Robustness in Large-Scale Random Networks

Minkyu Kim and Muriel Médard
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
Email: {minkyu, medard}@mit.edu

***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 reliability. In particular, we take the case of random regular networks for simplicity 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, and show that the second moment of degree over the first moment is an important shorthand for behavior of a network. Notably, 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. We present as an example the case of networks with a power-law degree distribution.**

***Index Terms*— Graph theory, combinatorics, network robustness, random graph**

I. Introduction

Providing resilient service against failures is a crucial issue for high-speed networks since a single failure may cause a severe loss of data. Today's high-speed networks are becoming increasingly complex and also dynamic in response to growing and shifting communication demands [11]. In such networks, the issue of reliability also becomes increasingly complex.

Restoration has been extensively researched for general mesh topologies, but very few analytical results are available. The typical approach is to give linear programming formulations or heuristic algorithms and to rely on simulations based on some standard networks for evaluating their performance (e.g., [15], [8]). While this type of method can provide numerical results for each network with a specific topology, it is often difficult to extrapolate these results to give an analytical view of how parameters scale as networks grow. Also, it may fail to provide concise rules to relate important network parameters, such as size and degree, to robustness.

Networks evolve over time, that is, nodes and links are added and deleted, or different networks can be interconnected. Furthermore, as networks become very large and change rapidly, they may grow in an increasingly uncontrolled fashion since they tend to no longer remain under the control of a single entity.

Our goal is to investigate the relation between reliability metrics and basic network parameters for very large networks that display randomness in topology. We use a random graph method to capture this phenomenon, where we compute reliability metrics in a probabilistic sense for a randomly chosen network from the set of networks with given size and degree constraints. In particular, for a randomly picked pair of nodes in a network, we consider

- length of the shortest path between the pair
- length of the shortest cycle including the pair, which represents the sum of the lengths of primary and backup paths
- probability that we can establish protected connections within a finite length bound using path or link protection

in terms of the size and the degree distribution of the network.

To this end, we first employ a random regular graph model, where the degree of each node is the same, for simplicity of exposition. Then we extend the graph model so as to deal with networks of arbitrary degree distributions and obtain generalized results applicable to a much wider family of networks.

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 (e.g., [15]). The speed of restoration is also considered [16], sometimes jointly with capacity [8]. Some other considerations are transparency, flexibility, and vulnerability [11].

In this paper, we are concerned with the length of paths in terms of the number of hops. While this parameter is less widely considered than 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 reasons. Also, path length indirectly affects efficiency and speed, i.e., a longer protection path requires a larger amount of resources, time and management complexity.

If we use path protection to protect the network against link (node) failure, then we have to establish a backup path which is link (node)-disjoint from source to destination. By Menger's theorem, the existence of such path between any two nodes is guaranteed in any edge (vertex)-redundant graph [17]. We see that the primary and backup paths form a cycle along the source and the destination. Also in link protection, the backup path around the failed link, together with the failed link itself, form a cycle. In light of these observations, the distribution and length of cycles in the graph are of natural interest.

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. Also, 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 of networks [6].

This paper is organized as follows: Section II considers the case of random regular networks, Section III generalizes the results to the case of networks of arbitrary degree distributions, Section IV presents as an example the case of networks with a power-law degree distribution, and Section V concludes with a summary of the results and a discussion of further work.

II. Regular Networks

In this section, we consider random regular networks, where each node has the same degree. This model, though seemingly too restrictive, can provide simplicity to our exposition, but also enough insight for results applicable to general networks. In the next section, we will find that many of the results can carry over, with minor modifications, to networks with arbitrary degree distributions.

A. Random Regular Graph Model

We represent each network by a graph, where each vertex corresponds to a node in the network and each edge to a link. By n we denote the number of vertices and by d the common degree of every vertex, where $3 \le d \le n-1$, and we assume that 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.

We here present the *configuration model*, which is a standard method for constructing random regular graphs uniformly [9], [20]. 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, $W = [n] \times [d]$. A *configuration* is a partition of W into $(dn/2)$ pairs. 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 (see Fig. 1). 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.

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

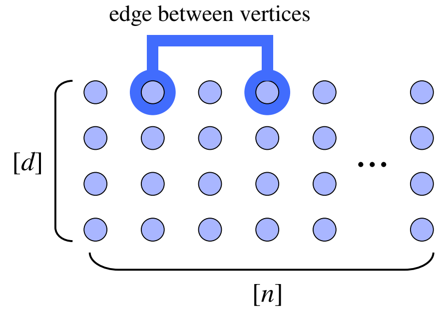


Fig. 1. Two-Dimensional Set W for Configuration Model

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 in the corresponding network we have no viable option to restore the connections but to recover the failed link or node itself. Therefore, there is no need to consider protection for such pairs. However, in random regular graphs constructed by the configuration model, it is known that such a phenomenon does not happen as n grows large.

For an event $\mathcal{E}_n$, we say that $\mathcal{E}_n$ holds *asymptotically almost surely* (*a.a.s.*) if $\Pr(\mathcal{E}_n) \to 1$ as n tends to infinity. Then, we have the following result regarding connectivity [19]:

Theorem 2.1: If $d \ge 3$ 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 after removing $(d-1)$ vertices from $G(n,d)$ for $d \ge 3$.

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 $k \ge 3$, Z_k's are asymptotically distributed as independent Poisson random variables [3]. More precisely,

Theorem 2.2: For each fixed j, a sequence of random variables $(Z_3, Z_4, ..., Z_j)$ converges *a.a.s.* 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 $\mathrm{E}(Z_{k\infty}) = \frac{(d-1)^k}{2k}$.

Note, however, that the previous theorem applies only for cycles of fixed length, that is, where the length of cycle does not grow with n. The case of long cycles of which length k is defined as a function of n, i.e., $k = k(n)$, is considered more recently by Garmo [7]. By counting the number of cycles on the two-dimensional set $W = [n] \times [d]$ and using Stirling's formula, Garmo calculates $\mathrm{E}(Z_k)$, $k = 3, ..., n$, as follows:

Lemma 2.3: Let k be an integer, $3 \le k \le n$, and $\lambda = k/n$. Then,

$$\mathrm{E}(Z_k) = \frac{(d-1)^k}{2k} \frac{1}{\exp\{\frac{1}{2}(\frac{d-2}{d}k - 1)\lambda + O(k\lambda^2)\} + O(\frac{1}{n})}.$$

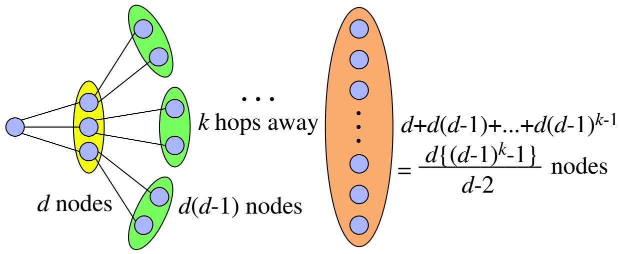


Fig. 2. Maximum Number of Nodes Within k Hops of s

In the above lemma, if k is fixed, then $\lambda \to 0$ as n tends to infinity, and thus $\mathrm{E}(Z_k) \to \frac{(d-1)^k}{2k}$ as shown in Theorem 2.2. On the other hand, if k grows with n, the exponent terms in the denominator, $\frac{1}{2}(\frac{d-2}{d}k - 1)\lambda + O(k\lambda^2)$, are no longer negligible, which leads $\mathrm{E}(Z_k)$ to become smaller than $\frac{(d-1)^k}{2k}$. Therefore, we see that $\frac{(d-1)^k}{2k}$ is an upper bound on $\mathrm{E}(Z_k)$ valid for all k, $3 \le k \le n$.

We will use these asymptotic results in the following discussion to quantify the reliability issues of networks represented by the configuration model.

B. Shortest Path

Throughout the remainder of this section, we assume that our network is a large random network which is d-regular graph generated by the 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.

Before proceeding, we present an important property of the model that we will use 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 the random selection of a graph and the other is the random pair selection. However, note that, by the symmetric structure of the configuration model, the value of X has no dependence on a specific pair. Hence, calculating 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.

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 $d + d(d-1) + d(d-1)^2 + \cdots + d(d-1)^{k-1} = d\{(d-1)^k - 1\}/(d-2)$ (see Fig. 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,

$$\Pr(L \le k) \le \min[1, \left(\frac{d\{(d-1)^k - 1\}}{d-2} \cdot \frac{1}{n-1}\right)].$$

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

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

where we assume that n is large.

For comparison, let us consider a related result by Newman *et al.* [14]. They give an asymptotic heuristic estimate of the typical length $\hat{L}$ of the shortest path between two randomly chosen nodes as follows:

$$\begin{aligned} \hat{L} &= \frac{\log[(n-1)(d^2 - 2d) + d^2] - \log d^2}{\log(d-1)} \\ &\sim \log_{d-1} n. \end{aligned}$$

They also note that this approximation may not be correct if all the vertices are not reachable from a randomly chosen vertex. However, if $d \ge 3$, 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.

Comparing this to the lower bound (1), we find that our lower bound matches the existing estimate for large n, and this may be viewed as an indication of its tightness.

C. 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.

Now we define an event Y_k that the pair is on a k-cycle (cycle of length k), i.e., there exists a k-cycle through the two nodes. Then, by the definition of X, $X \le k$ implies the pair is on a certain cycle no longer than k and we obtain the following inequality:

$$\Pr(Y_k) \le \Pr(X \le k) \le \sum_{i=3}^{k} \Pr(Y_i), \tag{2}$$

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

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{k=3}^{n} k \Pr(X=k) \\
&\geq \sum_{k=3}^{m-1} k\{\Pr(X \leq k) - \Pr(X \leq k-1)\} \\
&\qquad + m(1 - \Pr(X \leq m-1)) \qquad (3) \\
&\geq \sum_{k=3}^{m-1} k\{\Pr(Y_k) - \sum_{j=3}^{k-1} \Pr(Y_j)\} \\
&\qquad + m(1 - \sum_{j=3}^{m-1} \Pr(Y_j)) \qquad (4) \\
&= m - \sum_{k=3}^{m-1} \Pr(Y_k)\{\sum_{j=k+1}^{m} j - k\}, \qquad (5)
\end{aligned}$$

where m is an integer, $4 \leq m \leq n$. Note in (3) 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 (4) follows from (2). Since in (5) each $\Pr(Y_k)$ is multiplied by a negative number, 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 to 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:

$$\begin{aligned}
\Pr(Y_k) &= \sum_j \mathrm{E}(I_k | Z_k = j) \Pr(Z_k = j) \\
&= \sum_j \Pr(Y_k | Z_k = j) \Pr(Z_k = j), \qquad (6)
\end{aligned}$$

where the expectation and probability conditioned on Z_k are over the probability space of 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,

$$\Pr(Y_k | Z_k = j) = \frac{(\text{number of pair selections on } k\text{-cycle})}{\binom{n}{2}}. \qquad (7)$$

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 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 (6) and (7) that

$$\begin{aligned}
\Pr(Y_k) &\leq \sum_j \frac{j\binom{k}{2}}{\binom{n}{2}} \Pr(Z_k = j) \\
&= \frac{k(k-1)}{n(n-1)} \mathrm{E}(Z_k). \qquad (8)
\end{aligned}$$

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

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

Therefore,

$$\Pr(Y_k) \leq \frac{(k-1)}{2n(n-1)} (d-1)^k. \qquad (9)$$

Combining (5) and (9), we obtain

$$\mathrm{E}(X) \geq m - \sum_{k=3}^{m-1} \frac{(k-1)(d-1)^k}{2n(n-1)} \{\sum_{j=k+1}^{m} j - k\}, \qquad (10)$$

which is valid for any m, $4 \leq m \leq n$. We can calculate this lower bound numerically for various m. In Fig. 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 further grows.

We can collect these lower bounds for each n, which Fig. 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 in turn congruent to the lower bound (Eq. (1)) on the path length in the previous section.

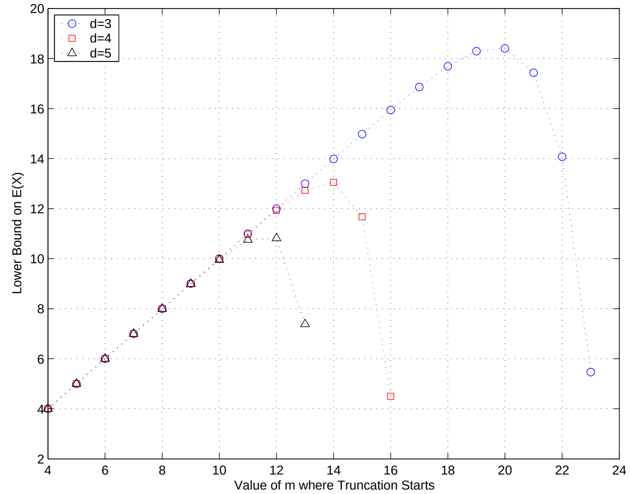


Fig. 3. Lower Bound on E(X) with respect to m for $n = 10,000$

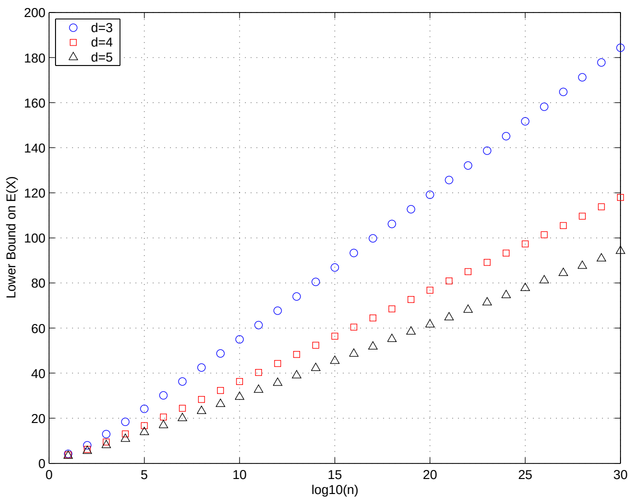


Fig. 4. Lower Bound with respect to $\log n$

This can be explained analytically as well. For fixed d and $m = m(n)$ that grows with n, let B be the terms on the right-hand side in (10). Then, by using manipulation of series,

$$\begin{aligned} B &= m - \sum_{k=3}^{m-1} \frac{(k-1)(d-1)^k}{2n(n-1)} \{ \sum_{j=k+1}^{m} j - k \} \\ &= m - \frac{1}{n^2} \cdot \Theta\left(\sum_{k=3}^{m-1} m^2 k (d-1)^k - k^3 (d-1)^k \right) \\ &= m - \frac{1}{n^2} \cdot \Theta(m^3 (d-1)^m). \end{aligned} \tag{11}$$

Let us suppose that $m = c \log n$ for a constant $c > 0$, where we can infer, by examining the value of m that gives the tightest lower bound for each n in Fig. 3, that the maximum may occur when m is approximately order of $\log n$. 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 $\Theta(m^3(d-1)^m)/n^2 \to 0$. Otherwise, if $c \geq \frac{2}{\log(d-1)}$, then B tends to below zero since the term $\Theta(m^3(d-1)^m)/n^2$ with a minus sign 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. Therefore, we conclude that the best case is when $B = \Theta(\log n)$, which is the tightest lower bound on $\mathrm{E}(X)$.

We also notice that, since the tightest lower bound occurs when $c \approx \frac{2}{\log(d-1)}$, the resulting bound is approximately $2 \log_{d-1} n$. Therefore, the lower bound on the shortest cycle turns out to be roughly twice the lower bound on the shortest path we obtained for regular graphs.

D. Probability of Short 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 I. Let a finite number l_{max} denote the maximum length of the paths allowed, and we want to compute the probability that we can protect the traffic using only such paths.

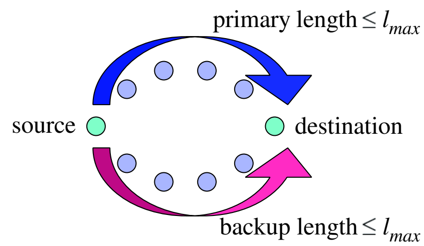


Fig. 5. Protection Cycle for Path Protection

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 Fig. 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 consider $\mathrm{E}(|C|)$, 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

$$\mathrm{E}(|C|) = \mathrm{E}[\sum_{c \in C} I_c] = \sum_{c \in C} \Pr[\exists c]. \tag{12}$$

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 by the same expression, which we denote by $p_k \sim (dn)^{-k}$ [9], i.e., it depends only on the number of edges. Therefore, we can calculate $\mathrm{E}(|C|)$ by calculating p_k and the number of protection cycles of length k, and summing their product over all possible length k's.

Now consider the number of protection cycles of length $k \leq (l_{max} + 1)$ 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

$$\begin{aligned} a_k &= \binom{n-2}{k-2} \frac{(k-1)!}{2} (d(d-1))^k \\ &\sim n^{k-2} \frac{(k-1)}{2} (d(d-1))^k, \end{aligned} \tag{13}$$

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 because it needs a path longer than l_{max}. Hence,

$$\begin{aligned} a_k &= \binom{n-2}{k-2} \frac{(k-2)!(2l_{max} - k + 1)}{2} (d(d-1))^k \\ &\sim n^{k-2} \frac{(2l_{max} - k + 1)}{2} (d(d-1))^k, \end{aligned} \tag{14}$$

where $k = (l_{max} + 2), ..., 2l_{max}$. Therefore, using (13) and

(14), we obtain

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

If we consider the probability that there exists at least a 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,

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

In the case of link protection, if we assume that there is a link between s and t, in order 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 exactly the same manner, we can calculate the expected number of such cycles around the pair. Hence, in the link protection case, we can bound the probability that there exists at least one protection cycle of length within a finite bound as follows:

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

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 the random networks described by the configuration model, it is highly unlikely to find a backup path within a finite range as the size of network grows very large.

III. General Networks

In this section, we present an extended version of the configuration model, by which we can overcome the limitation that the degree must be the same over all nodes. Then we show that most of our results for regular graphs carry over to more general networks based on the extended model.

A. Extended Graph Model

Molloy and Reed [12] and Newman *et al.* [14] present a random graph model with a given degree sequence, but they do not consider explicitly the randomization of degrees with a given degree distribution. Aiello *et al.* [1] use the same model as that we discuss here, however their analyses are limited to a specific degree distribution: the power-law distribution, which we will take as an example in the next section. We address our graph model in detail to demonstrate the extendability of our previous results for regular graphs.

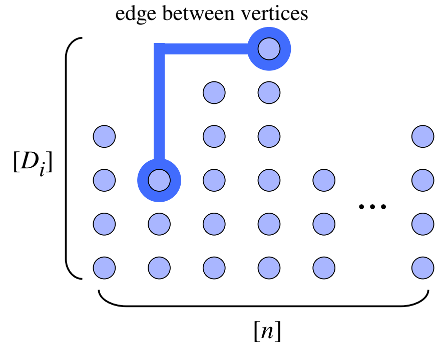


Fig. 6. 2-Dimensional Set W for Extended Model

In the configuration model, we consider a two-dimensional set $W = [d] \times [n]$ and partition the set into $(dn/2)$ pairs, then project onto the horizontal axis. A natural extension is to allow the degrees to vary over a finite range and keep the remaining procedure the same as before. Below we describe in detail the whole procedure of the extended model.

Suppose first that we are given a degree distribution for the graph, i.e., ith vertex has degree D_i, $i = 1, 2, ..., n$, each of which is defined to be an *identically independently distributed* (i.i.d.) random variable D such that

$$\Pr(D = d_j) = p_j \qquad j = 1, 2, ..., j_{max}, \tag{15}$$

where we assume $3 \leq d_1 < d_2 < \cdots < d_{j_{max}} < \infty$. 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 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. (Note that this regeneration causes D_n to be no longer i.i.d. with respect to other D_i's, but the effect of this regeneration is negligible.)
- Consider a two-dimensional set $W = [D_i] \times [n]$ consisting of $m = \sum_{i=1}^{n} D_i$ vertices (see Fig. 6).
- 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 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.

Note that in this model, by setting the minimum degree to be at least three, we can restrict ourselves to considering vertices of degree no less than three. Let us justify this exclusion in the context of protection in communication networks. For a node of degree one, there is only one link connecting the node to the network. Hence, if the link fails, there is no way but to simply fix the failed link to recover the connection. Also, it is easy to see that we cannot establish two link-disjoint paths starting or ending at a node of degree one. Hence, we do not need to consider nodes of degree one explicitly in both link and path protection.

For nodes of degree two, any such node should fall in the middle of links between a different pair of nodes. Hence, even if we ignore a node of degree two and merge its two links into one, there will be no changes in the topological structure of the network. Hence, in considering the asymptotic behavior of the length of paths, we can ignore nodes of degree two and later, if needed, we can add such nodes according to an appropriate distribution, which can be handled in a separate problem.

Regarding connectivity, [19] shows that a graph with any given collection of degrees lying between r and R, $3 \leq r \leq R \leq \infty$, is *a.a.s.* r-connected. Hence, if $d_{min} \geq 3$ is the minimum degree that each node can take, graphs constructed by the extended configuration model are *a.a.s.* d_{min}-connected.

B. Distribution of the Number of Cycles

We recall that the number of k-cycles in random regular graphs is asymptotically Poisson distributed. Interestingly, this property carries over to the case of general networks based on the extended model. Let the distribution of degree D be given as in (15) and denote the resulting graph by $G_e(n, D)$. Then we obtain the following theorem:

Theorem 3.1: 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 $(Z_3, Z_4, ..., Z_j)$ converges *a.a.s.* 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 $\mathrm{E}(Z_{k\infty}) = \frac{1}{2k}(\frac{\mathrm{E}(D^2)}{\mathrm{E}(D)} - 1)^k$.

Proof outline: Details of the proof are omitted for lack of space but can be found in [10]. This is an extension of the proof of the distribution of short cycles in random regular graphs [9]. First, consider a random multigraph with the given degree distribution. By conditioning on the number of nodes with degree d_i and using the strong law of large numbers, we can calculate each factorial moment. Averaging the results based on the degree distribution, we can show that each factorial moment converges *a.a.s.* to that of the desired joint Poisson random variables. Since Z_i's are independent, the distributions remain unchanged after conditioning $Z_1 = Z_2 = 0$. Hence, the result for a simple graph follows. □

Note that the above theorem holds only for fixed-length cycles. For length k which grows with n, we show that the expression of $E(Z_k)$ for fixed k is an upper bound on $E(Z_k)$, i.e.,

Lemma 3.2:

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

for $3 \leq k \leq n$.

Proof outline: A full derivation is omitted due to space limitations but also can be found in [10]. Even for $k = k(n)$, by considering the corresponding two-dimensional set, we can calculate the mean number of k-cycles exactly in a closed-form, which is, however, complicated. Applying Stirling's formula, we obtain a simpler asymptotic expression, and we can show that the desired inequality finally reduces to the log-sum inequality. □

We see that the crucial characteristic here is the second moment of degree over the first moment, $\{\mathrm{E}(D^2)/\mathrm{E}(D)\}$, which plays the same role as the degree in regular graphs (see Theorem 2.2). As we will see later, this parameter also has a crucial impact on the length of path and cycle.

C. Shortest Path

Throughout the remainder of this section, we consider a large network represented by a random graph $G_e(n, D)$, which is generated by the extended configuration model with sufficiently large n and the given degree distribution D satisfying the properties discussed in Section III-A.

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. As in the previous case of random regular graphs, Newman *et al.*'s heuristic approach to a related model [14] also applies to networks of general degree distribution, and we will discuss how their asymptotic estimate of typical length $\hat{L}$ relates to our results.

For a graph constructed by our model with general degree distribution D, $\hat{L}$ is given in [14] by

$$\hat{L} = \frac{\log[\{(n-1)(\mathrm{E}(D^2) - 2\mathrm{E}(D)) + \mathrm{E}(D)^2\}/\mathrm{E}(D)^2]}{\log[(\mathrm{E}(D^2) - \mathrm{E}(D))/\mathrm{E}(D)]} \tag{16}$$

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

$$\hat{L} = \frac{\log(\frac{n}{\mathrm{E}(D)})}{\log(\frac{\mathrm{E}(D^2)}{\mathrm{E}(D)} - 1)} + 1. \tag{17}$$

Interestingly, the second moment of degree over the first moment is also crucial here. As we increase the variance of degrees while maintaining the same mean degree, we have shorter $\hat{L}$. On the other hand, as pointed out in [14], two random graphs with completely different distributions of degrees, but the same value of the second moment of degree over the first moment, will have asymptotically the same mean path length.

D. Shortest Cycle

As before, we now consider a random variable X, defined to be the length of the shortest cycle including a randomly picked pair of nodes. We will show that the procedures for deriving a lower bound on $\mathrm{E}(X)$ for regular graphs apply here with slight modifications, which is expected because we can see that most of the previous results 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 details if there is no major change compared with the case of regular graphs.

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, $4 \leq m \leq n$,

$$\mathrm{E}(X) \geq m - \sum_{k=3}^{m-1} \Pr(Y_k)\{\sum_{j=k+1}^{m} j - k\}. \quad (18)$$

Also, we have the same expression as in (8) for an upper bound on $\Pr(Y_k)$ such that

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

Applying Lemma 3.2, we have, for graphs with a general degree distribution D,

$$\Pr(Y_k) \leq \frac{(k-1)}{2n(n-1)} \left(\frac{\mathrm{E}(D^2)}{\mathrm{E}(D)} - 1 \right)^k. \quad (19)$$

Hence, from (18) and (19), we obtain the following lower bound on $\mathrm{E}(X)$:

$$\mathrm{E}(X) \geq m - \sum_{k=3}^{m-1} \frac{(k-1)(\frac{\mathrm{E}(D^2)}{\mathrm{E}(D)} - 1)^k}{2n(n-1)} \{\sum_{j=k+1}^{m} j - k\}. \quad (20)$$

By similar numerical calculation and convergence analysis as in Section II-C, we conclude that the above bound is $O(\log n)$.

It is very interesting to see that, since it is multiplied by a negative value, as $\{\mathrm{E}(D^2)/\mathrm{E}(D)\}$ increases we have a smaller bound on $\mathrm{E}(X)$. Note that this fact is consistent with the results in the previous two sections: a larger value of $\{\mathrm{E}(D^2)/\mathrm{E}(D)\}$ yields shorter $\hat{L}$ (Eq. (17)) and more cycles (Theorem 3.1). This observation meets our expectation that, as the number of cycles increases, the length of the shortest cycle including a pair of nodes as well as the distance between the pair decreases.

Hence, the second moment of degree over the first moment should be dealt with as a special parameter that determines the asymptotic behavior of both the average length of paths and cycles, and the average number of cycles in random graphs.

E. *Finite Length Cycle*

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.

For path protection, we define a *protection cycle* as before, 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 $\mathrm{E}(|C|)$ where C denotes the set of all possible protection cycles including the pair. Similarly as in regular graphs, we can calculate $\mathrm{E}(|C|)$ by counting the number of appropriate cycles in the two-dimensional set of the extended configuration model. A full derivation can be found in [10]. Owing to space limitations we present only the result here.

$$\mathrm{E}(|C|) \sim \sum_{k=3}^{2l_{max}} \frac{1}{2n^2} \left(\frac{\mathrm{E}(D^2)}{\mathrm{E}(D)} - 1 \right)^k \times \min[k-1, 2l_{max} - k + 1]. \quad (21)$$

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,

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

Now, for link protection, 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. As in path protection, we calculate the expected number of such cycles around the pair and bound the probability that there exists at least one protection cycle of length within a finite bound as follows:

$$\frac{1}{(n\mathrm{E}(D))^3} \leq \Pr(\exists \text{protection cycle}) \leq \sum_{k=3}^{l_{max}+1} \frac{1}{2n^2} \left(\frac{\mathrm{E}(D^2)}{\mathrm{E}(D)} - 1 \right)^k.$$

Note from the results above that, also in random graphs of a general degree distribution, the probability that we 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.

IV. Example: Power-Law Degree Distribution

In this section, we take as an example the case of networks with a power-law degree distribution, which is often used as a description of large complex networks. Our aim here is not to discuss the validity of the power-law model, but to illustrate the flexibility and usefulness of our method by applying it to an existing model that is commonly used.

A considerable number of studies have been carried out on the structures and properties of many kinds of real-world complex network, e.g., [2], [18]. 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 [13]. More specifically, it is often of the form $p_k \sim k^{-\alpha}$ for some positive constant α, which is called a power-law distribution.

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. Among those few studied, a frequently considered network is the Internet. Faloutsos *et al.* [6] discover that several parameters of the Internet graph, such as outdegrees of a node or eigenvalues, display power-law distributions. However, more recently Chen *et al.* [4] show that, by using a different method for obtaining the Internet graph, degree distributions of the Internet graphs are also heavy-tailed but deviate significantly from a strict power-law. These observations indicate that, in considering communication networks, we do not need to limit ourselves to networks with the degree distribution of a power-law.

Now, let us consider networks with the degree distribution defined as

$$\Pr(D = k) = \begin{cases} 0 & \text{for } 0 \leq k \leq 2 \\ k^{-\alpha}/c & \text{for } 3 \leq k \leq k_{max}, \end{cases}$$

for given constant α. Here k_{max} is an often-used approximate of the maximum degree [13], given by $k_{max} = n^{1/(\alpha-1)}$, and c is a normalizing constant, given by $c = \sum_{k=3}^{k_{max}} k^{-\alpha}$. Note that this distribution is different from a strict power-law distribution in that nodes of degree below three are ignored for the reasons discussed in Section III-A. Also, this analysis is approximate – the degree constraint of our model is not precisely satisfied since here the maximum degree is a function of n, although the probability of such nodes decays significantly. Within the best of our knowledge, the overall effect of the unbounded maximum degree in graphs with a general degree distribution is not known precisely.

For networks of this degree distribution, Fig. 7 plots with respect to $\log n$ the parameters that we have considered. We choose three different α's around the experimental values in [6]. In particular, Fig. 7(a) shows the second moment of degree over the first moment of degree, which as we have discussed above closely relates to the average number of cycles as well as the average path/cycle length. We find that the value of the second moment of degree over the first moment increases with n, but decreases with α. In Fig. 7(b), we plot the asymptotic heuristic estimate of typical distance, $\hat{L}$ (Eq. (16)), and the lower bound on the expected length of the shortest cycle including a randomly picked pair of nodes, $E(X)$ (Eq. (20)). Here we see that, as we discussed before, those two parameters decrease as $\{\mathrm{E}(D^2)/\mathrm{E}(D)\}$ increases for the same n.

One may notice that those parameters produce surprisingly small values for the size of the network. Recall that the second moment of degree over the first moment is a crucial parameter that determines the asymptotic behavior of a network. For cycle length, this parameter corresponds exactly to the degree in a regular network (see (20) and (10)). For instance, if $\alpha = 2.6$ and $n = 10^{10}$, the network shows the characteristics of a regular graph with degree 825. This large value results partly from the fact that we ignore the nodes of degree less than three, and if we calculate the parameter again including the nodes of degree one and two, we now have 345, still a large value. Hence, we conclude that this property of high-connectivity and short path/cycle lengths is an inherent characteristic of networks with a power-law degree distribution.

Note, however, that we do not intend to suggest that the power-law distribution is appropriate for large-scale networks in practice, nor that our approach is limited to power-law distributed networks. Rather, we have shown that our approach is applicable to estimating the length of backup paths in general networks of arbitrary degree distributions.

V. Conclusions

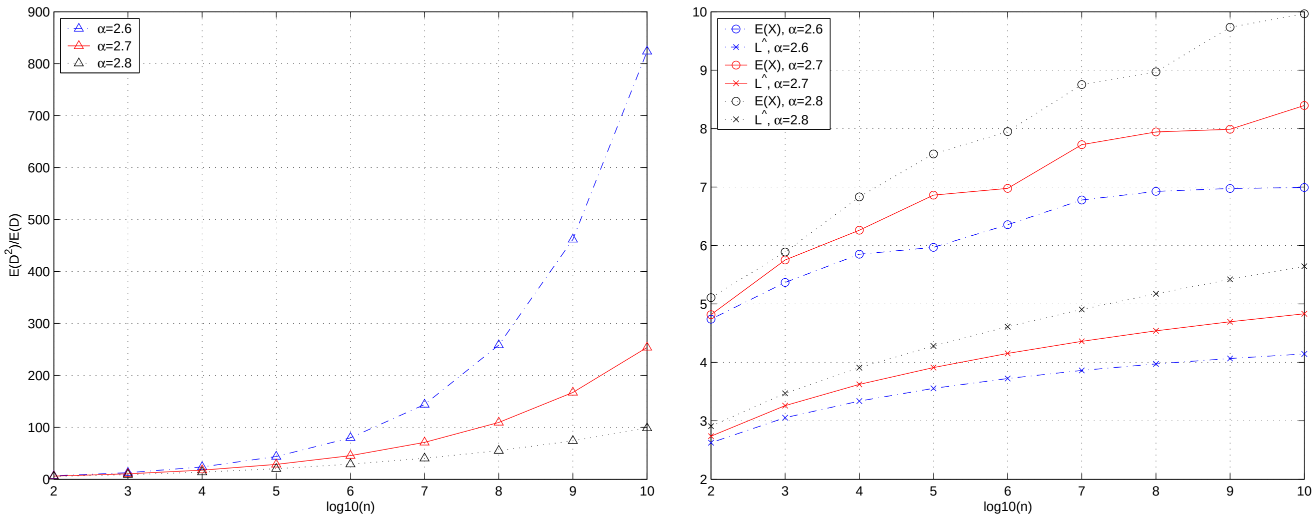
We have investigated the issue of protection in large-scale random networks by deriving several bounds on the parameters related to protection in terms of network parameters, such as size and degree. We first considered random regular networks described by the configuration model, and derived bounds on the mean path and cycle length and also the probability of finite-length protected connections. We then extended these results to general networks with arbitrary degree distributions. We 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. We found the second moment of degree over the first moment is a crucial parameter for the asymptotic behavior of a network.

The main contributions of this study 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 connectivity and length of backup paths, and basic network parameters, such as degree distribution and network size, for large networks. In addition, we established analytical results for the length of backup paths for path and link-based protection schemes. These results, though not complete yet, allow us to understand the applicability of standard protection preplanned approaches. Finally, we developed a unified framework for studying the issue of robustness in very general networks with arbitrary degree distributions.

There are several topics for further research. Our results indicate that both the shortest path and cycle between a pair of nodes may scale logarithmically with the size of the network. A formal analysis of the validity of this claim would be useful. Furthermore, we may extend our network model to allow time variability of networks that are evolving dynamically, which may provide an analytical tool for developing or evaluating specific algorithms or protocols for protection. The network model may be also modified to allow some dependency on proximity or localization, for instance to satisfy Rent's rule [5], or to explain transitivity or clustering in a network [13].

Acknowledgment

The authors would like to thank Douglas B. West, R. Srikant and Supratim Deb at the University of Illinois at Urbana-Champaign for helpful discussions.



(a) Second Moment of Degree over First Moment

(b) $\hat{L}$ and Lower Bound on $\mathrm{E}(X)$

Fig. 7. Networks of Power-Law Degree Distribution

References

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

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

[4] 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.

[5] 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.

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

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

[8] W. D. Grover and D. Stamatelakis, ”Cycle-oriented distributed preconfiguration: 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.

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

[10] M. Kim, *Robustness in large-scale random networks*, S.M. Thesis, Massachusetts Institute of Technology, 2003.

[11] 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.

[12] M. Molloy and B. Reed, ”A critical point for random graphs with a given degree sequence,” *Random Structures and Algorithms*, vol. 6, pp. 161-180, 1995.

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

[14] 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.

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

[16] 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.

[17] M. Stoer, *Design of Survivable Networks*, New York: Springer-Verlag, 1992.

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

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

[20] 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.