_j}) \right\} \leq 1, \qquad (3.20)$$

or equivalently, that the exponent is negative such that

$$(\frac{S}{2} - k)\ln(1 - \frac{2k}{S}) \leq \sum_{j=1}^{jmax} (n_j - k_j)\ln(1 - \frac{k_j}{n_j}). \qquad (3.21)$$

It is clear that $S > 2n$ since $S = \sum_{j=1}^{jmax} d_j n_j$ and each $d_j \geq 3$, where $2n$ is obtained by replacing each d_j by 2. Also, since k, the length of cycle, is clearly less than n, we find that $\frac{S}{2} > k$. Moreover, if we let $f(x) = (x - k)\ln(1 - \frac{k}{x})$ for given k, then it is easy to show that $f(x)$ is a non-increasing function for $x > k \geq 0$; $f'(x) = \ln(1 - \frac{k}{x}) + \frac{k}{x} \leq 0$. Hence, we have

$$(\frac{S}{2} - k)\ln(1 - \frac{2k}{S}) \leq (n - k)\ln(1 - \frac{k}{n})$$
$$= (\sum_{j=1}^{jmax} n_j - k_j)\ln\left(1 - \frac{\sum_{j=1}^{jmax} k_j}{\sum_{j=1}^{jmax} n_j}\right). \qquad (3.22)$$

Comparing (3.22) and the right-hand side of (3.21), we obtain the following result by log sum inequality [9]:

$$(\sum_{j=1}^{jmax} n_j - k_j)\ln\left(1 - \frac{\sum_{j=1}^{jmax} k_j}{\sum_{j=1}^{jmax} n_j}\right) \leq \sum_{j=1}^{jmax} (n_j - k_j)\ln(1 - \frac{k_j}{n_j}), \qquad (3.23)$$

which shows that (3.21), or equivalently (3.13), holds. Therefore, we have shown that the desired upper bound (3.11) on $E(Z_k)$ is valid for general k, i.e., the asymptotic mean of Z_k for fixed k in $G_e(n, D)$ is an upper bound on $E(Z_k)$ for general k.

3.2 Shortest Path

Throughout the remaining sections of this chapter, we assume that we are given an arbitrary degree distribution D satisfying the properties discussed in the previous section, and that our network is a large random graph $G_e(n, D)$ generated by the extended configuration model. Suppose n, the number of nodes, is large enough so that all the asymptotic properties in the previous section are assumed to hold, i.e., the deviation from the asymptotic behavior is assumed negligible.

In the previous chapter, we derived for random regular graphs a lower bound on the average distance between vertices. Also we addressed the fact that an approximate estimate of the average distance, given by Newman *et al.* [23], is close to our lower bound. In this section, we discuss how the approach of Newman *et al.* applies to graphs with general degree distributions, which may present much more interesting results than in the case of regular graphs.

Suppose we have a random graph $G_e(n, D)$, where the degree of each vertex follows the distribution of D, defined in (3.5). If we randomly pick a vertex s, then it is easy to see that the expected number of vertices adjacent to s is $z_1 = \sum_k k p_k = \mathrm{E}(D)$. To count the number of second neighbors, i.e., the number of vertices two hops away, we need to take two effects into account. First, a second neighbor can also be one of the first neighbors. However, in our graph model, the probability that one of the second neighbors is also a first neighbor decays as n^{-1}, and thus can be ignored in the limit of large n [24].

There is another effect, however, that we must consider: the degree distribution of a first neighbor of a randomly picked vertex is not the same as that of vertices on the graph as a whole [24]. This effect can be considered an example of the well-known issue of random incidence: in a renewal process, a randomly entered interval may be longer than the mean interarrival time because random incidence favors entry into large gaps [11]. Note, however, that this is not in the case of regular graphs. By a similar argument, if we randomly select an edge in a graph, there is a higher chance that the edge is connected to a high-degree vertex than to one with a low degree.

Moreover, we see that this probability is precisely proportional to the vertex's degree. Hence, if we let r be one of the first neighbors of s and define p'_k to be the degree distribution of r, then p'_k is given by

$$p'_k = \frac{kp_k}{\sum_j jp_j}.$$

Since it is obvious from the definition of r that s must be adjacent to r, the expected number of the first neighbors of r excluding s, which we denote by z', is given by

$$z_r = \sum_k kp'_k - 1 = \sum_k \left(\frac{k^2 p_k}{\sum_j jp_j} \right) - 1 = \frac{\mathrm{E}(D^2) - \mathrm{E}(D)}{\mathrm{E}(D)}. \tag{3.24}$$

Note that this is the average number of vertices two hops away from s via a particular first neighbor r. Multiplying this by the mean number of first neighbors of s, which is z_1, we obtain the mean number of second neighbors of s such that

$$z_2 = z_r \cdot z_1 = \frac{\mathrm{E}(D^2) - \mathrm{E}(D)}{\mathrm{E}(D)} \cdot \mathrm{E}(D) = \mathrm{E}(D^2) - \mathrm{E}(D). \tag{3.25}$$

The authors of [24] extend this calculation to further neighbors. The average number of vertices adjacent to each second neighbor, excluding the vertex we arrived along, is also given by (3.24), and indeed this is true at any distance m away from vertex s. Hence, the average number of neighbors at distance m from s is

$$z_m = \left(\frac{\mathrm{E}(D^2) - \mathrm{E}(D)}{\mathrm{E}(D)} \right) z_{m-1} = \left(\frac{\mathrm{E}(D^2) - \mathrm{E}(D)}{\mathrm{E}(D)} \right)^{m-1} \mathrm{E}(D). \tag{3.26}$$

Now, we can use the same argument given in Section 2.2.2. That is, when the total number of vertices within distance m is equal to n, the number of vertices in the whole graph, m is considered an approximate estimate of the average distance, which yields the following equation [23]:

$$1 + \sum_{m=1}^{\hat{L}} z_m = n. \tag{3.27}$$

Solving equations (3.26) and (3.27), we find that an asymptotic estimate $\hat{L}$ is

$$\begin{aligned}
\hat{L} &= \frac{\log[(n-1)(z_2 - z_1) + z_1^2] - \log z_1^2}{\log(z_2/z_1)} \\
&= \frac{\log[(n-1)(\mathrm{E}(D^2) - 2\mathrm{E}(D)) + \mathrm{E}(D)^2] - \log \mathrm{E}(D)^2}{\log[(\mathrm{E}(D^2) - \mathrm{E}(D))/\mathrm{E}(D)]},
\end{aligned} \tag{3.28}$$

and if $n \gg \mathrm{E}(D)$ and $\mathrm{E}(D^2) \gg \mathrm{E}(D)$, this reduces to

$$\hat{L} = \frac{\log(\frac{n}{\mathrm{E}(D)})}{\log(\frac{\mathrm{E}(D^2) - \mathrm{E}(D)}{\mathrm{E}(D)})} + 1. \tag{3.29}$$

Interestingly, the second moment of degree over the first moment, $\frac{\mathrm{E}(D^2) - \mathrm{E}(D)}{\mathrm{E}(D)}$, is also crucial to the average distance as in the mean number of cycles. Hence, this characteristic should be dealt with as a special parameter that determines the asymptotic behavior of both the average distance and the average number of cycles in random graphs.

3.3   Shortest Cycle

As in Section 2.3, we now consider a randomly picked pair of nodes and the random variable X, defined to be the length of the shortest cycle including the pair. It is clear that $X \geq 3$ since in $G_e(n, D)$ we do not allow cycles of length 1 or 2. We recall that, in random regular graphs with degree $d \geq 3$, there exists *a.a.s.* a Hamiltonian cycle (Theorem 2.1.4). However, for a general graph $G_e(n, D)$, it is not yet known whether a Hamiltonian cycle exists or not. Hence, in $G_e(n, D)$, X may not be well defined if there is no cycle including the pair, in which case we let X be infinite. If there is a pair of nodes corresponding to infinite X, then calculating the expectation of X over all pair selections is of no interest since it is simply infinite. Hence, we

need to restrict our graphs to those having a Hamiltonian cycle, which we hereafter assume.

Now, we show that the procedures to derive a lower bound on $E(X)$ for regular graphs apply here with slight modifications, which is easy to see because most of the results in the previous chapter do not depend on the fact that graphs are regular. For comprehensiveness, we recast important steps in the derivation, but also for conciseness, we omit the details if there is no major change in the argument.

If we let Y_k be the event that the pair is on a k-cycle, i.e., there exists a k-cycle around the pair, then, for any integer m, $4 \leq m \leq n$,

$$E(X) \geq m - \sum_{k=3}^{m-1} \Pr(Y_k)\{\sum_{j=k+1}^{m} j - k\}. \tag{3.30}$$

Also, we have the same expression as before for an upper bound on $\Pr(Y_k)$ such that

$$\Pr(Y_k) \leq \frac{k(k-1)}{n(n-1)} E(Z_k).$$

Since we have shown in the previous section that

$$E(Z_k) \leq \frac{1}{2k}\left(\frac{E(D^2) - E(D)}{E(D)}\right)^k,$$

we see now that, for graphs with a general degree distribution D,

$$\Pr(Y_k) \leq \frac{(k-1)}{2n(n-1)}\left(\frac{E(D^2) - E(D)}{E(D)}\right)^k. \tag{3.31}$$

Hence, from (3.30) and (3.31), we obtain the following lower bound on $E(X)$:

$$E(X) \geq m - \sum_{k=3}^{m-1} \frac{(k-1)(\frac{E(D^2)-E(D)}{E(D)})^k}{2n(n-1)}\{\sum_{j=k+1}^{m} j - k\}. \tag{3.32}$$

Then, by similar numerical calculation and convergence analysis as in Section 2.3, we can conclude that the above bound is $O(\log n)$.

62

3.4 Finite Length Cycle

Suppose we want to maintain the path lengths below a certain level in terms of the number of hops for the reasons mentioned in Section 1.1.1. Let a finite number l_{max} denote the maximum length of the paths allowed, and let us compute bounds on the probability that we can protect the network using only such paths. Again, if there is no major change in the argument compared with the case of regular graphs, we omit details and present only important steps.

3.4.1 Path Protection

We define a *protection cycle* as in the previous chapter, i.e., a cycle consisting of a primary and a backup path, each of which does not exceed l_{max}, from s to t. To compute the probability that there exists at least one protection cycle, we consider $E(|C|)$ where C denotes the set of all possible protection cycles including the pair.

In $G_e(n, D)$, let n_j be the number of vertices of degree d_j for $j = 1, 2, ..., j_{max}$. Conditioned on such $\{n_j\}_{j=1}^{j_{max}}$, the probability that a given set of k edges is contained in a random configuration is given in (3.9) such that

$$p_{k|\{n_j\}} \sim \left(\sum_{j=1}^{j_{max}} d_j n_j \right)^{-k}, \tag{3.33}$$

where k is fixed.

Now, we proceed to computing the number of protection cycles. Since this depends on the degrees of the randomly picked pair of nodes s and t, we first condition on D_s and D_t, the degrees of nodes s and t respectively, and later we will average the result over all values of D_s and D_t.

First, let us consider the number of protection cycles of length $k \leq (l_{max} + 1)$ on W, conditioned on $\{n_j\}_{j=1}^{j_{max}}$ and also on $D_s = d_u$ and $D_t = d_v$, $d_u \neq d_v$, which we denote by $a_{k|[\{n_j\}, d_u \neq d_v]}$. Since we need $(k - 2)$ intermediate nodes and allow any possible ordering of k nodes on the cycle, the number of possible protection cycles on

W is

$$a_{k|[\{n_j\},d_u\neq d_v]} = \frac{1}{2k} \sum_{k_1+\cdots+k_{jmax}=(k-2)} \frac{k!}{k_1!k_2!\cdots k_{jmax}!} \prod_{j\neq u,v} \frac{n_j!}{(n_j-k_j)!}(d_j(d_j-1))^{k_j}$$

$$\times \prod_{j=u,v} \frac{(n_j-1)!}{(n_j-1-k_j)!}(d_j(d_j-1))^{k_j+1}$$

$$\sim \frac{1}{2k}\, k(k-1) \sum_{k_1+\cdots+k_{jmax}=(k-2)} \frac{(k-2)!}{k_1!k_2!\cdots k_{jmax}!} \prod_{j=1}^{jmax} n_j^{k_j}(d_j(d_j-1))^{k_j}$$

$$\times d_u(d_u-1)d_v(d_v-1)$$

$$= \frac{k-1}{2}\, d_u(d_u-1)d_v(d_v-1)\left(\sum_{j=1}^{jmax} d_j(d_j-1)n_j\right)^{k-2}. \tag{3.34}$$

Note that the probability that randomly picked nodes s and t have degree d_u and d_v, $d_u \neq d_v$, is

$$\Pr(D_s = d_u,\ D_t = d_v,\ d_u \neq d_v) = \frac{n_u n_v}{\binom{n}{2}} \sim \frac{2 n_u n_v}{n^2}. \tag{3.35}$$

Now, if $d_u = d_v$,

$$a_{k|[\{n_j\},d_u=d_v]} = \frac{1}{2k} \sum_{k_1+\cdots+k_{jmax}=(k-2)} \frac{k!}{k_1!k_2!\cdots k_{jmax}!} \prod_{j\neq u} \frac{n_j!}{(n_j-k_j)!}(d_j(d_j-1))^{k_j}$$

$$\times \frac{(n_u-2)!}{(n_u-2-k_j)!}(d_u(d_u-1))^{k_j+2}$$

$$\sim \frac{1}{2k}k(k-1) \sum_{k_1+\cdots+k_{jmax}=(k-2)} \frac{(k-2)!}{k_1!k_2!\cdots k_{jmax}!} \prod_{j=1}^{jmax} n_j^{k_j}(d_j(d_j-1))^{k_j}$$

$$\times (d_u(d_u-1))^2$$

$$= \frac{k-1}{2}\, (d_u(d_u-1))^2\left(\sum_{j=1}^{jmax} d_j(d_j-1)n_j\right)^{k-2}, \tag{3.36}$$

and

$$\Pr(D_s = D_t = d_u) = \frac{\binom{n_u}{2}}{\binom{n}{2}} \sim \frac{n_u^2}{n^2}. \tag{3.37}$$

If we average the number of protection cycles of length k over all choices of d_u and d_v, from (3.34), (3.35), (3.36) and (3.37),

$$a_{k|\{n_j\}} \sim \sum_{u=v} \frac{n_u^2}{n^2} a_{k|[\{n_j\},d_u=d_v]} + \sum_{u\neq v} \frac{2n_u n_v}{n^2} a_{k|[\{n_j\},d_u\neq d_v]}$$

$$= \frac{k-1}{2} \frac{\sum_{u=1}^{j_{max}} (d_u(d_u-1)n_u)^2 + 2\sum_{u\neq v} d_u(d_u-1)n_u \cdot d_v(d_v-1)n_v}{n^2}$$

$$\times \left(\sum_{j=1}^{j_{max}} d_j(d_j-1)n_j \right)^{k-2}$$

$$= \frac{k-1}{2n^2} \left(\sum_{j=1}^{j_{max}} d_j(d_j-1)n_j \right)^k. \qquad (3.38)$$

On the other hand, if $(l_{max}+2) \leq k \leq 2l_{max}$, there exist some orderings on the cycle where s and t are located farther than l_{max} from each other, which we don't count. Similarly as in (2.17) and (3.38),

$$a_{k|\{n_j\}} \sim \frac{2l_{max} - k + 1}{2n^2} \left(\sum_{j=1}^{j_{max}} d_j(d_j-1)n_j \right)^k. \qquad (3.39)$$

Hence, from (3.33), (3.38) and (3.39), we obtain

$$\mathrm{E}(|C| \,|\{n_j\}) = \sum_{k=3}^{2l_{max}} a_{k|\{n_j\}} p_{k|\{n_j\}}$$

$$= \sum_{k=3}^{2l_{max}} \frac{1}{2n^2} \left(\frac{\left(\sum_{j=1}^{j_{max}} d_j(d_j-1)n_j\right)}{\sum_{j=1}^{j_{max}} d_j n_j} \right)^k \min[k-1, 2l_{max}-k+1]. \quad (3.40)$$

By averaging this over $\{n_j\}_{j=1}^{j_{max}}$, we now have the unconditional mean of $|C|$ such that

$$\mathrm{E}(|C|) = \sum_{k=3}^{2l_{max}} \frac{1}{2n^2} \left(\frac{\mathrm{E}(D^2) - \mathrm{E}(D)}{\mathrm{E}(D)} \right)^k \min[k-1, 2l_{max}-k+1]. \qquad (3.41)$$

If we consider the probability that there exists at least one protection cycle along the pair of nodes, it is bounded from above by $\mathrm{E}(|C|)$, which is a union bound including all possible protection cycles, and from below by the probability that there exists a

cycle of length 3 on W. Therefore,

$$\frac{1}{n^3}\frac{1}{\mathrm{E}(D)} \leq \Pr(\exists \text{protection cycle})$$

$$\leq \sum_{k=3}^{2l_{max}} \frac{1}{2n^2} \left(\frac{\mathrm{E}(D^2) - \mathrm{E}(D)}{\mathrm{E}(D)}\right)^k \min[k-1, 2l_{max} - k + 1].$$

3.4.2 Link Protection

Now assume there is a link between s and t. To ensure that traffic between the pair is recoverable by the link protection scheme, there must exist a cycle not exceeding $(l_{max} + 1)$ around the pair. In this section, we let a protection cycle refer to a cycle of this property and again let C be the set of all possible protection cycles.

To get the expectation of the cardinality of C, again, we need to sum $a_k p_k$ over all possible k, the length of cycle. Note in this case that s and t should be adjacent to each other and thus the longest cycle possible is of length $(l_{max} + 1)$. Similarly as in the case of path protection,

$$a_{k|\{n_j\}} \sim \frac{1}{2n^2} \left(\sum_{j=1}^{j_{max}} d_j(d_j - 1)n_j\right)^k.$$

and $p_{k|\{n_j\}}$ is given the same as before. Hence,

$$\mathrm{E}(|C| \,|\{n_j\}) = \sum_{k=3}^{l_{max}+1} a_{k|\{n_j\}} p_{k|\{n_j\}}$$

$$\simeq \sum_{k=3}^{l_{max}+1} \frac{1}{2n^2} \left(\frac{\left(\sum_{j=1}^{j_{max}} d_j(d_j - 1)n_j\right)}{\sum_{j=1}^{j_{max}} d_j n_j}\right)^k,$$

and by averaging this over $\{n_j\}_{j=1}^{j_{max}}$, we have the unconditional mean of $|C|$ such that

$$\mathrm{E}(|C|) = \sum_{k=3}^{l_{max}+1} \frac{1}{2n^2} \left(\frac{\mathrm{E}(D^2) - \mathrm{E}(D)}{\mathrm{E}(D)}\right)^k. \tag{3.42}$$

Note from the results above that, in random graphs of a general degree distribution as well as in random regular graphs (see (2.18) and (2.21)), the probability that we

66

find a backup path of finite length decays in the order of $\frac{1}{n^2}$ for both path and link protection schemes. In other words, it is highly unlikely to find a backup path within a finite range as the size of the network grows very large.

Chapter 4

Conclusions

4.1 Conclusions

We have investigated the issue of protection in large-scale random networks by deriving several bounds on the parameters related to protection in terms of the size of the network. Let us summarize the main results of this study.

We first considered random regular networks described by the configuration model. In such networks, for a randomly picked pair of nodes, we

- obtained a lower bound on the expected length of the shortest path between the pair and discussed the tightness of this bound

- presented a lower bound on the expected length of the shortest cycle including the pair, which represents the sum of the lengths of primary and backup paths

- calculated both upper and lower bounds on the probability that we can establish protected connections of finite lengths using path or link protection.

Also we considered a generalized graph model that yields random networks with arbitrary distributions of degrees. In such networks, we

- discussed asymptotic properties due to which the generalized model can be suited to describing the random networks for our analysis

- presented the distribution of the number of fixed-length cycles and derived an upper bound on the mean number of cycles of non-finite length

- showed that all the previous bounds for regular networks also hold for networks with general degree distributions with only minor changes while displaying the same asymptotic behavior.

The main contributions of this thesis are the following. First, we took an analytical approach toward the study of network protection by bringing the concept of randomness into network topologies. Our approach is crucial to understanding the relation between reliability metrics, such as connectedness and length of backup paths, and basic network parameters, such as degree distribution and network size, for large networks that display random growth. Our method can provide concise rules to relate those network parameters to robustness.

In addition, we established analytical results for the length of backup paths for path and link-based protection schemes rather than for the efficiency of backup capacity, upon which most studies concentrate. Our results show that the average lengths of primary and backup paths scale at least logarithmically with the size of the network, and that the protected connections of given finite length tend to be very rare. These results, though not complete yet because the upper bounds are currently missing, allow us to understand the applicability of standard protection preplanned approaches. That is, preplanned recovery methods may not be an attractive option for very large networks since they would require extensive maintenance of alternate routes, and the routes' lengths would not allow for transparent and rapid recovery. Also, this may suggest that a distributed and real-time approach is preferable.

Finally, we developed a unified framework for studying the issue of robustness in very general random networks with arbitrary degree distributions. Notably, we showed that the expression of the mean number of fixed-length cycles is an upper bound on the mean number of cycles of non-finite length. We can use our approach to consider general models for large random networks.

4.2 Directions for Further Research

There are several topics for further research. First, we need to continue deriving some probabilistic results for the asymptotic behavior of path lengths. In particular, our results indicate that, since the average distance between a pair of nodes scales logarithmically, the expected length of the shortest cycle including the pair may also do so. To complete this argument, in addition to the lower bound on cycle length that we have derived, we need an upper bound.

Also, as briefly mentioned in the previous section, we can use our analytical results to consider more specific models for large graph topology, such as power law distribution of degree, and examine the impact of degree distribution on reliability and path lengths in a more practical sense. In fact, we can model large networks with almost any kind of distribution of degrees and provide more concrete results.

Furthermore, we can extend our network model to allow the time variability of networks that are evolving, which may provide an analytical tool for developing or evaluating specific algorithms or protocols for protection. We can also modify the model to allow some dependency on proximity or localization, for instance to satisfy Rent's rule [8], or to explain transitivity or clustering in a network [25].

71

Bibliography

[1] W. Aiello, F. Chung, and L. Lu, "A random graph model for massive graphs," *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pp. 171-180, 2000.

[2] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378-382, 2000.

[3] A.-L. Barabási, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509-512, 1999.

[4] E. A. Bender, and E. R. Canfield, "The asymptotic number of labeled graphs with given degree sequences," *Journal of Combinatorial Theory*, Series A, vol. 24, pp. 296-307, 1978.

[5] B. Bollobás, *Random Graphs*, 2nd edition, Academic Press, 2001.

[6] G. N. Brown, W. D. Grover, J. B. Slevinsky, and M. H. MacGregor, "Mesh/Arc networking: An architecture for efficient survivable self-healing networks," *Proceedings of the IEEE International Conference on Communications*, vol. 1, pp. 471-477, 1994.

[7] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, "The origin of power laws in Internet topologies revisited," *Proceedings of IEEE INFOCOM*, pp. 608-617, 2002.

[8] P. Christie, and D. Stroobandt, "The interpretation and application of Rent's rule," *IEEE Transactions on Very Large Scale Integraion (VLSI) Systems*, vol. 8, pp. 639-648, 2000.

[9] T. M. Cover, and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[10] J. Doucette, W. D. Grover, and T. Bach, "Bi-criteria studies of mesh network restoration path-length versus capacity tradeoffs," *Proceedings of Optical Fiber Communication Conference and Exhibit*, vol. 2, pp. TuG2 -T1-3, 2001.

[11] A. W. Drake, *Fundamentals of Applied Probability Theory*, McGraw-Hill, 1967.

[12] G. Ellinas, S. Rong, A. Hailemariam, and T. E. Stern, "Protection cycle covers in optical networks with arbitrary mesh topologies," *Optical Fiber Communication Conference*, vol. 3, pp. 213-215, 2000.

[13] P. Erdős, "Some remarks on the theory of graphs," *Bulletin of the American Mathematical Society*, vol. 53, pp. 292-294, 1947.

[14] M. Faloutos, P. Faloutos, and C. Faloutos, "On power-law relationships of the Internet topology," *Proceedings of ACM SIGCOMM*, pp. 251-262, 1999.

[15] H. Garmo, "The asymptotic distribution of long cycles in random regular graphs," *Random Structures and Algorithms*, vol. 15, pp. 43-92, 1999.

[16] W. D. Grover, "Case studies of survivable ring, mesh and mesh-arc hybrid networks," *Proceedings of IEEE GLOBECOM*, pp. 633-638, 1992.

[17] W. D. Grover, and D. Stamatelakis, "Cycle-oriented distributed proconfiguration: Ring-like speed with mesh-like capacity for self-planning network reconfiguration," *Proceedings of the IEEE International Conference on Communications*, vol. 1, pp. 537-543, 1998.

[18] F. Jaeger, "A survey of the double cycle cover conjecture," *Cycles in Graphs*, Annals of Discrete Mathematics, vol. 115, pp. 1-12, North-Holland, 1985.

[19] S. Janson, T. Łuszak, and A. Ruciński, *Random Graphs*, John Wiley & Sons, 2000.

[20] S. S. Lumetta, and M. Médard, "Towards a deeper understanding of link restoration algorithms for mesh networks," *Proceedings of IEEE INFOCOM*, vol. 1, pp. 367-375, 2001.

[21] B. D. McKay, "The expected eigenvalue distribution of a large regular graph", *Linear Algebra and Its Applications*, vol. 40, pp. 203-216, 1981.

[22] M. Médard, R. A. Barry, S. G. Finn, W. He, and S. Lumetta, "Generalized loop-back recovery in optical mesh networks," *IEEE/ACM Transactions on Networking*, vol. 10, pp. 153-164, 2002.

[23] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical Review E*, vol. 64, 026118, 2001.

[24] M. E. J. Newman, "Random graphs as models of networks," *Handbook of Graphs and Networks*, S. Bornholdt and H. G. Schuster (eds.), Wiley-VCH, Berlin, 2003.

[25] M. E. J. Newman, "The structure and function of networks," *SIAM Review*, vol. 45, pp. 167-256, 2003.

[26] S. Ramamurthy, and B. Mukherjee, "Survivable WDM mesh networks, part I - Protection," *Proceedings of IEEE INFOCOM*, pp. 744-751, 1999.

[27] S. Ramamurthy, and B. Mukherjee, "Survivable WDM mesh networks, part II - Restoration," *Proceedings of the IEEE International Conference on Communications*, vol. 3, pp. 2023-2030, 1999.

[28] R. Ramaswami, and K. N. Sivarajan, *Optical Networks: A Practical Perspective*, Morgan Kaufmann Publishers, 1998.

[29] R. W. Robinson, and N. C. Wormald, "Almost all regular graphs are hamiltonian," *Random Structures and Algorithms*, vol. 5, pp. 363-374, 1994.

[30] A. Ruciński, and N. C. Wormald, "Random graph process with degree restrictions," *Combinatorics, Probability and Computing*, vol. 1, pp. 169-180, 1992.

[31] A. Ruciński, and N. C. Wormald, "Random graph processes with maximum degree 2," *Annals of Applied Probability*, vol. 7, pp. 183-199, 1997.

[32] A. Ruciński, and N. C. Wormald, "Connectedness of graphs generated by a random d-proces," *Journal of the Australian Mathematical Society*, vol. 72, pp. 67-85, 2002.

[33] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, pp. 268-276, 2001.

[34] C. Thomassen, "On the complexity of finding a minimum cycle cover of a graph," *SIAM Journal on Computing*, vol. 26, pp. 675-677, 1997.

[35] N. C. Wormald, "The asymptotic connectivity of labelled regular graphs," *Journal of Combinatorial Theory*, Series B, vol. 31, pp. 156-167, 1981.

[36] N. C. Wormald, "Models of random regular graphs," *Surveys in Combinatorics, 1999, J.D. Lamb and D.A. Preece, eds. London Mathematical Society Lecture Note Series*, vol. 276, pp. 239–298, Cambridge University Press, 1999.

[37] T. H. Wu, *Fiber Network Service Survivability*, Artech House, 1992.