# Complex Networks: Structure, Percolation and Optimization

Tomer Kalisky
Department of Physics
Ph.D. Thesis

Submitted to the Senate of Bar-Ilan university
Ramat-Gan, Israel January 1, 2006

This work was carried out under the supervision of Professor Shlomo Havlin (Department of Physics), Bar-Ilan University.

# Acknowledgments:

I wish to thank my supervisor, Shlomo Havlin, for his guidance in science and in general. Throughout my research I have enjoyed his constant trust, help and support.

I wish to thank my collaborators Yosef Ashkenazy, Reuven Cohen, Sameet Sreenivasan, Lidia Braunstein, Sergey Buldyrev, Eugene Stanley, Ehud Perlsman, Yaakov Strelniker and Nilly Madar, from which I have learned a lot through our joint work.

I wish to thank and all my colleagues from Shlomo Havlin's group, Ronny Bartsch, Shay Moshel, Avi Gozolchiani, Shay Carmi, Valerie Livina, and Igor Zhitkov.

I wish to thank all members of the Bar-Ilan Physics department for the enjoyable years I have spent there.

Finally, I wish to thank my wife Beena for her constant support and clever advice, which have helped me surpass difficulties I have encountered in my work.

# Contents

# List of Figures

# List of Tables

# Abstract

In the last few years there has been much interest in the subject of complex networks. It was realized that many real-world systems have a web-like structure of many nodes interconnected by links, and may be described as networks. The most famous example is the Internet, which is composed of many interacting computers. Other examples are social networks that describe people and the relations between them, and biological networks that describe the interactions between proteins and genes a living cell. All these systems show collective behavior that is not evident from the structure of their individual components.

Recently it was found that most real world networks do not behave according to the classical random graph model, which was introduced by Erdös and Rényi in the 1960's. Instead, the number of links emerging from each node – the node "degree" $k$ – is distributed according to a power-law $P(k) \sim k^{-\lambda}$ (with $2 < \lambda < 3$), as opposed to the Poisson degree distribution of the Erdös-Rényi model. This new class of networks, termed "scale-free", was found to have many anomalous properties not found in the classical model. For example, they were found to have a much shorter average distance, and to be highly resilient to random breakdown of nodes or links.

However, most real world networks also have a "cost" or "weight" associated with each link. Such networks are called "weighted" or "disordered". For example, in the Internet each link has a certain bandwidth, thus limiting the rate of information transfer along this link. When routing information between two

distant computers in the network, we are interested in the rout ("path") with minimal cost – the “optimal path”. In this dissertation we wish to answer the following questions: What is the structure of the optimal path? How does the optimal path behave for different types of disorder? And how are the different optimal paths distributed? The main paradigm in this dissertation is that such "optimization" problems on weighted random graphs may be solved using percolation theory.

The dissertation is organized as follows: In Chapters 1 and 2 we give a brief summary of previous results and common methods used in the field of complex networks. Chapter 2 includes a new simplified formulation for the method of generating functions as applied to percolation on random graphs.

In Chapter 3 we proceed to introduce a new structural characterization of networks, which we term “Tomography”. We examine the structure of “layers” of nodes around the maximally connected node. It is shown that the distance distribution of all nodes from the maximally connected node of the network consists of two regimes. The first is characterized by rapid growth in the number of nodes, and the second decays exponentially. We also show analytically that the nodes degree distribution at each layer is a power law with an exponential cut-off. Empirical results from the Internet show a similar behavior to our model.

We next study the percolation transition in Erdös-Rényi and scale-free graphs. It is known from percolation theory on lattices that the critical probability for the percolation transition is not a sharp threshold; actually it is a region of non-zero width $\Delta p_c$ for systems of finite size. In Chapter 4 we show that for

complex networks $\Delta p_c \sim p_c / l$, where $l \sim N^{\nu_{opt}}$ is the average length of the percolation cluster, and N is the number of nodes in the network. For Erdos-Renyi graphs $\nu_{opt} = 1/3$, while for scale-free networks with a degree distribution $P(k) \sim k^{-\lambda}$ and $3 < \lambda < 4$, $\nu_{opt} = (\lambda - 3)/(\lambda - 1)$. We show analytically and numerically that the "survivability" $S(p,l)$, which is the probability of a cluster to survive l chemical shells at conduction probability p, behaves near criticality as $S(p,l) = S(p_c,l) \cdot \exp\left[\frac{1}{p_c}(p - p_c)l\right]$. Thus for probabilities inside the region $|p - p_c| < \frac{p_c}{l}$ the behavior of the system is indistinguishable from that of the critical point.

Finally, in Chapter 5, we apply percolation theory to optimization problems in networks. We first consider Erdös-Rényi graphs with random weights associated with each link, and show that any disordered random graph contains an inherent scale-free network, which we term the "supernode network". We then show that the minimum spanning tree in Erdös-Rényi graphs is related to this network, and is composed of percolation clusters which are interconnected by a set of links that create a scale-free tree.

We then use the above results to study the behavior of the optimal path in disordered graphs. With each link $i$ we associate a weight $\tau_i = \exp(a r_i)$, where $r_i$ is a random number taken from a uniform distribution between 0 and 1, and the parameter $a$ controls the strength of the disorder. In these systems it was found that the optimal path length $l(a)$ scales as $N^{\nu_{opt}}$ for strong disorder and as $\log N$

for weak disorder. We show that that the average length of the optimal path, for intermediate values of the parameter $a$, obeys the following scaling form: $l(a) = l_\infty \cdot F\left(\frac{1}{p_c}\frac{l_\infty}{a}\right)$, where $l_\infty$ is the optimal path length in strong disorder ($a \to \infty$), and $p_c$ is the percolation threshold. The optimal paths lengths distribution obeys a similar form: $P(l, N, a) = \frac{1}{l_\infty} G\left(\frac{l}{l_\infty}, \frac{1}{p_c}\frac{l_\infty}{a}\right)$. These relations are supported by numerical simulations for Erdös-Rényi and scale-free graphs. Hence, a single "control parameter" $Z \equiv \frac{1}{p_c}\frac{l_\infty}{a}$ controls both the average and distribution of the optimal paths lengths.

We explain this by showing that the optimal path follows the percolation cluster up to a characteristic length $\xi = ap_c$, after which it takes a “shortcut” link (as in the “small-world” model) outside the cluster. The control parameter $Z \equiv \frac{1}{p_c}\frac{l_\infty}{a}$ is actually the number of shortcuts. Thus, the behavior of weighted random graphs is similar to percolation systems near the critical threshold, which have a fractal behavior up to a certain correlation-length $\xi$.

To summarize, in this dissertation we studied structure, percolation, and optimization problems in disordered complex networks. We analyzed the structure of minimal spanning trees, and the average and distribution of optimal-paths lengths. This work demonstrates how practical optimization problems in networks can be solved using methods from statistical physics and percolation theory.

# Chapter 1: Introduction

## 1.1. Networks

Networks (or in mathematical terms, graphs) are mathematical entities composed of nodes and links connecting them. Many systems in nature may be described as “networks” [1-3], for example the Internet, which is composed of millions of interconnected computers. Other examples include social networks, which are used to describe people and the social connections between them (for example terrorist networks), and biological networks, which describe the complex interactions between genes and proteins in a living cell.

The main function of a network is to enable effective communication between distant nodes of the network. For example, in the Internet, any two computers that are connected to the network can (in principle) communicate with each other. In social networks, any two people in the world can usually reach each other through a short chain of social acquaintances. This is possible because the average distance along the network is extremely small: usually the average number of links connecting any two nodes on the network scales logarithmically with the network size. This is also known as the "small world" property.

### 1.1.1. Erdös-Réyni graphs

Graph theory is rooted in the 18$^{th}$ century. Leonhard Euler conducted one of the first studies in graph theory in 1736, thereby solving the problem known as “the

seven bridges of Königsberg". The mathematicians Pal Erdös and Alfred Réyni proposed a special class of graphs called "random graphs" in the 1960's. In the Erdös-Réyni (ER) model [4, 5], we start with N nodes, and then connect any two nodes with some small probability p. The number of links k emerging from every node, also called the node "degree", is a random variable distributed according to a Poisson distribution with the average degree $\langle k \rangle \approx pN$.

Erdös and Réyni found that as the probability p is increased from 0 to 1, the graph topology changes abruptly from a loose collection of small connected clusters ("components") to a system dominated by a single cluster, also called the "Giant Component". The point of transition, also termed "The Critical Point", occurs when the average degree $\langle k \rangle \approx pN = 1$. Below the critical point the graph is composed of small clusters with a negligible number of loops (trees), and the size of the largest cluster is proportional to $\log N$. As the probability p increases, trees of increasing order (size) appear. At the critical point an abrupt change occurs: A "giant component" appears, its size is proportional to $N^{2/3}$, trees of all possible order are present, and loops of all orders appear. Above the critical point the giant component dominates the system: Its size is proportional to N, and loops are present inside it[1]. Another property is that the average distance between any two nodes inside the giant component is proportional to $\log N$ [4, 6]. Hence we get the "small world" property above the critical point.

[1] However, the other clusters (also termed the "finite components") still have a negligible number of loops.

### 1.1.2. Scale-free networks

As mentioned above the degree distribution in ER graphs is Poissonian, that is: $P(k) = e^{-z}\frac{z^k}{k!}$, where $P(k)$ is the fraction of nodes having degree $k$, and $z = \langle k \rangle$ is the average degree. However, most networks in the real world show a degree distribution that is clearly not Poissonian, rather it has a form of a power law: $P(k) \sim k^{-\lambda}$ where the exponent $\lambda$ is usually in the range $2 \le \lambda \le 3$. These networks are termed "scale-free" (SF). The power law degree distribution implies that the degrees of the nodes are not homogeneous as in ER graphs; rather there are a few nodes with a very large degree ("hubs") and many nodes with very small degree. Scale-free degree distributions were observed in communication networks such as the Internet [5, 7, 8], in biological networks [9, 10], in social networks [11], and many other systems.

Barabasi proposed a simple model for generating SF networks [5]. In the Barabasi-Albert (BA) model we start with an initial number of nodes. We then add new nodes and connect them to existing nodes, where the probability that a new node will connect to an already existing node is proportional to the existing node's degree ("Preferential Attachment"). A different method to construct SF networks is the "Molloy-Reed" construction[2]: initially the degree of each node is chosen according to a scale-free distribution, where each node is given a number of open links or "stubs" according to its degree. Then, stubs from all nodes of the network are interconnected randomly to each other.

---

[2] This model was actually proposed by Bollobás.

The exact form of the degree distribution is usually taken to be [12]:

$$P(k) = ck^{-\lambda}, \qquad k = m,...,K \tag{1.1}$$

where m and K are the minimal and maximal degrees, and $c \approx (\lambda - 1)m^{\lambda-1}$ is a normalization constant[3]. For real networks with finite size, the highest degree K depends on network size N: $K \approx mN^{1/(\lambda-1)}$, thus creating a "natural" cutoff for the highest possible degree[4].

Because of their in-homogeneity, scale-free networks have anomalous properties: networks with $2 < \lambda < 3$ have an average distance which is proportional to $\log\log N$ (which is much shorter than the average distance for ER graphs, which scales as $\log N$ ) [13], They are also highly resilient to random failure of nodes or links (because their percolation threshold vanishes for large systems - see below) [12], and they have different critical exponents than ER graphs (see below) [14].

---

[3] The normalization constant is found using the relation [12]:

$$1 = \sum_{k=m}^{\infty} P(k) \approx c \int_{k=m}^{\infty} k^{-\lambda} dk = \frac{c}{-\lambda+1}\left(0 - m^{-\lambda+1}\right) = c \cdot \frac{m^{-\lambda+1}}{\lambda-1},$$

and thus: $c \approx (\lambda - 1)m^{\lambda-1}$.

[4] An estimate of this cutoff can be found by the assumption that the tail of the distribution above K is of the order of one site [12]:

$$\frac{1}{N} = \sum_{k=K}^{\infty} P(k) \approx c \int_{k=K}^{\infty} k^{-\lambda} dk = c \frac{k^{-\lambda+1}}{-\lambda+1}\Bigg|_{K}^{\infty} = (\lambda-1)m^{\lambda-1} \cdot \frac{1}{-\lambda+1}\left(0 - K^{-\lambda+1}\right) = m^{\lambda-1}K^{-\lambda+1},$$

which gives: $K \approx mN^{1/(\lambda-1)}$.

# 1.2. Percolation

## 1.2.1. Percolation theory

Various phenomena in nature may be described by percolation theory, for example: forest fires, epidemic spreading and conduction in disordered systems [15]. A typical percolation system consists of a d-dimensional grid of length L, in which the nodes or links are removed with some probability $1-p$, or are considered “conducting" with probability $p$ [15, 16]. Below some critical probability $p_c$ the system becomes disconnected into small clusters, i.e., it becomes impossible to cross from one side of the grid to the other by following the conducting links. However, above $p_c$ a spanning cluster emerges, similar to the giant component in random graphs. Percolation may be considered as a geometrical phase transition exhibiting universality, critical exponents, and an upper critical dimension at $d_c$=6 – see Figures 1-1 and 1-2.

Some of the system parameters which were found to have universal behavior are the size of the largest cluster[5]: $P_\infty \sim \left(p-p_c\right)^\beta$, the correlation length: $\xi \sim \mid p-p_c \mid^{-\nu}$, the mean size of finite clusters: $S \sim \mid p-p_c \mid^{-\gamma}$, and the cluster sizes distribution: $n_s(p) \sim s^{-\tau} \cdot f\left(\left(p-p_c\right)s^\sigma\right)$. The exponents $\beta,\gamma,\nu$ etc. are called "critical exponents". Table 1-1 gives the values of some of the critical exponents for percolation systems. The "universality" of the critical exponents means that they do not depend on the local structure of the grid (i.e. square, triangle etc.),

[5] $P_\infty$ is defined as the probability that a randomly chosen node resides in the giant component.

rather only on the *dimension* d of the system. The critical threshold $p_c$, though, is not a universal quantity. Notice also that for dimensions above the *upper critical dimension* $d_c$=6 the critical exponents do not change.

Stanley [15] found that at the percolation threshold $p_c$ the spanning cluster is a fractal, i.e., the mass of the spanning cluster scales as: $M \sim L^{d_f}$, where L is the geometrical length of the system and $d_f$ is the *fractal dimension* which is generally not an integer in the case of fractals. The fractal dimension $d_f$ is also a universal quantity (see Table 1-1). Above and below the percolation threshold the fractal behavior is still present only for length scales below the correlation length $\xi$, that is:

$$M \sim \begin{cases} L^{d_f} & L < \xi \\ L^{d} & L > \xi \end{cases}.$$

At criticality, $\xi \to \infty$.

When measuring the distance between two points on a percolation cluster we can distinguish between the "geometrical distance" L (or $r$) and the "chemical distance" $l$. The geometrical distance is the regular Euclidian distance between two points in a d-dimensional lattice, whereas the chemical distance is the distance along the links of the percolation cluster[6]. Thus we also define a "chemical dimension" $d_l$, which gives the relation between the chemical size of the percolation cluster and its mass: $M \sim r^{d_f} \sim l^{d_l}$. In random graphs, it is more

---

[6] For lattices with dimension $d \geq d_c = 6$ a path along the percolation cluster is uncorrelated, and like a random walk obeys $r^2 \sim l$. Therefore, above the upper critical dimension, $\nu_l = 2\nu_r$ (see [15, 16] for details)

convenient to refer to the chemical distance because the geometrical distance has no meaning.

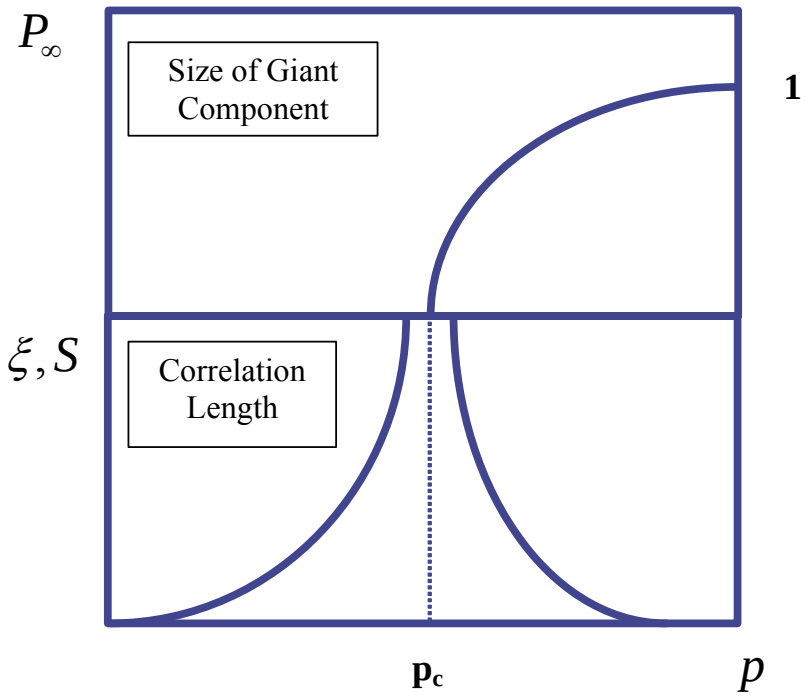
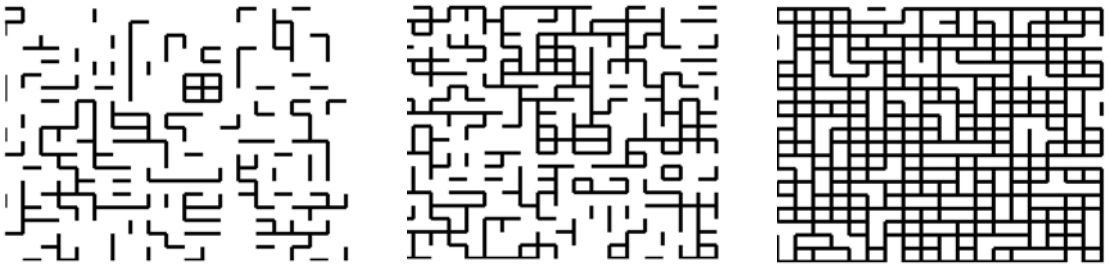



**Figure 1-1:** A sketch of the dependence of the size of the giant component, the mean finite cluster size, and the correlation length, on the conduction probability $p$. At the critical threshold $p_c$ the giant component appears, clusters of all order (size) are present, and the correlation length diverges (i.e., it is proportional to the system size L).



$p < p_c$ $p = p_c$ $p > p_c$

**Figure 1-2:** A demonstration of percolation on a two-dimensional grid. Below the percolation threshold the system is composed of small clusters. At $p = p_c$ a spanning cluster with a fractal structure appears, and clusters of all sizes are present. Well above $p_c$ the finite clusters merge into the giant component. In this particular case (square grid, d=2, bond percolation) $p_c = 1/2$.

| Exponent | d=2 | d=3 | d=4 | d=5 | d=6-ε | ER Network (d≥6) | SF Network ($3<\lambda<4$) |
|---|---|---|---|---|---|---|---|
| $\beta$ | 5/36 | 0.417 | 0.64 | 0.84 | $1-\varepsilon/7$ | 1 | $\frac{1}{\lvert\lambda-3\rvert}$ |
| $\gamma$ | 43/18 | 1.795 | 1.44 | 1.18 | $1+\varepsilon/7$ | 1 | 1 |
| $\nu_r$ (geometric) | 4/3 | 0.875 | 0.68 | 0.57 | $\frac{1}{2}+\frac{5\varepsilon}{84}$ | 1/2 | - |
| $\nu_l$ (chemical) | | | | | | 1 | 1 |
| $\sigma$ | 36/91 | 0.45 | 0.48 | 0.49 | $\frac{1}{2}+O(\varepsilon^2)$ | 1/2 | $\frac{\lvert\lambda-3\rvert}{\lambda-2}$ |
| $\tau$ | 187/91 | 2.18 | 2.31 | 2.41 | $\frac{5}{2}-\frac{3\varepsilon}{14}$ | 5/2 | $\frac{2\lambda-3}{\lambda-2}$ |
| $d_l$ | 1.678 | 1.84 | | | | 2 | $\frac{\lambda-2}{\lambda-3}$ |
| $d_f$ | 91/48 | 2.524 | 3.06 | 3.54 | $4-\frac{10\varepsilon}{21}$ | 4 | $2\frac{\lambda-2}{\lambda-3}$ |
| $d_c$ | 6 | 6 | 6 | 6 | 6 | 6 | $2\frac{\lambda-1}{\lambda-3}$ |

**Table 1-1:** Critical exponents for percolation systems of different dimensions

## 1.2.2. Percolation in networks

Percolation theory may be applied to random graphs. Randomly removing a fraction $1-p$ of the links of an ER graph corresponds to changing the average degree $\langle k\rangle$ to $\langle k_{diluted}\rangle = p\langle k\rangle$ in the diluted graph. According to the results of

Erdös, a giant component exists as long as $\langle k_{diluted} \rangle = p\langle k \rangle \geq 1$. Thus the critical probability for the emergence of a giant component is: $p_c = 1/\langle k \rangle$.

In finite dimensional grids, the number of neighbors each node may have increases with the dimension d. In this sense, networks correspond to systems of infinite dimension, because the number of neighbors each node may have increases with the size of the system. To summarize, the process of randomly diluting links (or nodes) from a network may be described by percolation theory, where the system dimension is taken to be $d = \infty$. Note that all systems above the upper critical dimension $d_c = 6$ behave the same near the critical point.

It was found by Albert *et al.* [17] that the Internet, which can be modeled as a SF network with $\lambda \approx 2.5$, is relatively resilient random removal of nodes. Cohen *et al*. studied the corresponding problem of percolation on SF networks [12, 14]. It was found that the general expression for the critical threshold is:

$$p_c = \frac{\langle k \rangle}{\langle k(k-1) \rangle} = \frac{1}{\kappa_0 - 1} \tag{1.2}$$

Where $\kappa_0 - 1$ is the *branching factor*, i.e., the average number of links emerging from a node reached by following a random link, and $\kappa_0 = \langle k^2 \rangle / \langle k \rangle$. For SF networks whose degree distribution is described by equation (1.1),

$$\kappa_0 = \left( \frac{2-\lambda}{3-\lambda} \right) \frac{K^{3-\lambda} - m^{3-\lambda}}{K^{2-\lambda} - m^{2-\lambda}} \sim \begin{cases} m, & if \quad \lambda > 3 \\ m^{\lambda-2} K^{3-\lambda}, & if \quad 2 < \lambda < 3 \\ K, & if \quad 1 < \lambda < 2 \end{cases}, \tag{1.3}$$

where $K = mN^{1/(\lambda-1)}$. In the case of $2 < \lambda < 3$, the second moment of the degree distribution diverges, and thus $\kappa_0$ diverges with system size. The result of this is that $p_c \sim N^{-\frac{3-\lambda}{\lambda-1}} \to 0$ for large N.

It was also found [14], using the method of generating functions (see Chapter 2.2), that the critical exponents in SF networks with $3 < \lambda < 4$ are different than for ER graphs, thus creating a different *universality class*[7], see Table 1-1.

In Chapter 4 we will show that for networks of finite size N the percolation threshold is not a sharp point; rather it has a width $\Delta p_c$, such that in the range $p \in \left[ p_c - \Delta p_c, p_c + \Delta p_c \right]$ all properties of criticality (e.g. a negligible number of loops) are present.

# 1.3. Optimization in networks

## 1.3.1. Weighted networks

The function of most real world networks is to connect distant nodes, either by transfer of information (e.g. the Internet), or through transportation of people and goods (such as networks of roads and airlines). In many cases there is a “cost" or a “weight" associated with each link, and the larger the weight on a link, the harder it is to traverse this link. In this case, the network is called “disordered" or

[7] For $\lambda \geq 4$ the critical exponents are equal to those of ER graphs. For the anomalous regime $2 < \lambda < 3$ the critical exponents are not fully known yet.

"weighted" [18]. For example, in the Internet each link between two routers has a bandwidth or delay time, in a transportation network some roads may have only one lane while others may be highways allowing for large volumes of traffic. When modeling random weighted networks, it is commonly assumed that each link is associated with a weight $w_i = \exp(ar_i)$, where $r_i$ is a random number taken from a uniform distribution between 0 and 1, and the parameter $a$ controls the strength of the disorder. Networks with large values of $a$ can be interpreted as networks with large fluctuations in link weights: both very small and very large weights are present.

### 1.3.2. Optimal path and the Shortest Path Tree (SPT)

We define the *optimal path* to be the path between two nodes on the network such that the sum of weights along this path is minimal. This path may be much longer than the *shortest-hopcount-path*, which is the shortest path without regarding the weights on the links. One of the most popular algorithms for finding the optimal path is Dijkstra's Algorithm [19], which builds a tree of shortest paths - the *Shortest Paths Tree* (SPT) - starting from some chosen node to all nodes of the network.

### 1.3.3. The Minimum Spanning Tree (MST)

Another type of optimal tree that can be built on the network is the *Minimum Spanning Tree (*MST). This is a tree that spans the whole graph, and whose total weight is minimal. There are two algorithms to build this tree: Prim's algorithm and Kruskal's algorithm [19]. It was shown [20] that any path between two nodes

along the MST is the path with the minimal *barrier* (i.e., minimal maximal-weight) possible between those two nodes.

In Chapter 5 we will describe the structure of the MST using percolation theory. We will show that the MST is composed of percolation clusters interconnected by a scale-free tree. The optimal paths in networks with strong disorder ($a \to \infty$) follow the MST, and therefore their behavior can also be described by percolation theory. Moreover, in networks with intermediate disorder (i.e. intermediate values of the parameter $a$) the optimal path follows the MST up to some characteristic length $\xi = ap_c$. In this sense there is a similarity between percolation systems near the critical point and disordered systems with large fluctuations on the weights of the links.

## 1.4. Outlook and applications

In the last years, extensive research has been done on networks. Models for generating networks were developed [5] and many important structural properties were found (e.g. [5, 12, 13, 17]). However, most real world networks have a weight associated with each link, and this was not taken into account.

In this dissertation we approach this problem by studying the relation between percolation and optimization. We first study structural properties of networks and their behavior near the percolation critical point. Then, we apply percolation theory to optimization problems on weighted networks, and explain the structure of minimum spanning trees and the behavior of the optimal paths on the network. The ideas presented here demonstrate how optimization problems

in networks can be solved using statistical physics and percolation theory. From a practical point of view, our results may be helpful for devising new searching and routing algorithms on the Internet, and for generating efficient multicast trees [21, 22].

# Chapter 2: Methods and Previous Results

In this Chapter we present main results and common methods from the field of complex networks, which will be used throughout this dissertation.

## 2.1. Chemical length of percolation cluster in networks:

Let us consider a grid of dimension d=6, in which a fraction $1-p_c$ of the links have been removed. According to percolation theory the mass of the percolation cluster scales as:

$$M \sim L^{d_f} \sim l^{d_l}$$

Where L is the geometric size of the system[8] and $l$ is the chemical length of the percolation cluster. $l$ is actually proportional to the average distance between any two points on the percolation cluster when we are restricted to go only along links lying inside the percolation cluster. The total number of nodes in the system is:

$$N \sim L^d$$

Thus:

$$l \sim L^{\left(d_f/d_l\right)} \sim \left[N^{(1/d)}\right]^{\left(d_f/d_l\right)} \sim N^{\left(d_f/d_l d\right)}$$

[8] The percolation cluster spans the system. Therefore L is also the geometric size of the percolation cluster.

Or: $l \sim N^{\nu_{opt}}$, where: $\nu_{opt} \equiv \frac{d_f}{d_l d}$.

Because of universality, this relation holds also for random graphs (where only the chemical distance $l$ has a meaning). Random graphs may be treated as infinite dimensional systems ($d \geq d_c = 6$)[9].

***For ER graphs:***

$d = 6$, $d_f = 4$, and $d_l = 2$ (see Table 1-1). Thus:

$$\nu_{opt} \equiv \frac{d_f}{d_l d} = \frac{4}{2 \cdot 6} = \frac{1}{3}.$$

***For SF graphs:*** (with[10] $3 < \lambda < 4$)

$d = 2\frac{\lambda - 1}{\lambda - 3}$, $d_f = 2\frac{\lambda - 2}{\lambda - 3}$, and $d_l = \frac{\lambda - 2}{\lambda - 3}$ (see Table 1-1). Thus

$$\nu_{opt} \equiv \frac{d_f}{d_l d} = \frac{2\frac{\lambda - 2}{\lambda - 3}}{\frac{\lambda - 2}{\lambda - 3} \cdot 2\frac{\lambda - 1}{\lambda - 3}} = \frac{\lambda - 3}{\lambda - 1}.$$

The mean distance between two points in a random graph scales as $\log N$, where N is the number of nodes in the graph [6]. However, when the graph is diluted such that a fraction $1 - p_c$ of its links are removed, the average distance increases dramatically to $l \sim N^{1/3}$ (for ER graphs) or $l \sim N^{(\lambda-3)/(\lambda-1)}$ (for SF graphs with $3 < \lambda < 4$), which is much longer than $\log N$ [18]. This conforms with results

---

[9] For random graphs $d = \infty > d_c$. This means that at the percolation threshold, the diluted graph can be embedded into a space with dimension $d_c$. This embedding dimension is $d_c = 6$ for ER graphs and $d_c = 2(\lambda - 1)/(\lambda - 3)$ for SF graphs (with $3 < \lambda < 4$) – see Table 1-1.

[10] Note that for $\lambda \geq 4$, SF graphs behave like ER graphs (near the critical point).

for finite dimensional systems [23], in which typical paths along the percolation cluster were found to follow long and twisted trajectories – much longer than the Euclidian distance - because of the fractal nature of the percolation cluster.

## 2.2. Generating functions

A general method for studying percolation on random graphs with arbitrary degree distributions was first introduced by Molloy and Reed [24, 25], who suggested viewing the giant component as being “explored”, and used differential equations for the number of unexplored links and nodes to find the size of the giant component[11] and the degree distribution of the residual graph (i.e., the finite clusters) [12].

An alternative method is to use generating functions [26]. This powerful tool was introduced by Newman, Watts, and Strogatz [27], who used this method to calculate the size of the giant component and the cluster sizes distribution in SF networks. Later, Cohen *et al*. used this method to find the percolation critical exponents [14] (see Table 1-1).

---

[11] We used a similar procedure in Chapter 3.

[12] We note that the problem of percolation on networks can be solved analytically due to the fact that we are dealing with a system whose dimension d is above the upper critical dimension ($d_c$=6), or equivalently, due to the fact that the number of loops (at criticality) is negligible. This is also termed in physics as the "mean-field" regime.

### 2.2.1. Definition

Given a network with a degree distribution $P(k)$, which is the probability that a randomly chosen node will have $k$ neighbors, we can define the probability generating function:

$$G_0(x) = \sum_{k=0}^{\infty} P(k) x^k .$$

The coefficient of $x^k$ is the probability that a randomly chosen node will have *k* neighbors. Note that $G_0(1) = \sum_{k=0}^{\infty} P(k) = 1$ because the degree distribution is normalized. The average degree is given by $\langle k \rangle = \sum_{k=1}^{\infty} kP(k) = G_0'(1)$.

Now, choose a random *link* and consider a node at one of its ends. The probability $P_1(k)$ that a node thus reached by *following a random link* will have a degree *k* is[13]: $P_1(k) = \frac{1}{\langle k \rangle} kP(k)$. For our purposes it is useful to consider only the number of *outgoing* links – not including the incoming link. The corresponding generating function is:

$$G_1(x) = \sum_{k=1}^{\infty} P_1(k) x^{k-1} = \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} kP(k) x^{k-1}$$

Note that $G_1(x) = \frac{G_0'(x)}{G_0'(1)}$, and that $G_1(x)$ is normalized such that $G_1(1) = 1$.

---

[13] This is equal to the fraction of edges in the network that connect to a node of degree *k*, which is proportional to $k \cdot NP(k)$ (there are $NP(k)$ nodes of degree *k*, and each one has *k* such links connected to it).

### 2.2.1.1. Example: Erdös-Réyni graphs

Take for example an ER graph. In the Erdös-Réyni (ER) model, we start with N nodes, and then connect any two nodes with some small probability p. The degree distribution is:

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k} \approx e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!},$$

where $\langle k \rangle \approx pN$ and N$\to\infty$. Thus, the number of links, $k$, emerging from every node is a random variable distributed according to a Poisson distribution with the average degree $\langle k \rangle \approx pN$. The generating functions are:

$$G_0(x) = \sum_{k=0}^{\infty} P(k) x^k = e^{-\langle k \rangle} \sum_{k=0}^{\infty} \frac{(\langle k \rangle x)^k}{k!} = e^{-\langle k \rangle} e^{x\langle k \rangle} = e^{\langle k \rangle (x-1)},$$

and:

$$G_1(x) = \frac{G_0'(x)}{G_0'(1)} = \frac{\langle k \rangle e^{\langle k \rangle (x-1)}}{\langle k \rangle} = e^{\langle k \rangle (x-1)}.$$

Notice that in ER graphs $G_0(x) = G_1(x)$. However, this is not true in general (for example, in SF networks).

## 2.2.2. Percolation

Given a random graph with a degree distribution $P(k)$, we can find the probability generating functions $G_0(x)$ and $G_1(x)$. Assume that a fraction $1-p$ of the links is randomly removed. How do the functions $P(k)$, $G_0(x)$, and $G_1(x)$ change ?

After randomly removing a fraction $1-p$ of the links (bond percolation), the probability for a randomly chosen node to have k remaining links in the diluted graph is given by [12]:

$$\tilde{P}(k)=\sum_{k_0=k}^{\infty}P(k_0)\binom{k_0}{k}p^k(1-p)^{k_0-k}$$

The corresponding probability generating functions $\tilde{G}_0(x)=\sum_{k=0}^{\infty}\tilde{P}(k)\cdot x^k$ and $\tilde{G}_1(x)=\sum_{k=1}^{\infty}\tilde{P}_1(k)\cdot x^{k-1}$ in the diluted graph are:

$$\tilde{G}_0(x)=\sum_{k=0}^{\infty}\tilde{P}(k)\cdot x^k=\sum_{k=0}^{\infty}\left[\sum_{k_0=k}^{\infty}P(k_0)\binom{k_0}{k}p^k(1-p)^{k_0-k}\right]\cdot x^k=$$
$$=\sum_{k_0=0}^{\infty}P(k_0)\sum_{k=0}^{k_0}\binom{k_0}{k}(xp)^k(1-p)^{k_0-k}=$$
$$=\sum_{k_0=0}^{\infty}P(k_0)(1-p+px)^{k_0}=G_0(1-p+px)$$

And[14]:

$$\tilde{G}_1(x)=\frac{\tilde{G}_0'(x)}{\tilde{G}_0'(1)}=\frac{pG_0'(1-p+px)}{pG_0'(1)}=\frac{G_0'(1-p+px)}{G_0'(1)}=G_1(1-p+px)$$

[14] It can be shown that: $\tilde{G}_1(x)=G_1(1-p+px)=\frac{1}{\langle k\rangle}\sum_{k=1}^{\infty}kP(k)\cdot(1-p+px)^{k-1}$ is equal to $\tilde{G}_1(x)=\frac{1}{\langle k_{diluted}\rangle}\sum_{k=1}^{\infty}k\tilde{P}(k)\cdot x^{k-1}$, where $\langle k\rangle$ and $P(k)$ are the average and distribution of the degrees in the original graph, and $\langle k_{diluted}\rangle=p\langle k\rangle$ and $\tilde{P}(k)$ are the average and distribution of the degrees in the diluted graph.

For example, in ER graphs, $\tilde{G}_1(x) = \tilde{G}_0(x) = e^{\langle k \rangle([1-p+px]-1)} = e^{\langle k \rangle p(x-1)}$. This means that the dilution of links in an ER graph changes the average degree from $\langle k \rangle$ to $\langle k_{diluted} \rangle = \langle k \rangle p$ .

In the following sections, when dealing with percolation on networks, we will refer to the generating functions $\tilde{P}(k)$, $\tilde{G}_0(x)$, and $\tilde{G}_1(x)$ of the *diluted* graph unless otherwise specified.

### 2.2.2.1. Calculating cluster sizes

Consider a random graph below the percolation threshold. In this regime all clusters ("components") are trees, that is, the number of loops is negligible[15] [4, 5]. Let $H_1(x)$ be the generating function for the size of a component reached by choosing a random link and following it to one of its ends. $H_1(x)$ actually describes the distribution sizes of a randomly chosen "branch".

$H_1(x)$ obeys the following self-consistent relation [27]:

$$H_1(x) = x\tilde{G}_1(H_1(x)) \tag{2.1}$$

This equation means that the probability to reach a branch of $i$ nodes (by following a link) is equal to the probability of following a link, reaching a single node, and then branching out to reach a total of $i-1$ nodes through zero, one, two, etc. branches (see Fig. 2-1).

[15] Our analysis regards only the *finite* clusters which have a tree-like structure. Below the percolation threshold all clusters are finite, while above the percolation threshold an infinite cluster (the "giant component") appears. The giant component has loops and has to be explicitly excluded from the following analysis.

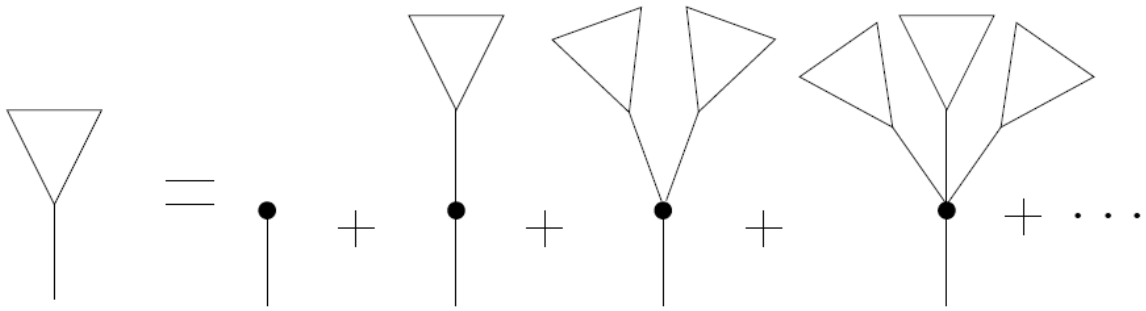


**Figure 2-1:** A graphical sketch of the recursion relation (2.1). The probability to reach a branch consisting of $i$ nodes is equal to the sum of probabilities to reach a single node, to reach a node connected to a single branch of size $i-1$, to reach a node connected to two branches having a total size of $i-1$, etc.

In order to demonstrate this, let us write: $H_1(x)=\sum_{i=0}^{\infty} h_i x^i$ and $\tilde{G}_1(x)=\sum_{k=1}^{\infty} P_1(k)\,x^{k-1}$ .

Substituting into Eq. (2.1) we get:

$$\begin{aligned} h_0+h_1x+h_2x^2+\ldots = x\Big[ & P_1(1)+ \\ & +P_1(2)\cdot\left(h_0+h_1x+h_2x^2+\ldots\right)+ \\ & +P_1(3)\cdot\left(h_0+h_1x+h_2x^2+\ldots\right)^2+ \\ & +P_1(4)\cdot\left(h_0+h_1x+h_2x^2+\ldots\right)^3+ \\ & +\ldots\Big] \end{aligned}$$

We will take for our demonstration the coefficients of $x^0$, $x^1$, $x^2$, and $x^3$ (See Figure 2-2).

***The coefficients of*** $x^0$ ***:***

$$h_0=0$$

The probability to follow an existing link and to reach no node is zero.

***The coefficients of*** $x^1$ ***:***

$$h_1 = P_1(1) + P_1(2)\cdot h_0 + P_1(3)\cdot h_0^2 + P_1(4)\cdot h_0^3 + \ldots =$$
$$= P_1(1)$$

The probability to reach a branch consisting of a single node is equal to the probability to follow a link into a node of degree 1 (There is also the possibility to reach a node with higher degree, in which every outgoing link leads to a zero-sized branch, but the probability for this is zero, as we have previously shown).

***The coefficients of*** $x^2$***:***

$$h_2 = P_1(2)\cdot h_1 + P_1(3)\cdot 2h_0h_1 + P_1(4)\cdot 3h_0^2h_1 + \ldots =$$
$$= P_1(2)\cdot h_1 =$$
$$= P_1(2)\cdot P_1(1)$$

A two-node branch is reached if we follow a link to a node of degree 2, and follow its single outgoing link into a branch of size 1 – see Fig. 2-2(b).

***The coefficients of*** $x^3$***:***

$$h_3 = P_1(2)\cdot h_2 + P_1(3)\cdot\left(2h_0h_2 + h_1^2\right) + P_1(4)\cdot\left(3h_0h_1^2 + 3h_2h_0^2\right) + \ldots =$$
$$= P_1(2)\cdot h_2 + P_1(3)\cdot h_1^2 =$$
$$= P_1(2)\cdot P_1(2)\cdot P_1(1) + P_1(3)\cdot P_1(1)\cdot P_1(1)$$

A three node branch can be built in two ways, as shown in Fig. 2-2(c):

a) By following a link to a node of degree 2 in which the single outgoing link reaches a branch of size 2.

b) By following a link into a node of degree 3, in which every one of the two outgoing links reaches a branch consisting of a single node.

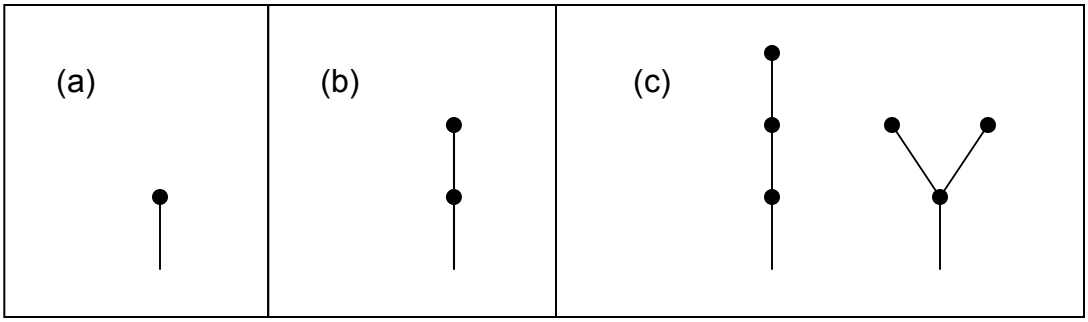



**Figure 2-2:** The possible ways for building a branch of (a) one, (b) two, and (c) three nodes, as described by Eq. (2.1).

If we start from a random node, we have one branch at each end of every link emerging from that node, and all these branches constitute the cluster ("component") in which our node resides in. We set $H_0(x)$ to be the generating function describing the size of a component in which a randomly chosen node resides. Similar to Eq. (2.1) we can write [27]:

$$H_0(x) = x\tilde{G}_0\left(H_1(x)\right) \tag{2.2}$$

Given the generating functions $\tilde{G}_0(x)$, and $\tilde{G}_1(x)$, In order to find the cluster sizes (which are given by $H_0(x)$) we have to solve Eq. (2.1) for $H_1(x)$ and substitute into Eq. (2.2).

The above equations describe the finite components, which have a tree-like structure. Below the percolation threshold, $H_0(1) = 1$ (and also $H_1(1) = 1$), because this is the probability for a randomly chosen node to belong to a finite cluster (or branch) of any size. However, above the percolation threshold $H_0(x)$ is not normalized, because it does not include the giant component. The size of the giant component is given by:

$$P_\infty = 1 - H_0(1) = 1 - \tilde{G}_0(u)$$

where $u \equiv H_1(1)$ obeys the self consistent relation:

$$u = \tilde{G}_1(u)$$

Or in terms of the probability distribution function (of the diluted graph):

$$P_\infty(p) = 1 - \sum_{k=0}^{\infty} \tilde{P}(k) u^k \quad \text{where:} \quad u = \frac{1}{p\langle k \rangle} \sum_{k=1}^{\infty} k \tilde{P}(k) u^{k-1} .$$

where p is the dilution probability[16].

### 2.2.2.2. The critical threshold for percolation

Assume our graph is diluted such that it is below the percolation threshold (i.e. $p < p_c$). The cluster sizes distribution is described by the two equations: $H_1(x) = x\tilde{G}_1(H_1(x))$ and $H_0(x) = x\tilde{G}_0(H_1(x))$.

Assume that $H_0(x) = \sum_{i=0}^{\infty} h_i x^i$. The average cluster size (to which a randomly chosen node belongs) is: $\langle s \rangle = \sum_{i=1}^{\infty} i h_i = H_0'(1)$. Using Eq. (2.2) we get:

$$\langle s \rangle = H_0'(1) = 1 + \tilde{G}_0'(1) \cdot H_1'(1)$$

From Eq. (2.1) we have: $H_1'(1) = 1 + \tilde{G}_1'(1) H_1'(1) \Rightarrow H_1'(1) = \frac{1}{1 - \tilde{G}_1'(1)}$. Thus:

$$\langle s \rangle = 1 + \tilde{G}_0'(1) \cdot \frac{1}{1 - \tilde{G}_1'(1)} .$$

---

[16] Note that the average degree in the diluted graph is $\langle k_{diluted} \rangle = p\langle k \rangle$, where $\langle k \rangle$ is the average degree of the original graph.

At the percolation threshold components of all orders appear, i.e. the average cluster size diverges. It can be seen that the average cluster size diverges when $\tilde{G}_1'(1)=1$. Thus the percolation threshold is given by:

$$1=\frac{d}{dx}\tilde{G}_1(x)\bigg|_{x=1}=\frac{d}{dx}G_1(1-p+px)\bigg|_{x=1}=G_1'(1-p+px)\cdot p\bigg|_{x=1}=G_1'(1)\cdot p$$

Thus:

$$p_c=\frac{1}{G_1'(1)}=\frac{1}{\frac{1}{\langle k\rangle}\sum_{k=2}^{\infty}k(k-1)P(k)}=\frac{\langle k\rangle}{\langle k(k-1)\rangle}.$$

This conform well with previous results [12]. For example, for ER graphs: $G_1(x)=e^{\langle k\rangle(x-1)}$ and $G_1'(x)=\langle k\rangle e^{\langle k\rangle(x-1)}$, resulting in the well known result: $p_c=\frac{1}{G_1'(1)}=\frac{1}{\langle k\rangle}$.

## 2.3. Summary and Conclusions

In this Chapter we gave a brief review of previous results, and have introduced the main methods that will be used in this dissertation. The first result, derived from percolation theory, shows that random graphs, when diluted to the critical threshold, lose their relatively small average distance (also known as the "small world" property), i.e., the average path length between two points increases exponentially from $\log N$ to $N^{\nu_{opt}}$.

In the second part of this Chapter we introduced a simplified formulation of the method of generating functions, and showed how this method may be used

to derive the critical probability for percolation, and the size of the giant component for random graphs with an arbitrary degree distribution. In the next Chapters we will apply these methods to Erdös-Réyni and scale-free networks.

# Chapter 3: Tomography of scale-free networks

## 3.1. Introduction

In this Chapter we study the structure of scale-free networks with a degree distribution of the form $P(k) \sim k^{-\lambda}$. We examine the "Tomography" [28, 29] of these networks, i.e. the structure of layers around a single network node. It is shown that the distance distribution of all nodes from the maximally connected node of a random scale-free network consists of two regimes. The first is characterized by a rapid growth in the number of nodes, and the second decays exponentially. We also show analytically that the nodes degree distribution at each layer is a power law with an exponential cut-off. Empirical results from the Internet show a similar behavior to our model.

The Chapter is organized as follows: first, we describe the process of generating the network, and define our terminology. Then, we analyze the degree distribution at each layer surrounding the maximally connected node. The results presented here are based on [28, 29].

## 3.2. Model description

We base our construction on the Molloy-Reed model [24, 25], also described in the introduction. The construction process tries to gradually expose the network,

similar to the method introduced in [13], thus creating a hierarchy in the Molloy-Reed model, and enabling us to define layers in the graph.

We start by setting the number of nodes in the network, N. We then choose the nodes degrees according to the scale-free distribution function $P(k) = cK^{-\lambda}$, where $c \approx (\lambda - 1)m^{\lambda - 1}$ is the normalizing constant and the degree k is in the range $[m, K]$, for some chosen minimal degree m and the natural cutoff $K = mN^{1/(\lambda - 1)}$ of the distribution [12]. At the first stage each node in the network has a given number of outgoing links, which we term "open connections" (or "stubs"), according to its chosen degree. Let us define V as the set of N chosen nodes, C as the set of unconnected outgoing links from the nodes in V, and E as the set of edges in the graph. Using these definitions, the set of links in E is empty at this point, while the set of outgoing open links in C contains all unconnected outgoing links in the graph. In the Molloy-Reed construction described in the introduction, the links in C are randomly matched, such that at the end of the process, C is empty, and E contains all the matched links $\langle u, v \rangle$ that connect pairs of nodes $u, v \in V$ in the network.

Instead, here we proceed as follows: we start from the maximal degree node, which has a degree K, and connect it randomly to K available open connections, thus removing these open connections from C (see Fig. 3-1(a)). The nodes that were now connected to the maximal degree node are termed as the first "layer". We have actually *exposed* the first layer (or chemical shell) of nodes, indexed as l=1. We now continue to fill out the second layer l=2 in the same way: We connect all open connections emerging from nodes in layer 1 to

randomly chosen open connections from C. These open connections may be chosen from nodes of layer 1 (thus creating a loop) or from other links in C. We continue until all open connections emerging from layer 1 have been connected, thus filling layer l=2 (see Fig. 3-1(b)). Generally, to form layer l+1 from an arbitrary layer l, we randomly connect all open connections emerging from l to either other open connections emerging from l or chosen from the other links in C (see Fig. 3-1(c)). Note, that when we have formed layer l+1, layer l has no more open connections. The process continues until the set of open connections, C, is empty.

## 3.3. Theory

We proceed now to evaluate the probability for nodes with degree k to reside outside the first l layers, denoted by $P_l(k)$. The number of open connections outside layer l, is given by:

$$T_l = N\sum_{k=1}^{\infty} kP_l(k)$$

The probability that a detached node with degree k will be connected to an open connection emerging from layer l is: $\frac{k}{\chi_l + T_l}$, where $\chi_l$ is the number of open connections emerging from layer l (see Fig. 3-1(b)). Therefore, the conditional probability for a node with degree k to be also outside layer l+1, given that it is outside layer l, is the probability that it does not connect to *any* of the $\chi_l$ open connection emerging from layer l, that is:

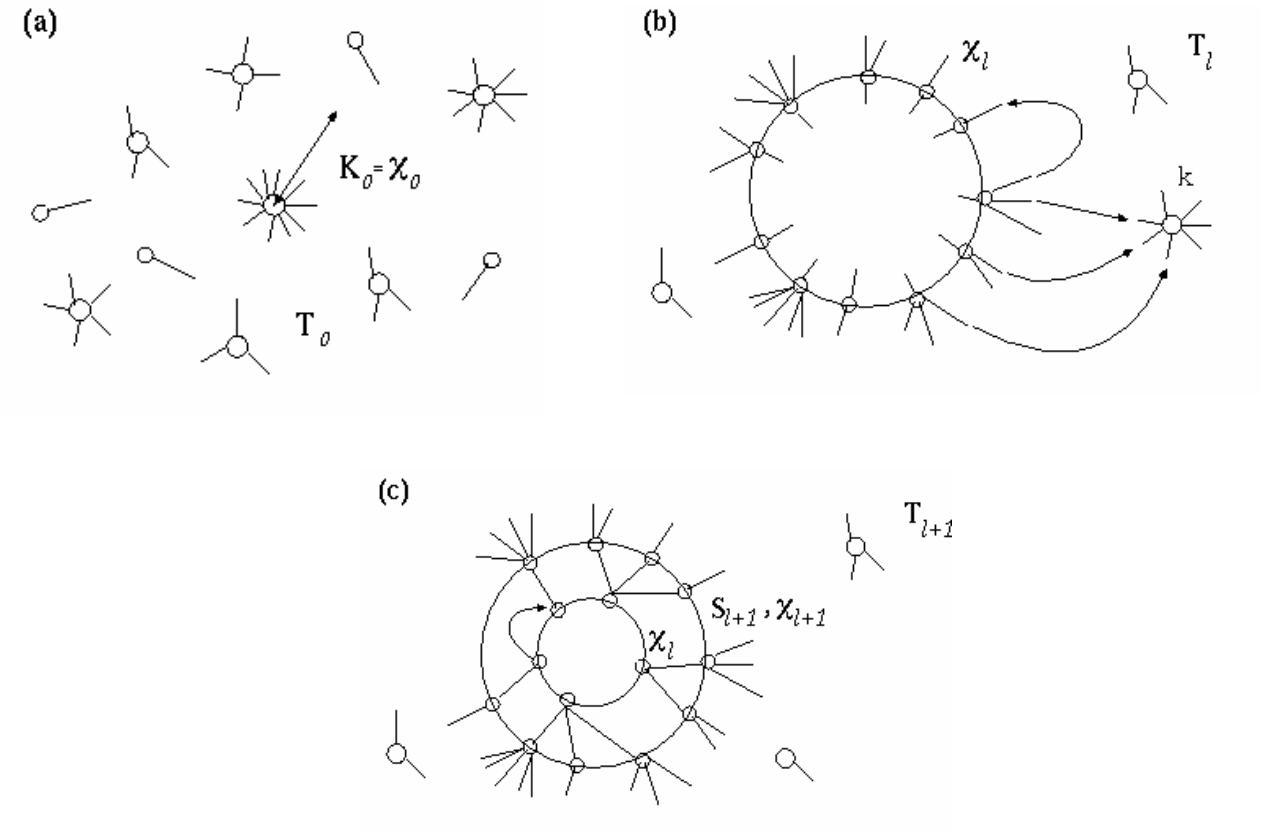



**Figure 3-1:** Illustration of the exposure process. The large circles denote exposed layers of the giant component, while the small circles denote individual sites. The sites outside the large circles have not been reached yet. (a) We begin with the highest degree node and fill out layer 1. (b) In the exposure of layer l+1 any open connection emerging from layer l may connect to any open node ($T_l$ connections) or loop back into layer l ($\chi_l$ connections). (c) The number of connections emerging from layer l+1 is the difference between $T_l$ and $T_{l+1}$, after reducing $S_{l+1}$, which is the number of incoming connections from layer l.

(3.1)

$$P\left(k,l+1\,|\,l\right)=\left[1-\frac{k}{\chi_l+T_l}\right]^{\chi_l}\approx\exp\left(-\frac{k}{1+T_l/\chi_l}\right)$$

for large enough values of $\chi_l$.

Thus, the probability that a node of degree k will be outside layer l+1 is:

$$P_{l+1}\left(k\right)=P_l\left(k\right)P\left(k,l+1\,|\,l\right)=P_l\left(k\right)\exp\left(-\frac{k}{1+T_l/\chi_l}\right)$$

Thus we derive the exponential cutoff:

$$P_l\left(k\right)=P\left(k\right)\exp\left(-\frac{k}{K_l}\right) \tag{3.2}$$

where:

$$\frac{1}{K_{l+1}}=\frac{1}{K_l}+\frac{1}{1+\frac{T_l}{\chi_l}}, \tag{3.3}$$

gives the evolution of the cutoff[17] with $l$.

---

[17] The exponential cutoff may be derived also using the following "mean field" approximation: Each node is treated independently, where the interaction between nodes is inserted through the expected number of incoming connections. At each node, the process is treated as equivalent to randomly distributing $\chi_l$ independent points on a line of length $\chi_l+T_l$ and counting the resultant number of points inside a small interval of length k. Thus, the number of incoming connections $k_{in}$ from layer l to a node with k open connections is distributed according to a Poisson distribution with an expected value of:

$$\left\langle k_{in}\right\rangle=\frac{k}{\chi_l+T_l}\chi_l\,,$$

and:

$$P_{l+1}\left(k_{in}\,|\,k\right)=e^{-\left\langle k_{in}\right\rangle}\frac{\left\langle k_{in}\right\rangle^{k_{in}}}{k_{in}!}$$

The probability for a node with k open connections *not* to be connected to layer l, i.e. to be outside layer l+1 also, is:

Now let us find the behavior of $\chi_l$ and $S_l$, where $S_{l+1}$ is the number of links incoming to the l+1 layer (and approximately[18] equals $N_{l+1}$, the number of nodes in the l+1 layer). The number of incoming connections to layer l+1 ($S_{l+1}$) equals the number of connections emerging from layer l ($\chi_l$), minus the number of connections looping back into layer l. The probability for a link to loop back into layer l is:

$$P(loop \mid l) = \frac{\chi_l}{\chi_l + T_l}$$

and therefore:

$$S_{l+1} = \chi_l \left(1 - \frac{\chi_l}{\chi_l + T_l}\right) \tag{3.4}$$

The number of connections emerging from all the nodes of layer l+1 is $T_l - T_{l+1}$. This is also the sum of the number of incoming connections from layer l into layer l+1, which is equal to $S_{l+1}$, and the number of outgoing connections $\chi_{l+1}$. Therefore:

$$\chi_{l+1} = T_l - T_{l+1} - S_{l+1} \tag{3.5}$$

---

$$P(k, l+1 \mid l) = P_{l+1}(k_{in} = 0 \mid k) = e^{-\langle k_{in} \rangle} = \exp\left(-\frac{k}{1 + T_l / \chi_l}\right)$$

Thus the total probability to find a node of degree k outside layer l+1 is:

$$P_{l+1}(k) = P_l(k) P(k, l+1 \mid l) = P_l(k) \exp\left(-\frac{k}{1 + T_l / \chi_l}\right)$$

and one obtains the exponential cutoff.

[18] This holds true assuming that almost no site in layer l+1 is reached by two connections from layer l. This is justified in the case where m=1 (where most of the nodes have only one incoming link), and also for the first layers in case of m>1 (in which loops are not yet formed).

At this point we have the following relations: $T_{l+1}(K_{l+1})$ (Equations (3.1) and (3.2)), $S_{l+1}(\chi_l,T_l)$ (3.4), $K_{l+1}(K_l,\chi_l,T_l)$ (Eq. (3.3)), and $\chi_{l+1}(T_l,T_{l+1},S_{l+1})$ (Eq. (3.5)). These relations may be solved numerically to give the degree distribution and number of nodes at each layer[19]. Approximate analytical results for the limit $N \to \infty$ can be found in [13].

## 3.4. Simulations

In order to check our analysis we simulated a SF network with $2 < \lambda < 3$, and used the Breadth-First-Search (BFS) algorithm [19] to extract the number of nodes at each layer starting from the maximal connected node. The BFS algorithm is implemented as follows (see Fig. 3-2): we start from a chosen node and insert it into a queue, assigning it a layer number l=0. We then extract it from the other side of the queue and take all its neighbors and insert them into the end of the queue, assigning them a layer number l=1. The algorithm continues in the same way: at each iteration we extract the first node (with layer number l) from the beginning of the queue and insert its neighbors to the end of the queue, assigning them a layer number l+1. However, each node that is inserted into the queue is marked so that it will not be inserted again. In this way the algorithm "explores" the graph layer by layer until there are no more nodes.

Figure 3-3 shows results from simulations (symbols) for the number of nodes on layer l, which can be seen to be in agreement with the analytical curves

---

[19] We begin with $K_0$=K (the natural cutoff of the network), $\chi_0 = K_0 = K$, and $P_0(k) = ck^{-\lambda}$.

of $S_l$ (lines). We can see that starting from a given layer l=L the number of nodes decays exponentially. We believe that the layer index L is related to the radius of the graph, which scales as $\log\log N$ for scale-free networks with $2<\lambda<3$ [13]. It can be seen that $S_l$ is a good approximation for the number of nodes at layer l. This is true in cases when only a small fraction of sites in each layer l have more than one incoming connection. An example for this case is when m=1 so that most of the sites in the network have only one connection. Figure 3-4 shows results for $P_l(k)$ with similar agreement. Note the exponential cutoff that becomes stronger with l (i.e. $K_l$ is a monotonically decreasing function of l).

It is important to note that the simulation results give the probability distribution for the giant component, while the analytical reconstruction gives the probability distribution for the whole graph. This may explain the difference in the probability distributions for lower degrees (and large l): many low degree nodes are not connected to the giant component and therefore the probability distribution derived from the simulation is smaller for low values of k.

In another study, Brunet *et al*. [30] analyzed the layer structure for the Barabasi-Albert model [17] and its randomized variations [31]. They found that in the Barabasi-Albert model there are more nodes in the first layers due to dissortative mixing [32], which means that in Barabasi-Albert model there is a tendency for high degree nodes to connect to high degree nodes.

## 3.5. Empirical results

Figures 3-5 and 3-6 show similar analysis for a "cut" (i.e., a large subnetwork) of the Internet at router level, taken from the Lucent mapping project [33]. The actual probability distribution is not a pure power law; rather it can be approximated by $\lambda = 2.3$ for small degrees and $\lambda = 3$ at the tail. Our analytical reconstruction of the layer statistics assumes $\lambda = 3$, because the tail of a power law distribution is the important factor in determining properties of the system. This method results in a good reconstruction for the number of nodes in each layer, and a qualitative reconstruction of the probability distribution in each layer. Similar behavior was found in real maps of multicast trees [22, 29].

In general, large degree nodes of the network mostly reside in the lower layers, while the layers further away from the source node are populated mostly by low degree nodes [22]. This implies that the tail of the distribution affects the lower layers, while the distribution function for lower degrees affects the outer layers. Thus the deviations in the analytical reconstruction of the number of nodes per layer for the higher layers may be attributed to the deviation in the assumed distribution function for low degrees (that is: $\lambda = 3$ instead of $\lambda = 2.3$).

Note that our model does not take into account the correlations in node degrees, which were observed in the Internet [30, 32], and hierarchical structures [34]. This may also explain the deviation of our measurements from the model predictions.

## 3.6. Summary and conclusions

In this Chapter we introduced a structural characterization for networks, which we termed "Tomography". The Tomography of a network reveals the statistical properties of layers inside it (in particular, the number of nodes in each layer and their degree distribution). We show that, when applied to real world networks, the Tomography reveals their non-random nature with respect to the Molloy-Reed model. In general, the Tomography can be used as a simple local test for evaluation of different modeling schemes for the Internet [29], and also for the design of better network algorithms that take advantage of the network structure, for example: in estimation of the expected number of clients in a multicast trees [22].

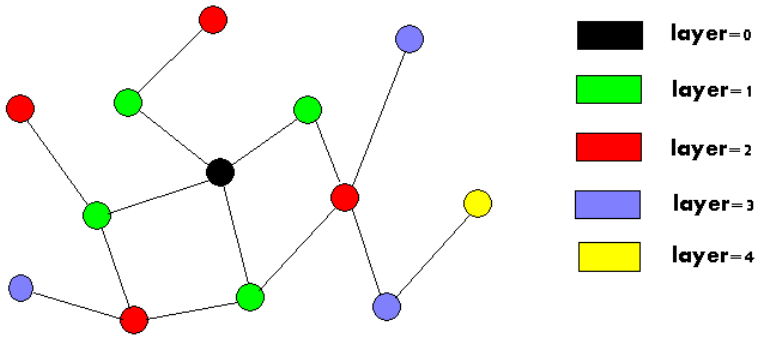



**Figure 3-2:** An illustration of extracting the layers in the network using the BFS algorithm. Starting from the maximal connected node, the graph is "exposed" layer by layer until there are no more nodes.

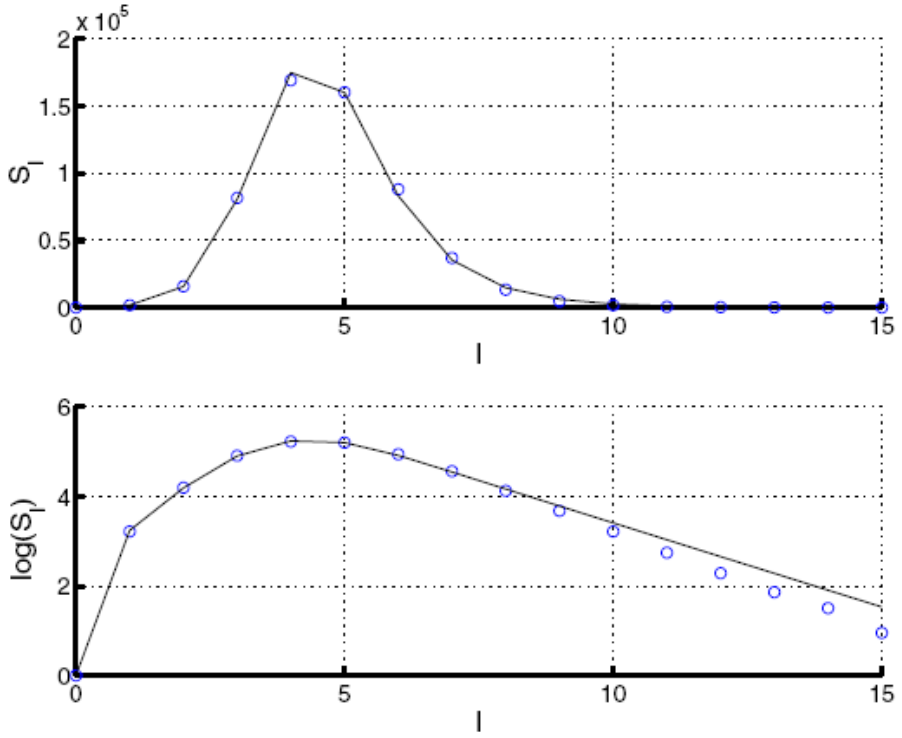



**Figure 3-3:** Approximate number of nodes ($S_l$) vs. layer index l for a network with $N=10^6$ nodes, $\lambda=2.85$, and m=1. Symbols represent simulation results while solid lines are a numerical solution for the derived recursive relations. Bottom: from the semi-log plot we see that there is an exponential decay of $S_l$ for layers $l>L$ starting from a given layer L, which we believe is related to the radius of the graph.

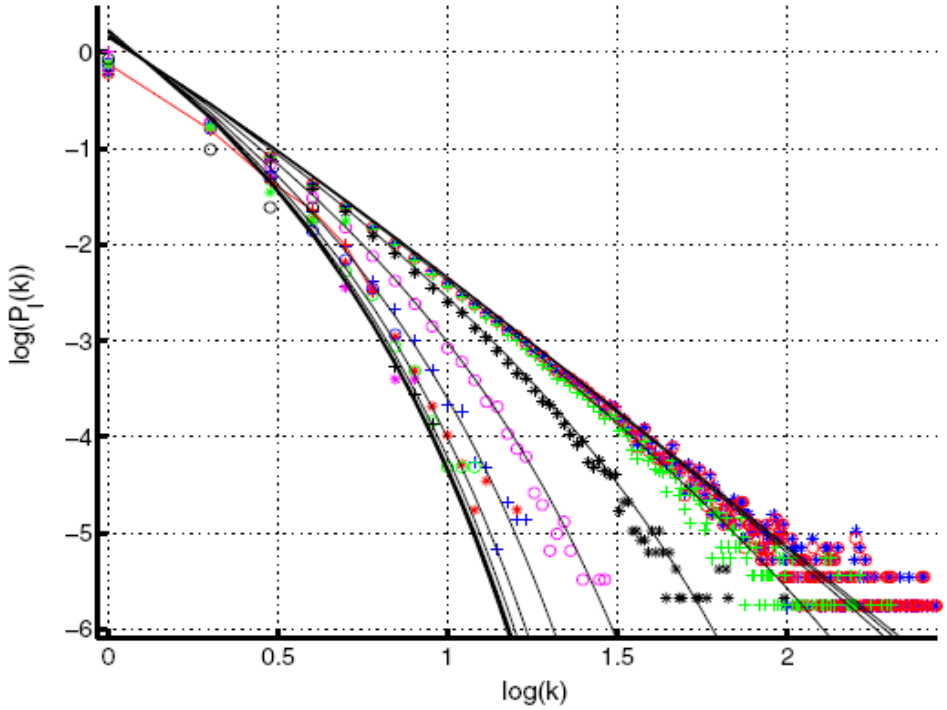



**Figure 3-4:** Log-log plot of $P_l(k)$ for different layers l=0,1,2,... (from top to bottom), for a network with $N=10^6$ nodes, $\lambda=2.85$, and m=1. Symbols represent simulation results while solid lines are a numerical solution of the derived recursive relations.

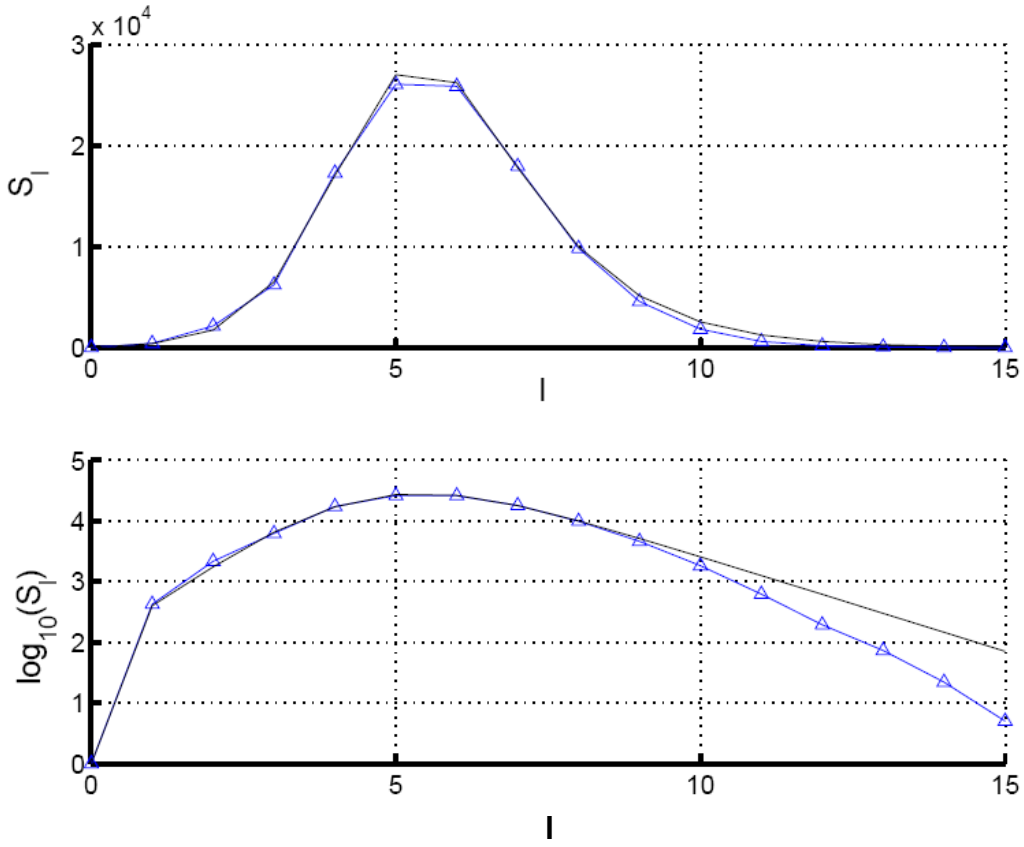



**Figure 3-5:** Real data results. Shown is the number of nodes at each layer for a router level cut of the Internet with N =112,969 nodes. Analytical reconstruction for $S_l$ was done with $\lambda = 3$, and $m = 1$.

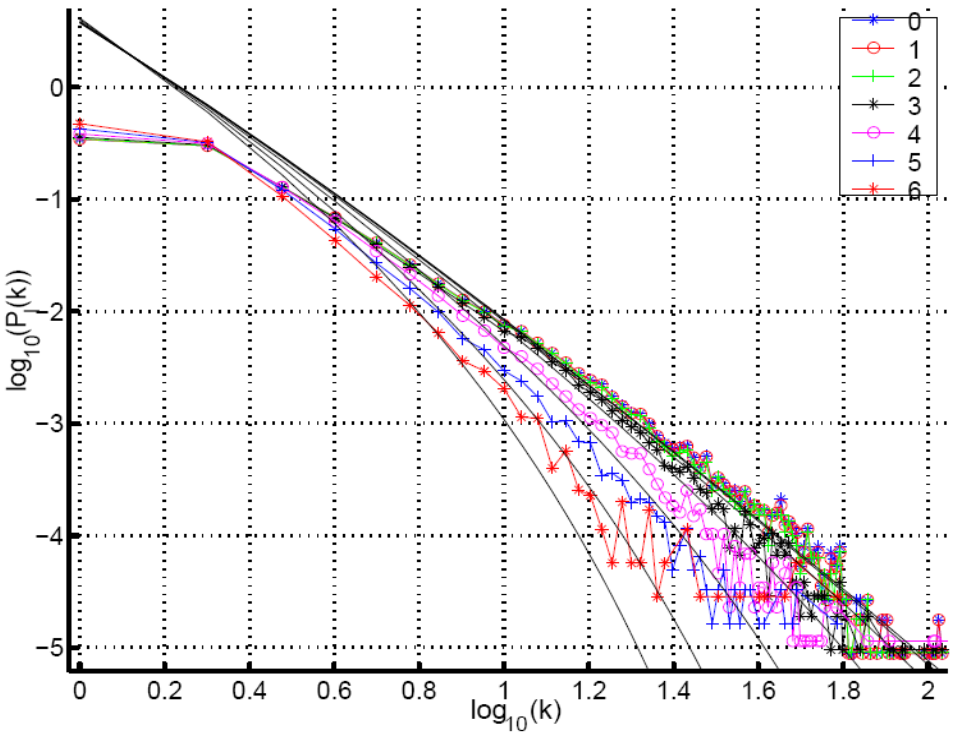



**Figure 3-6:** Real data results. Shown is a log-log plot of $P_l(k)$ for different layers $l = 0,1,2,...$ for a router level cut of the Internet with N =112,969 nodes. Qualitative analytical reconstruction was done with $\lambda = 3$, and $m = 1$. The exponential cutoff can be seen at the tail of the distribution.

# Chapter 4: Width of percolation threshold in complex networks

It is known that the critical probability for the percolation transition is not a sharp threshold; actually it is a region of non-zero width $\Delta p_c$ for systems of finite size. In this Chapter we will show that for complex networks $\Delta p_c \sim p_c / l$, where $l \sim N^{\nu_{opt}}$ is the average (chemical) length of the percolation cluster, and N is the number of nodes in the network. For Erdös-Réyni (ER) graphs $\nu_{opt} = 1/3$, while for scale-free (SF) networks with a degree distribution $P(k) \sim k^{-\lambda}$ and $3 < \lambda < 4$, $\nu_{opt} = (\lambda - 3)/(\lambda - 1)$. We show analytically and numerically that the survivability $S(p,l)$, which is the probability of a cluster to survive l chemical shells at conduction probability p, behaves near criticality as $S(p,l) = S(p_c,l) \cdot \exp\left[\frac{1}{p_c}(p - p_c)l\right]$. Thus for probabilities inside the region $|p - p_c| < \frac{p_c}{l}$ the behavior of the system is indistinguishable from that of the critical point. The results presented here are based on [35].

## 4.1. Introduction

In this Chapter we will study the behavior of networks near the percolation threshold. The problem of percolation on networks has been studied extensively (e.g. [5]). Using percolation theory we can describe the resilience of the network

to breakdown of sites or links [12, 36], epidemic spreading [5, 37, 38], and properties of optimal paths [18].

A typical percolation system consists of a d-dimensional grid of length L, in which the nodes or links are removed with some probability $1-p$, or are considered “conducting" with probability $p$ (e.g. [15, 16]). Below some critical probability $p_c$ the system becomes disconnected into small clusters, i.e., it becomes impossible to cross from one side of the grid to the other by following the conducting links. Percolation is considered a geometrical phase transition exhibiting universality, critical exponents, upper critical dimension at $d_c$=6 etc. It was noted by Coniglio [39] that for systems of finite size L the transition from connected to disconnected state has a "width" $\Delta p_c \sim \frac{1}{L^{1/\nu}}$ where $\nu$ is the critical exponent related to the correlation length[20].

Percolation on networks has been studied also from a mathematical viewpoint [4-6]. It was found that in Erdös-Réyni (ER) graphs with an average degree $\langle k \rangle$ the percolation threshold is: $p_c = \frac{1}{\langle k \rangle}$. Below $p_c$ the graph is

---

[20] To see this, consider a percolation system in a d-dimensional lattice. At the percolation threshold $p_c$ the spanning cluster spans the whole system ($\xi \to \infty$). Near the percolation threshold a typical cluster spans a length of $\xi \sim \left| p - p_c \right|^{-\nu}$, which is the "correlation length". If the system has a finite geometrical length L, then for probabilities close enough to $p_c$ such that $L < \xi \sim \left| p - p_c \right|^{-\nu}$, the typical cluster size $\xi$ is larger than the system size, and thus in effect there is a cluster spanning the system. Thus, for probabilities inside the range $\left| p - p_c \right| < \Delta p_c \sim L^{-1/\nu}$ the behavior of the system is indistinguishable from its behavior at the critical point.

composed of small clusters (most of them trees). As $p$ approaches $p_c$ trees of increasing order appear. At $p = p_c$ a giant component emerges and loops of all orders abruptly appear. However, for graphs of finite size N the percolation threshold has a finite width $\Delta p_c \sim \frac{1}{N^{1/3}}$ [4], meaning that all attributes of criticality are present in the range $p \in \left[ p_c - \Delta p_c, p_c + \Delta p_c \right]$. For example: The number of loops is negligible below[21] $p_c + \Delta p_c$.

In this Chapter we study the Survivability of the network near the critical threshold. The survivability $S(p,l)$ is defined to be the probability of a connected cluster to "survive" up to l chemical shells in a system with conductance probability p [40] (i.e the probability that there exists at least one node at chemical distance l from a randomly chosen node on the same cluster). At the critical point $p_c$, the survivability decays as a power-law: $S(p_c,l) \sim l^{-x}$, where $x$ is a universal exponent[22].

Below $p_c$ the survivability decays exponentially to zero, while above $p_c$ it decays (exponentially) to a constant. Here we will derive analytically and numerically the functional form of the survivability above and below the critical

---

[21] O. Riordan and P. L. Krapivsky (private communication).

[22] For directed percolation we have $x = \beta / \nu$ [41]. This conforms with our results for networks, as will be derived in this Chapter, because $\beta = 1$ for ER graphs and $\beta = \frac{1}{\lambda - 3}$ for SF networks (with $3 < \lambda < 4$), while $\nu_l = 1$ for both cases (see Table 1-1).

point. We will show that near the critical point $S(p,l) = S(p_c,l) \cdot \exp\left[\frac{1}{p_c}(p-p_c)l\right]$. Thus, given a system that has a maximal chemical length l at the percolation threshold, for probabilities inside the range $|p-p_c| < \frac{p_c}{l}$ the behavior of the system is indistinguishable from that of the critical point. Hence we get $\Delta p_c < \frac{p_c}{l}$.

The maximal chemical length l at the critical threshold, i.e. the length of the percolation cluster, was found to be: $l \sim N^{\nu_{opt}}$ [18] where N is the number of nodes in the network (see Chapter 2). For Erdös-Réyni (ER) graphs $\nu_{opt} = \frac{1}{3}$, while for scale-free (SF) networks with a degree distribution $P(k) \sim k^{-\lambda}$ and $3 < \lambda < 4$, $\nu_{opt} = (\lambda - 3)/(\lambda - 1)$.

## 4.2. General formalism

Consider a random graph with a degree distribution P(k), i.e., a randomly chosen node has a probability P(k) to have k links. The probability to reach a node of degree k by following a randomly chosen link is $P_1(k) = \frac{1}{\langle k \rangle} kP(k)$ [27] where $\langle k \rangle$ is the average degree. Accordingly, we write the two corresponding probability-generating functions (e.g. [27]):

$$G_0(x) = \sum_{k=0}^{\infty} P(k) x^k$$

And:

$$G_1(x)=\frac{G_0'(x)}{G_0'(1)}=\frac{1}{\langle k\rangle}\sum_{k=1}^{\infty}kP(k)\cdot x^{k-1}=\sum_{k=1}^{\infty}P_1(k)\cdot x^{k-1}$$

Where $G_1(x)$ describes the probability that a node reached by following a random link has k outgoing links, not including the incoming link. For example, in ER graphs: $G_1(x)=G_0(x)=e^{\langle k\rangle(x-1)}$.

After randomly removing a fraction $1-p$ of the links (bond percolation), the probability for a randomly chosen node to have k remaining links in the diluted graph is given by [12]:

$$\tilde{P}(k)=\sum_{k_0=k}^{\infty}P(k_0)\binom{k_0}{k}p^k(1-p)^{k_0-k}$$

The corresponding probability generating functions $\tilde{G}_0(x)=\sum_{k=0}^{\infty}\tilde{P}(k)\cdot x^k$ and $\tilde{G}_1(x)=\sum_{k=1}^{\infty}\tilde{P}_1(k)\cdot x^{k-1}$ in the diluted graph are:

$$\begin{aligned}\tilde{G}_0(x)&=\sum_{k=0}^{\infty}\left[\sum_{k_0=k}^{\infty}P(k_0)\binom{k_0}{k}p^k(1-p)^{k_0-k}\right]\cdot x^k=\\&=\sum_{k_0=0}^{\infty}P(k_0)\sum_{k=0}^{k_0}\binom{k_0}{k}(xp)^k(1-p)^{k_0-k}=\\&=\sum_{k_0=0}^{\infty}P(k_0)(1-p+px)^{k_0}=G_0(1-p+px)\end{aligned}$$

And:

$$\tilde{G}_1(x)=\frac{\tilde{G}_0'(x)}{\tilde{G}_0'(1)}=\frac{pG_0'(1-p+px)}{pG_0'(1)}=G_1(1-p+px)$$

For example, in ER graphs, $\tilde{G}_1(x)=\tilde{G}_0(x)=e^{\langle k\rangle([1-p+px]-1)}=e^{\langle k\rangle p(x-1)}$.

We next define $M_l(x) = m_0 + m_1 x + m_2 x^2 + \ldots$ to be the generating function for the number of sites that exists on layer (i.e. chemical shell) l starting from a random node on the diluted graph, and $N_l(x) = n_0 + n_1 x + n_2 x^2 + \ldots$ to be the corresponding function for the number of sites that exists on layer l from a node *reached by following a random link.* In order to find $M_L(x)$ for some layer $L \gg 1$ we can write the following recursive relations [27, 42]:

$$N_1(x) = \tilde{G}_1(x)$$

For $1 \le l < L-1$:

$$N_{l+1}(x) = \tilde{G}_1\left(N_l(x)\right) \tag{4.1}$$

And similarly, for the final layer:

$$M_L(x) = \tilde{G}_0\left(N_{L-1}(x)\right) \tag{4.2}$$

Eq. (4.1) means that the probability $n_i^{(l+1)}$ for reaching a branch having i nodes at layer l+1 is composed of the probability of reaching a node by following a link, and then reaching i nodes at layer l by following all possible branches emerging from that node - see sketch in Fig. 4-1.

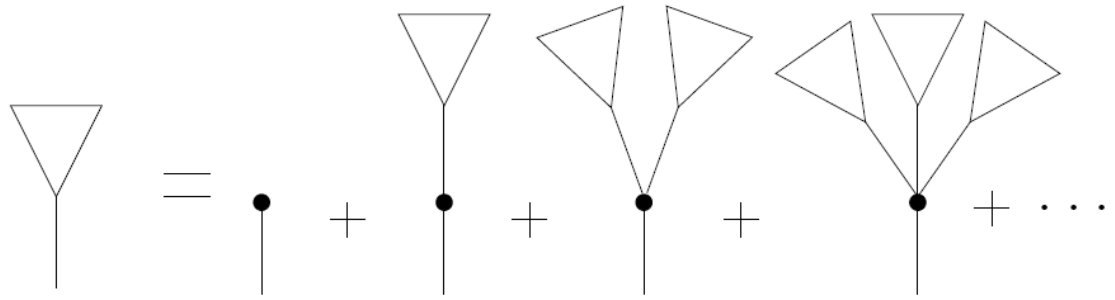


**Figure 4-1:** A graphical sketch of the recursion relation (4.1). The probability to reach a branch having i nodes at layer l+1 may be represented as the sum of probabilities to reach a single node, to reach a node connected to a single branch having i nodes at layer l, to reach a node connected to two branches having a total of i nodes at layer l, etc.

As a simple demonstration, let us evaluate the probability $n_0^{(l+1)}$ to encounter zero nodes at layer l+1 of a branch. Taking the zeroth power in Eq. (4.1) we have: $n_0^{(l+1)} = \tilde{P}_1(1) + \tilde{P}_1(2) \cdot n_0^{(l)} + \tilde{P}_1(3) \cdot \left[ n_0^{(l)} \right]^2 + \ldots$, which means that the probability to reach zero nodes at layer l+1 (by following a link) is composed of the probability $\tilde{P}_1(1)$ to reach a node with no emerging branch, the probability $\tilde{P}_1(2) \cdot n_0^{(l)}$ to reach a node that has a single emerging branch with zero nodes at layer l, the probability $\tilde{P}_1(3) \cdot \left[ n_0^{(l)} \right]^2$ to reach a node having two branches such that both of them have zero nodes at layer l etc. (see Fig. 4-1). Similarly, Eq. (4.2) refers to $M_L(x)$, which gives the probability for the number of nodes at layer L reached by starting from a random node, rather than by following a random link [27].

Notice that $M_L(0) = m_0$ is the probability that there are zero nodes at layer L from a random node, i.e., the probability to die before layer L. Thus $\varepsilon_L = 1 - M_L(0)$ is the probability to survive up to layer L. Similarly, $\varepsilon_l = 1 - N_l(0)$ (where $1 \le l < L-1$) is the probability for a *branch* to survive up to layer l. From Eq. (4.1) we have:

$$N_{l+1}(0) = \tilde{G}_1\left(N_l(0)\right)$$

$$1 - \varepsilon_{l+1} = \tilde{G}_1\left(1 - \varepsilon_l\right) = G_1\left(1 - p + p\left[1 - \varepsilon_l\right]\right)$$

Thus for $1 \le l < L-1$:

$$\varepsilon_{l+1} = 1 - G_1\left(1 - p\varepsilon_l\right) \tag{4.3}$$

And for the final layer L we have (Eq. (4.2)):

$$\varepsilon_L = 1 - G_0\left(1 - p\varepsilon_{L-1}\right), \tag{4.4}$$

which gives the survivability at layer L [42].

# 4.3. Erdös-Réyni graphs

For Erdös-Réyni (ER) graphs: $G_0(x) = G_1(x) = e^{\langle k \rangle (x-1)}$ and Eq. (4.3) gives:

$$\begin{aligned}
\varepsilon_{l+1} &= 1 - e^{\langle k \rangle \left([1 - p\varepsilon_l] - 1\right)} = 1 - e^{-p\langle k \rangle \varepsilon_l} = \\
&= 1 - \left[1 - p\langle k \rangle \varepsilon_l + \frac{p^2 \langle k \rangle^2}{2} \varepsilon_l^{\,2} - \ldots\right] = \\
&= \frac{p}{p_c}\varepsilon_l - \frac{p^2 \langle k \rangle^2}{2} \varepsilon_l^{\,2} + \ldots
\end{aligned}$$

Where $p_c = \frac{1}{\langle k \rangle}$. Setting $\delta = p - p_c$, we get:

$$\varepsilon_{l+1} = \frac{p_c + \delta}{p_c}\varepsilon_l - \left(p_c + \delta\right)^2 \frac{\langle k \rangle^2}{2}\varepsilon_l{}^2 + \ldots \approx$$
$$\approx \varepsilon_l + \frac{\delta}{p_c}\varepsilon_l - \frac{1}{2}\varepsilon_l{}^2$$

where we have left only terms of second order[23] in $\varepsilon_l, \delta$. We thus get:

$$\frac{d\varepsilon_l}{dl} \approx \varepsilon_{l+1} - \varepsilon_l = -\frac{1}{2}\varepsilon_l{}^2 + \frac{\delta}{p_c} \cdot \varepsilon_l .$$

At criticality, $\delta = 0$, and the solution to this equation is: $\varepsilon_l \sim l^{-1}$. The additional term suggests the following solution near criticality: $\varepsilon_l \sim l^{-1} \cdot \exp\left(\frac{1}{p_c}\delta l\right)$. Note that for ER graphs Equations (4.3) and (4.4) are the same, and thus the survivability $\varepsilon_L$ at the final iteration also has the same form: $\varepsilon_l \sim L^{-1} \cdot \exp\left(\frac{1}{p_c}\delta L\right)$. The above result can be written as:

$$S\left(p,l\right) = S\left(p_c,l\right) \cdot \exp\left(\frac{1}{p_c}\left(p - p_c\right)l\right) \qquad (4.5)$$

In order to check this result we numerically calculated the survivability $S\left(p,l\right)$ near $p_c$ according to the recursive relations (4.3) and (4.4)[24].

Fig. 4-2(a) shows the survivability $S\left(p,l\right)$ for different values of p. For $p = p_c$ the survivability decays as a power law, while above and below there is an exponential decay, either to zero (for $p < p_c$) or to a constant (for $p > p_c$). Fig. 4-

---

[23] We assume that $p < p_c$ and thus $\varepsilon_l \ll 1$ for large $l$.

[24] We start with $\varepsilon_0 = 1$, and use Eq. (4.3) also for $l = 0$.

2(b) shows that all curves of the survivability $S(p,l)$ from (a) can be rescaled such that they all collapse. Moreover, scaled survivabilities from all different graphs with different values of $\langle k \rangle$ (i.e., different values of $p_c$) also collapse on the same curve. However, equation (4.5) is true only below the percolation threshold where there is no giant component. Above the percolation threshold there is an exponential decay to a non-zero constant, and the generalized expression is:

$$S(p,l) = S(p_c,l) \cdot \exp\left( -\frac{1}{p_c} |p - p_c| l \right) + P_\infty \qquad (4.6)$$

where $P_\infty$ is the probability for a randomly chosen node to be inside the percolation cluster[25]. Indeed, setting $\varepsilon_{l+1} = \varepsilon_l$ in the recursive relation $\varepsilon_{l+1} = 1 - e^{-p\langle k \rangle \varepsilon_l}$, the resulting "steady state" solution is $\varepsilon_l = P_\infty$ [4, 6].

## 4.4. Scale-free graphs

Scale-free graphs can be taken to have a degree distribution of the form $P(k) = ck^{-\lambda}$, where $c \approx (\lambda - 1) m^{\lambda - 1}$ and m is the minimal degree [12]. In order to solve equation (4.3) we have to evaluate:

$$G_1(1 - p\varepsilon_l) = \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} kP(k)(1 - p\varepsilon_l)^{k-1}$$

---

[25] $S(p, l \to \infty)$ is the probability that if we start from a randomly chosen site, we will survive an infinite chemical distance. This equals to the probability $P_\infty$ that the chosen site resides in the giant component. In ER graphs $P_\infty$ obeys the transcendental equation: $P_\infty = 1 - e^{-\langle k \rangle p P_\infty}$ [4, 6].

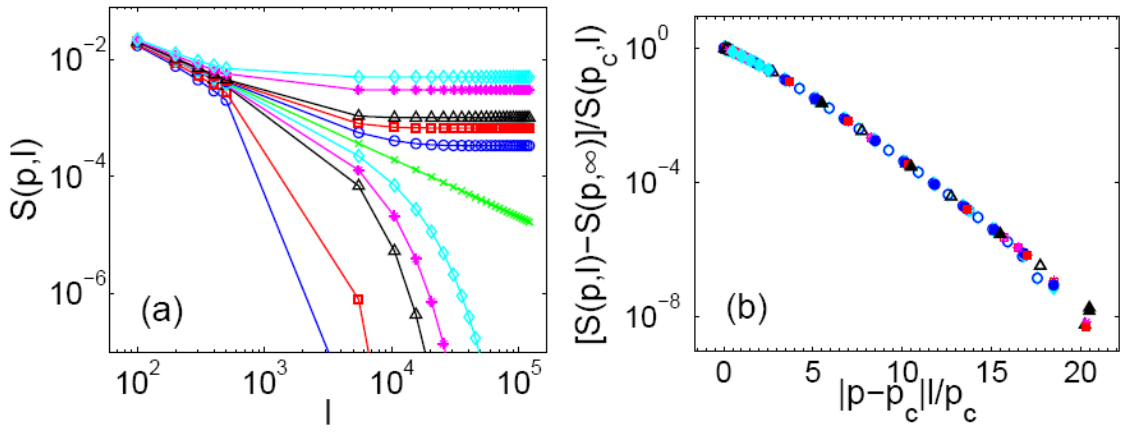



**Figure 4-2:** (a) The survivability $S(p,l)$ for an ER graph with $\langle k\rangle = 5$, numerically calculated for different values of $p$: $p_c$, $p_c \pm 5\times10^{-4}$, $p_c \pm 3\times10^{-4}$, $p_c \pm 1\times10^{-4}$, $p_c \pm 6.66\times10^{-5}$, and $p_c \pm 3.33\times10^{-5}$. For $p = p_c$ the survivability decays to zero according to a power law: $S(p_c,l) \sim l^{-1}$. For $p < p_c$, $S(p,l) \to 0$, while for $p > p_c$, $S(p,l) \to Const$. The decay is exponential (to zero or to a constant) according to equations (4.5) and (4.6). (b) Scaling of the survivability for different values of p, l, and $\langle k\rangle$. Shown is $\left[S(p,l) - S(p,\infty)\right]/S(p_c,l)$ vs. $|p-p_c|l/p_c$ for ER graphs with $\langle k\rangle = 5$ (unfilled symbols) and $\langle k\rangle = 10$ (filled symbols). The collapse of all curves on an exponential function (for large l) supports the scaling relations (4.5) and (4.6).

Expanding by powers of $\varepsilon$, and inserting $P(k) = ck^{-\lambda}$ with $3 < \lambda < 4$, we get [43]

(See also appendix A):

$$\sum_{k=1}^{\infty} kP(k)(1-\varepsilon)^{k-1} = \langle k\rangle - \langle k(k-1)\rangle\varepsilon + \frac{c}{2}\Gamma(4-\lambda)\varepsilon^{\lambda-2}$$

Thus equation (4.3) becomes:

$$\varepsilon_{l+1} = 1 - \frac{1}{\langle k\rangle}\left[\langle k\rangle - \langle k(k-1)\rangle p\varepsilon_l + \frac{c}{2}\Gamma(4-\lambda)\cdot(p\varepsilon_l)^{\lambda-2}\right] =$$

$$= \frac{p}{p_c}\varepsilon_l - \frac{c}{2\langle k\rangle}\Gamma(4-\lambda)\cdot p^{\lambda-2}\varepsilon_l^{\lambda-2}$$

where $p_c = \frac{\langle k \rangle}{\langle k(k-1) \rangle}$ [12]. Taking $p = p_c + \delta$, and substituting $A = \frac{c}{2\langle k \rangle} \Gamma(4-\lambda)\, p_c^{\lambda-2}$, we get:

$$\varepsilon_{l+1} = \frac{p_c + \delta}{p_c} \varepsilon_l - \frac{c}{2\langle k \rangle} \Gamma(4-\lambda) \cdot (p_c + \delta)^{\lambda-2} \varepsilon_l^{\lambda-2} =$$
$$= \varepsilon_l + \frac{\delta}{p_c} \cdot \varepsilon_l - A\left[1 + \frac{\delta}{p_c}\right]^{\lambda-2} \varepsilon_l^{\lambda-2} \approx$$
$$\approx \varepsilon_l - A\varepsilon_l^{\lambda-2} + \frac{\delta}{p_c}\left[\varepsilon_l - A(\lambda-2) \cdot \varepsilon_l^{\lambda-2}\right]$$

For large $l$, $\varepsilon_l \ll 1$. Taking into account that $\lambda - 2 > 1$ we have $\varepsilon_l^{\lambda-2} \ll \varepsilon_l$. Therefore:

$$\frac{d\varepsilon_l}{dl} \approx \varepsilon_{l+1} - \varepsilon_l = -A\varepsilon_l^{\lambda-2} + \frac{\delta}{p_c} \cdot \varepsilon_l$$

For $\delta = 0$ the solution is $\varepsilon_l \sim l^{-x}$ with $x = 1/(\lambda - 3)$. The additional term suggests the following solution near criticality: $\varepsilon_l \sim l^{-x} \cdot \exp\left(\frac{1}{p_c}\delta l\right)$. The last iteration (Eq. (4.4)) can be shown to give the same behavior for $\varepsilon_L$. A similar form can be found also for[26] $\lambda > 4$. The scaling form for SF networks is confirmed by numerical simulations as shown in Figures 4-3(a) and (b).

[26] In this range the behavior is similar to ER graphs [14].

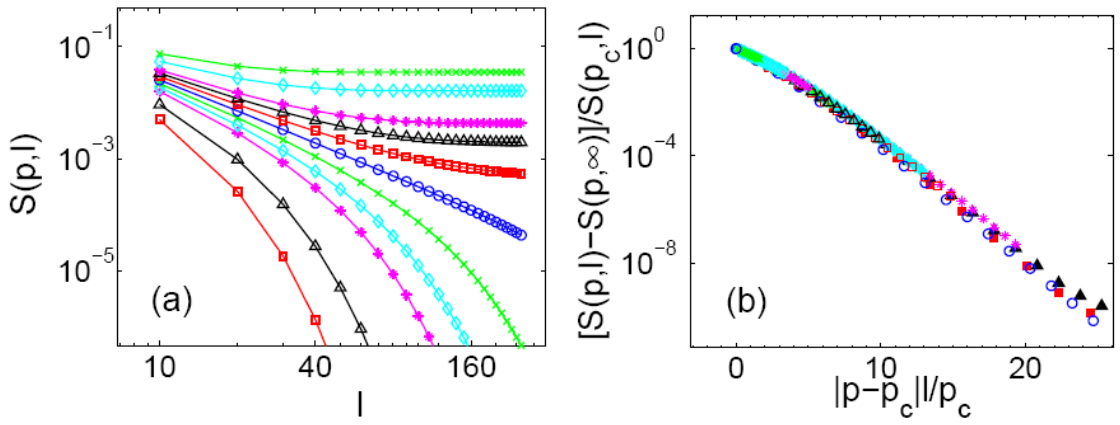



**Figure 4-3:** (a) The survivability $S(p,l)$ for a SF network with $\lambda = 3.5$, numerically calculated for different values of $p$: $p_c$, $p_c \pm 6\times10^{-2}$, $p_c \pm 4\times10^{-2}$, $p_c \pm 2\times10^{-2}$, $p_c \pm 1.33\times10^{-2}$, and $p_c \pm 6.66\times10^{-3}$. For $p = p_c$ the survivability decays to zero according to a power law: $S(p_c,l) \sim l^{-2}$. For $p \neq p_c$, $S(p,l)$ decays exponentially (to zero or to a constant) according to equations (4.5) and (4.6). (b) Scaling of the survivability for different values of p, l, and $\lambda$. Shown is $\left[S(p,l)-S(p,\infty)\right]/S(p_c,l)$ vs. $\left|p-p_c\right|l/p_c$ for SF graphs with $\lambda = 3.5$ (filled symbols) and $\lambda = 5$ (unfilled symbols). For all cases m=2. Due to numerical difficulties only curves with $p < p_c$ are shown.

## 4.5. Summary and conclusions

We have shown analytically and numerically that the survivability in ER and SF graphs scales according to equations (4.5) and (4.6) near the critical point. Thus, the scaling form of the survivability near the critical probability obeys the following scaling relation (for $p < p_c$):

$$S(p,l) = S(p_c,l)\cdot\exp\left(\frac{p-p_c}{\Delta p_c}\right)$$

where $\Delta p_c \sim \frac{p_c}{l}$. Given a system with a maximal chemical length $l$ at criticality, for all values of conductivity p inside the range $\left[p_c - \Delta p_c, p_c + \Delta p_c\right]$ the survivability behaves similar to the power law $S\left(p_c, l\right) \sim l^{-x}$ found at $p = p_c$. Thus, the width of the critical threshold is $\Delta p_c \sim \frac{p_c}{l}$, where l is the chemical length of the percolation cluster. For ER graphs, $l \sim N^{1/3}$, while for SF networks with $3 < \lambda < 4$, $l \sim N^{(\lambda-3)/(\lambda-1)}$.

It is important to remember that real world networks do not have an infinite number of nodes. Some networks are rather small, for example, the transcription regulation network of the bacteria *E. coli* has roughly 400 nodes [44]. The importance of the results presented in this Chapter is that they allow us to use percolation theory also in real world systems of finite size. Furthermore, the scaling form of the survivability can be used in the modeling of epidemic spreading on networks [38].

## 4.6. Appendix 4.A: Derivation of $G_1\left(1-\varepsilon\right)$

We wish to evaluate $G_1\left(1-\varepsilon\right) = \frac{1}{\langle k \rangle}\sum_{k=1}^{\infty} kP\left(k\right)\left(1-\varepsilon\right)^{k-1}$ for a degree distribution $P\left(k\right) = ck^{-\lambda}$, where $c \approx \left(\lambda - 1\right)m^{\lambda-1}$ and $\varepsilon$ is small. We expand the function: $F(\varepsilon) = \sum_{k=1}^{\infty} kP(k)(1-\varepsilon)^{k-1}$ by series according to $\varepsilon$. For a function with $n-1$ derivatives at $\varepsilon = 0$, the Taylor expansion is:

$$F(\varepsilon) = F(0) + F'(0)\varepsilon + \frac{1}{2}F''(0)\varepsilon^2 + ... + \frac{1}{(n-1)!}F^{(n-1)}(0)\varepsilon^{n-1} + R_n$$

Now:

$$F(0) = \sum_{k=1}^{\infty} kP(k) = \langle k \rangle$$

$$F'(0) = -\sum_{k=2}^{\infty} k(k-1)P(k) = -\langle k(k-1) \rangle$$

$$F''(0) = \sum_{k=3}^{\infty} k(k-1)(k-2)P(k) = \langle k(k-1)(k-2) \rangle$$

...

The remainder of the series expansion is: $R_n = \frac{\varepsilon^n}{n!}F^{(n)}(\tilde{\varepsilon})$, where: $0 < \tilde{\varepsilon} < \varepsilon$. In our case we take $\tilde{\varepsilon} \approx \varepsilon$ because $\varepsilon$ is small. In the case of a power-law (scale-free) distribution $P(k) \sim k^{-\lambda}$ with $3 \le \lambda \le 4$, the 2nd derivative diverges. Thus we can expand only up to order n=2. The remainder is:

$$R_2 = \frac{\varepsilon^2}{2}F^{(2)}(\varepsilon) = \frac{\varepsilon^2}{2}\int_{k=0}^{\infty} k(k-1)(k-2)ck^{-\lambda}(1-\varepsilon)^{k-3}dk \approx \frac{\varepsilon^2}{2}c\int_{k=0}^{\infty} k^{3-\lambda}e^{-\varepsilon(k-3)}dk\,.$$

Taking $x = \varepsilon k$ gives:

$$R_2 = \frac{\varepsilon^2}{2}ce^{3\varepsilon}\int_{k=0}^{\infty} k^{3-\lambda}e^{-\varepsilon k}dk = \frac{\varepsilon^2}{2}ce^{3\varepsilon}\int_{x=0}^{\infty}\frac{x^{3-\lambda}}{\varepsilon^{3-\lambda}}e^{-x}\frac{dx}{\varepsilon} = \frac{\varepsilon^2}{2}ce^{3\varepsilon}\frac{1}{\varepsilon^{4-\lambda}}\int_{x=0}^{\infty} x^{3-\lambda}e^{-x}dx \approx$$

$$\approx \frac{c}{2}\varepsilon^{\lambda-2}\Gamma(4-\lambda)$$

Thus we get:

$$\sum_{k=1}^{\infty} kP(k)(1-\varepsilon)^{k-1} = \langle k \rangle - \langle k(k-1) \rangle \varepsilon + \frac{c}{2}\Gamma(4-\lambda)\varepsilon^{\lambda-2},$$

as required.

# Chapter 5: Optimization in weighted networks

Optimization problems in networks are important for many applications in computer science. For example, when constructing a communications network or implementing a routing algorithm, one wants the system to be optimal in the sense that the data packets will arrive with minimal delay time and with minimal waste of network resources. In this Chapter we will show how the problem of optimization can be approached using tools from percolation theory and statistical physics. The results presented here are based on [35, 45-47].

## 5.1. Introduction

Many real world systems exhibit a web like structure and may be treated as “networks.” Examples may be found in physics, sociology, biology, and engineering [1-3]. The function of most real world networks is to connect distant nodes, either by transfer of information, e.g., the Internet, or through transportation of people and goods such as networks of roads and airlines. In many cases there is a “cost” or a “weight” associated with each link and the larger the weight on a link, the harder it is to traverse this link. In this case, the network is called "disordered" or "weighted" [18]. For example, in the Internet, each link between two routers has a bandwidth or delay time; in a transportation networks, some roads may have only one lane while others may be highways allowing for large volumes of traffic. Another important example is in biological

networks such as the transcription regulation network of bacteria, where the weights represent chemical binding affinities between proteins and the DNA. These weights are fine-tuned by evolution in order to adapt to the environment [48].

Consider two nodes A and B on such a disordered network. In general, there will be a large number of paths connecting A and B. Among these paths, there is usually a single path for which the sum of the weights along the path is minimal. This path is called the "*optimal path*", and all optimal paths emerging from a certain node and reaching all nodes of the network create a *shortest-path-tree* (SPT) from that node. A standard algorithm for finding the optimal path (and the SPT) is *Dijkstra's* algorithm [19]. The optimal path may be much longer than the *shortest-hopcount-path* [23], i.e. the path with the minimum number of links between nodes A and B, as demonstrated if Fig. 5-1.

The problem of finding the optimal path is one example for an "optimization" problem. Another example is the *minimum-spanning-tree* (MST), which is a tree that reaches all nodes of the graph and for which the sum of the weights of all the links (total weight) is minimal. In this Chapter we will explain the behavior of the optimal path, the SPT and the MST and their dependence on the nature of the disorder.

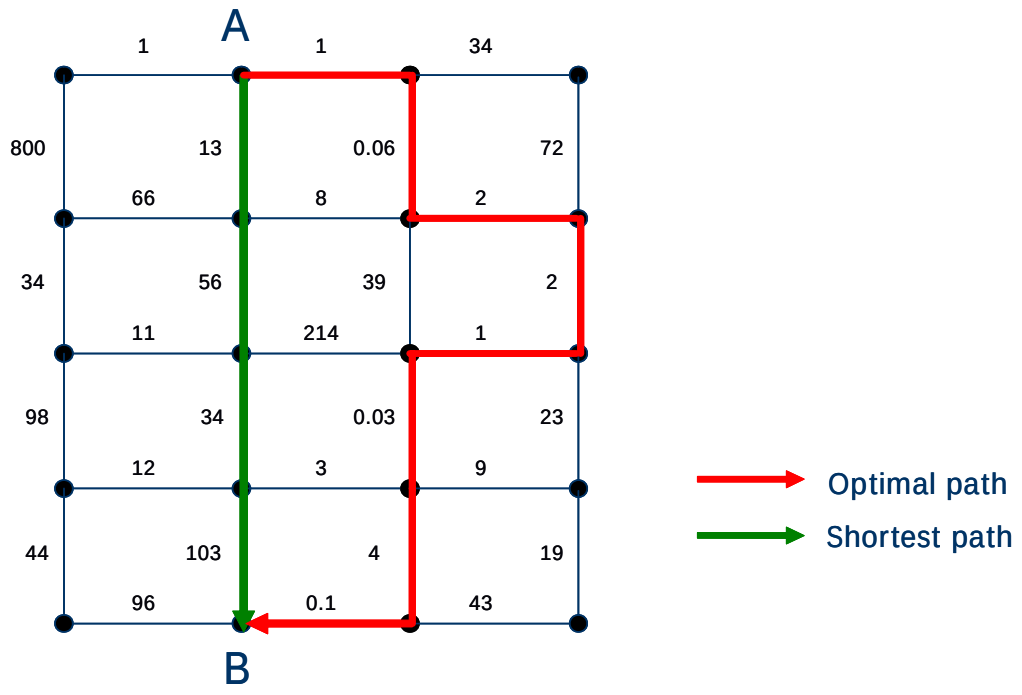



**Figure 5-1:** A sketch of a two-dimensional weighted network. In order to cross from node A to B along the *optimal path* we have to take a long detour, whereas the *shortest-hopcount-path* is a straight line.

## 5.2. Scale-free substructures emerging from weighted networks

In this subsection we will show that any weighted random graph has an internal scale-free structure. We will then show that the minimum spanning tree (MST) is related to this network, and is composed of percolation clusters, which we regard as “super-nodes”, interconnected by a scale-free tree. We will then use these results to understand the structure of optimal paths in disordered networks.

### 5.2.1. Scale-free topology

Scale-free topology is very common in natural and man-made networks. Examples vary from social contacts between humans to technological networks

such as the World Wide Web or the Internet [1-3]. Scale free (SF) networks are characterized by a power law distribution of connectivities, i.e. the degree distribution is $P(k) \sim k^{-\lambda}$, where k is the degree of a node and the exponent $\lambda$ controls the broadness of the distribution. Many networks are observed to have values of $\lambda$ around 2.5. For values of $2<\lambda<3$ the second moment of the distribution, $\langle k^2 \rangle$, diverges, leading to several anomalous properties [12, 13], for example, the percolation threshold in such networks vanishes ( $p_c \to 0$ ), and the radius (i.e., the average of the shortest-hopcount-path) scales as $r \sim \log\log N$, a phenomenon also known as "ultra-small world".

### 5.2.2. The clusters network

Consider an Erdös-Réyni (ER) graph with N nodes and an average degree $\langle k \rangle$, thus having a total of $N\langle k \rangle / 2$ links. To each link we assign a weight chosen randomly and uniformly from the range $[0,1]$. We define "black" links to be those links with weights below the percolation threshold $p_c = 1/\langle k \rangle$ [4]. Two nodes belong to the same cluster if they are connected by black links (Fig. 5-2(a)). From percolation theory [15, 16] follows that the number of clusters containing s nodes scales as a power law, $n_s \sim s^{-\tau}$, with $\tau = 2.5$ for ER networks[27]. We next merge

[27] ER networks can be regarded as having an infinite dimension since space does not play any role. For example, in a d-dimensional rectangular grid, each node may have $2^d$ neighbors, whereas in random graphs the number of possible neighbors of each node diverges with network size.

all nodes inside each cluster into a single “supernode”. We define a new “supernode network” (Fig. 5-2(b)) of $N_{SN}$ supernodes[28]. The links between two supernodes (see Figs. 5-2(a) and 5-2(b)) have weights larger than $p_c$.

The supernode network is scale-free with a degree distribution $P(k) \sim k^{-2.5}$. This can be explained as follows: every node in a supernode has the same (finite) probability to be connected to a node outside the supernode. Thus, the degree k of each supernode is proportional to the cluster size s, which obeys $n_s \sim s^{-\tau}$. Hence, $P(k) \sim k^{-\lambda}$ with $\lambda$=2.5, as supported by simulations shown in Fig. 5-3.

Next, we check if the supernode network is a random SF network (thus having anomalous properties such as ultra-small world etc.). We link-randomize the supernode network, in a way that preserves the degree distribution [31], This is done by the following algorithm: choose randomly two links A→B and C→D, and switch the endpoints, that is, disconnect those two links and connect A→D and C→B (Fig. 5-4(b)). We observe (Fig. 5-4(a)) that the "tomography" of the resulting randomized network, i.e. the number of nodes at each chemical shell from the maximal degree node (see Chapter 3), is the same as in the original supernode network, thus indicating that the supernode network is indeed random.

---

[28] $N_{SN} = N/2$ in Erdos-Renyi networks, as will be shown below.

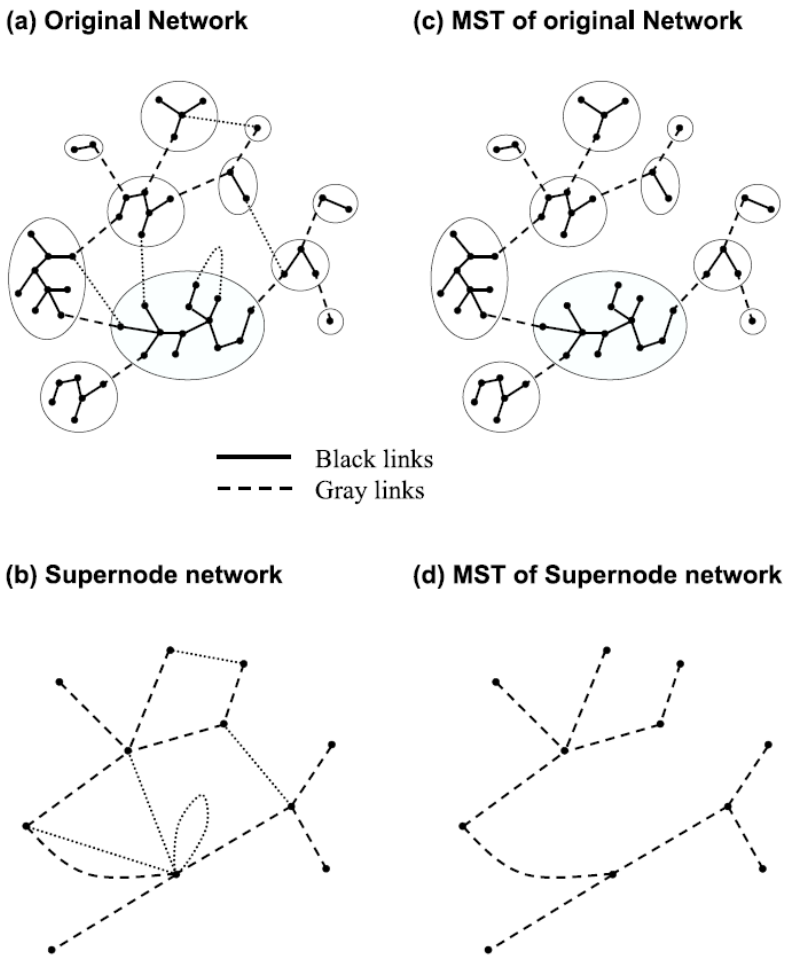



**Figure 5-2:** Sketch of the "supernode network". (a) The original ER network, partitioned into percolation clusters whose sizes s are power-law distributed, with $n_s \sim s^{-\tau}$, where $\tau = 2.5$ for ER graphs. The "black" links are the links with weights below $p_c$, the "dotted" links are the links that are removed by the bombing algorithm, and the "gray" links are the links whose removal will disconnect the network (and therefore are not removed even though their weight is above $p_c$). (b) The "supernode network": the nodes are the clusters in the original network and the links are the links connecting nodes in different clusters (i.e., "dotted" and "gray" links). The supernode network is scale-free with $P(k) \sim k^{-\lambda}$ and $\lambda = 2.5$. Notice the existence of self loops and double connections between the same two supernodes. (c) The minimum spanning tree (MST), which is composed of black and gray links only. (d) The MST of the supernode network ("gray tree"), which is obtained by bombing the supernode network (thereby removing the "dotted" links), or equivalently, by merging the clusters in the MST to supernodes. The gray tree is also scale-free, with $\lambda = 2.5$.

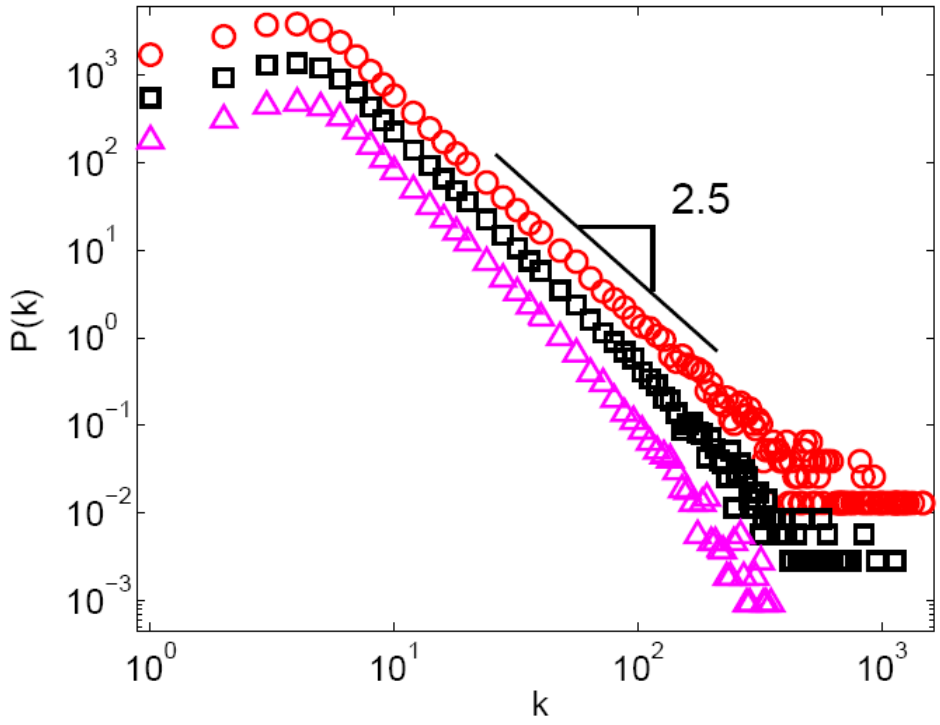



**Figure 5-3:** The degree distribution of the supernode network of Fig. 5-2(b), where the supernodes are the percolation clusters, and the links are the links with weights larger than $p_c$ (O). The distribution exhibits a scale-free tail with $\lambda = 2.5$. If we choose a threshold less than $p_c$, we obtain the same power law degree distribution with an exponential cutoff. The different symbols represent slightly different threshold values: $p_c - 0.03$ (□) and $p_c - 0.05$ (Δ). The original ER network has N=50,000 nodes and an average degree $\langle k \rangle = 5$. Note that for $k \approx \langle k \rangle$ the degree distribution has a maximum.

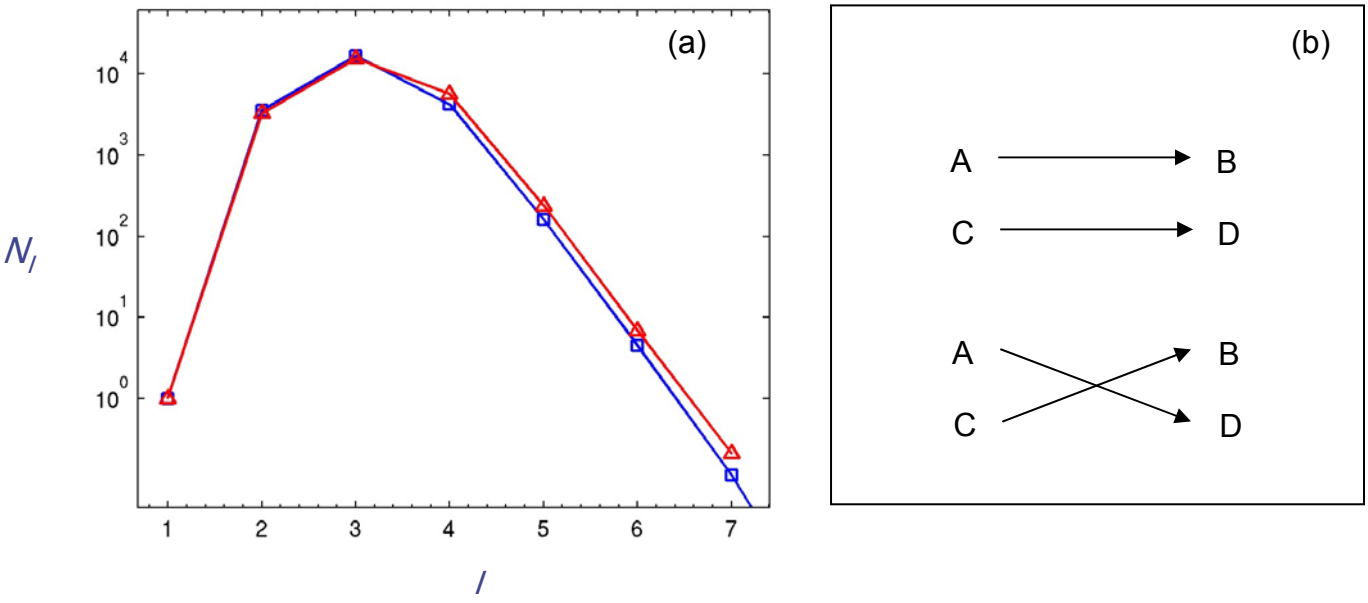



**Figure 5-4:** (a) The "tomography" of the supernode network (□) and the link-randomized supernode network (Δ). It can be seen that the number of nodes at each layer is the same, thus indicating that the supernode network is random. The tomography here was obtained by averaging over many realizations of the network. Hence, for large $l$, we get a fractional number of nodes on average. (b) An illustration of the link-randomizing algorithm. Note that the number of links connected to each one of the nodes A,B,C, and D does not change in the switching process. Thus the degree of each node is preserved.

### 5.2.3. The minimum spanning tree – definition and algorithms

We next show that the minimum spanning tree (MST) of an ER graph is related to the supernode network, and therefore also exhibits scale-free properties. The MST on a weighted graph is a tree that reaches all nodes of the graph and for which the sum of the weights of all the links (total weight) is minimal [19].

Standard algorithms [19] for finding the MST are Prim's algorithm which resembles invasion percolation [15], and Kruskal's algorithm which resembles normal percolation. We first explain ***Prim's algorithm:***

(a) Create a tree containing a single vertex, chosen arbitrarily from the graph.

(b) Create a set containing all the edges in the graph.

(c) Remove from the set an edge with minimum weight that connects a vertex in the tree with a vertex not in the tree.

(d) Add that edge to the tree.

(e) Repeat steps (c-d) until every edge in the set connects two vertices in the tree.

Note that two nodes in the tree cannot be connected again by a link, thus forbidding loops to be formed.

Prim's algorithm starts by choosing a random node in the network, and then growing outward to the "cheapest" link which is adjacent to the starting node. Each link which is "invaded" is added to the growing cluster (tree), and the process is iterated until every site has been reached. Bonds can only be invaded if they do not produce a loop, so that the tree structure is maintained [20]. This process resembles *invasion percolation* in the Physics literature.

A direct consequence of the invasion process is that a path between two points A and B on the MST is the path whose maximum weight is minimal, i.e., the *minimal-barrier path*[29]. This is because if there were another path with a smaller barrier (i.e. maximal weight link) connecting A and B, the invasion process would have chosen that path to be on the MST instead.

The minimal-barrier path is important in cases where the "bottleneck" link is important. For example, in streaming video broadcast on the Internet [21, 22], it is important that each link along the path to the client will have enough capacity to support the transmission rate, and even one link with not enough bandwidth can become a bottleneck and block the transmission. In this case we will choose the minimal-barrier path rather than the optimal path[30].

An equivalent algorithm for generating the MST is ***Kruskal's algorithm:***

(a) Create a forest F (a set of trees), where each vertex in the graph is a separate tree.

(b) Create a set S containing all the edges in the graph.

(c) While S is nonempty:

- Remove an edge with minimum weight from S.
- If that edge connects two different trees, then add it to the forest, combining two trees into a single tree.
- Otherwise discard that edge.

---

[29] This path is sometimes referred to as the "min-max path" [49].

[30] In this example the weight of the link is related to its inverse capacity (the capacity of a link is roughly proportional to its bandwidth) and is also proportional to its delay time. The optimal path is thus the path with a minimum total delay time, which is less important for video broadcast.

Note that an edge cannot connect a tree to itself, thus forbidding loops to be formed.

Kruskal's algorithm resembles the percolation process because we add links to the forest according to increasing order of weights. The forest is actually the set of percolation clusters growing as the conductance probability $p$ (i.e. the weight of the link that is currently being added) is increasing (as long as $p < p_c$, see below).

It was noted by Dobrin *et al*.[20] that the geometry of the MST depends only on the unique ordering of the links of the network according to their weights. It does not matter if the weights are nearly the same or wildly different, it is only their *ordering* that matters. Given a network with weights $\{r_i\}$ on the links, any transformation which preserves the ordering of the weights (e.g., the link which has the fiftieth largest energy is the same before and after the transformation) leaves the MST geometry unaltered. This property is termed "universality" of the MST. Thus, given a network with weights $\{r_i\}$, with $r_i$ being a random variable distributed uniformly in the range $[0,1]$, a transformation of the weights to $\{\tau_i\}$ such that $\tau_i = \exp(ar_i)$ (for $a > 0$) will leave the MST unchanged.

## 5.2.4. Structure of the minimum spanning tree

### 5.2.4.1. The bombing optimization algorithm

An equivalent algorithm for finding the MST in a weighted random network is the "*bombing optimization algorithm*" [18]. We start with the original ER network and

remove links in order of descending weights. If the removal of a link disconnects the graph, we restore the link and mark it “gray” [50]; otherwise the link (shown dotted in Fig. 5-2(a)) is removed. The algorithm ends and an MST is obtained when no more links can be removed without disconnecting the graph.

Let us apply the bombing algorithm on a weighted ER graph, with weights distributed uniformly between 0 and 1. Because the weights are randomly distributed, the bombing algorithm resembles a percolation process in which the links (above some weight $p$) are removed with probability $1-p$. However, in the bombing algorithm only links that close a loop can be removed (otherwise the graph will be disconnected). It is known that for ER networks the loops are negligible below criticality [5], i.e., the percolation clusters have almost no loops. Therefore the bombing process does not modify the percolation clusters — where the links have weights below $p_c$. Thus, the bombing modifies only links *outside* the clusters, so actually it is only the links of the supernode network that are bombed. Hence the MST resulting from bombing is composed of percolation clusters connected by gray links (Fig. 5-2(c)).

From the MST of Fig. 5-2(c) we now generate a new tree, the MST of the supernode network, which we call the “gray tree”, whose nodes are the supernodes and whose links are the gray links connecting them (see Fig. 5-2(d)). Note that bombing the original ER network to obtain the MST of Fig. 5-2(c) is equivalent to bombing the supernode network of Fig. 5-2(b) to obtain the gray tree, because the links inside the clusters are not bombed.

To summarize, the MST on a weighted ER network (with weights $r_i$ distributed uniformly between 0 and 1) is composed of two components (see Fig. 5-2(c) and (d)): (i) the percolation clusters, which consist of links with weights below $p_c$ (we will refer to these as the "black" links because they are not affected by the bombing), and (ii) the "gray tree" – a tree connecting the percolation clusters, and whose links have weights above $p_c$. We will next describe the properties of these two components.

Note that because of the universality of the MST, the above description applies to any random distribution of weights that can be mapped monotonously to the uniform distribution between 0 and 1. For example, in the case where the weights are distributed uniformly in the range $[a,b]$, or where the weights are given by the expression: $\tau_i = \exp(ar_i)$, where $r_i$ is distributed uniformly in the range $[0,1]$ and $a > 0$. MST's on SF networks have a similar structure [46].

In the following two subsections we will describe the properties of the percolation clusters and the gray tree, which together compose the MST.

### 5.2.4.2. The percolation clusters

The percolation clusters are described by percolation theory for $d \geq d_c = 6$. The average length along the largest cluster scales as $l_{perc} \sim N^{1/3}$ for ER networks and as $l_{perc} \sim N^{(\lambda-3)/(\lambda-1)}$ for SF networks [18] (See Chapter 2).

The number of clusters $N_{cl}$ at the critical threshold (which is also the number of supernodes $N_{sn}$ in the supernodes network and the gray tree) may be

calculated as follows: The MST consists of a total of $N-1$ links – one incoming link per node except for the tree root [19]. These may be divided into two groups: gray and black links (see Fig. 5-2(c)). The original network has a total of $N\langle k\rangle/2$ links[31]. The bombing algorithm leaves a fraction $p_c$ of "black" links inside the percolation clusters. Hence $N_{black}=N\langle k\rangle p_c/2$. The number of gray links is $N_{gray}=N_{cl}-1$ (see Fig. 5-2(d)). Thus:

$$N_{MST}=N_{black}+N_{gray}$$

$$N-1=N_{black}+\left(N_{cl}-1\right)$$

$$N_{cl}=N-N_{black}=N-N\langle k\rangle p_c/2=N\left(1-\langle k\rangle p_c/2\right)$$

For ER graphs, $p_c=1/\langle k\rangle$, thus $N_{cl}=N/2$, whereas for SF networks we get: $N_{cl}=N\left(1-\frac{1}{2}\frac{\langle k\rangle^2}{\langle k(k-1)\rangle}\right)$.

### 5.2.4.3. The "gray tree"

We find (Fig 5-5(a)) that the gray tree has also a scale-free degree distribution P(k), with $\lambda$=2.5 - same as the supernode network[32]. We also find

---

[31] By definition, the average degree is $\langle k\rangle=2E/N$, where *E* is the total number of links and *N* is the total number of nodes.

[32] MST's on scale-free networks with $\lambda$=2.5 were found to retain the original network's degree distribution [51, 52, 53]. The origin of this phenomenon is yet unknown.

(see Fig. 5-5(b)) that the average path length $l_{gray}$ on the gray tree scales as[33] $l_{gray} \sim \log(N_{sn}) = \log(N/2) \sim \log N$ [47]. Note that even though the gray tree is scale-free, it is not ultra-small [13], since the average length does not scale as $\log\log N$. This is because the gray tree is not a random scale-free tree; rather it is *optimal* (i.e. it is the MST of the supernode network).

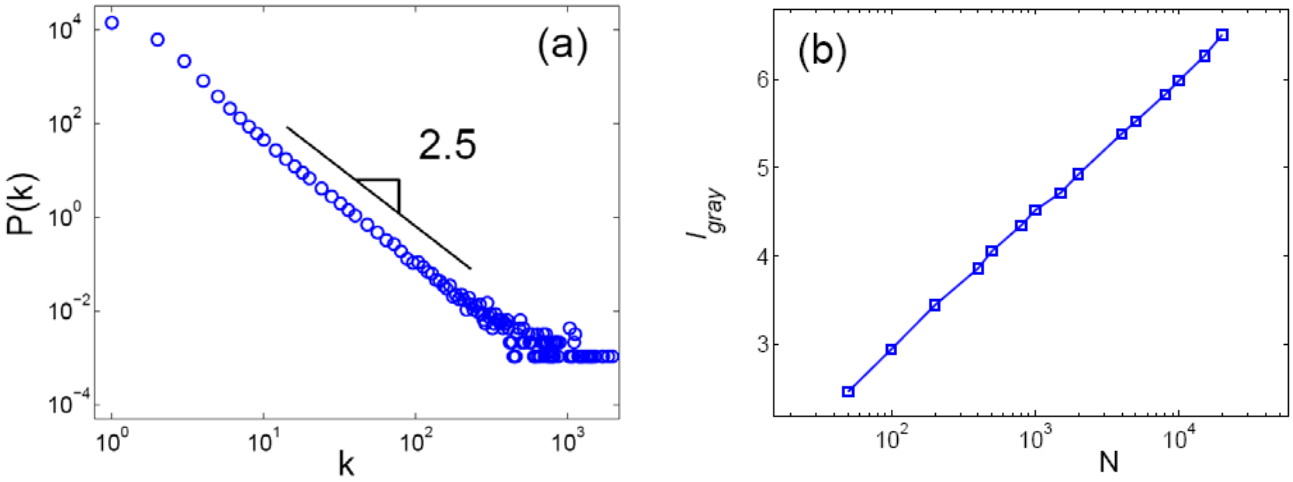



**Figure 5-5:** (a) The degree distribution of the "gray tree" (the MST of the supernode network, shown in Fig. 5-2(d)), in which the supernodes are percolation clusters and the links are the gray links. The distribution exhibits a scale-free tail with $\lambda = 2.5$, same as the supernode network. (b) The average path length $l_{gray}$ on the gray tree as a function of original network size. It is seen that $l_{gray} \sim \log(N_{sn}) \sim \log N$.

[33] Although Braunstein et al. [18] found that the length of the optimal path is $l_{opt} \sim (\log N)^{\lambda-1}$, for SF networks with 2<λ<3 in the strong disorder limit, this is valid only when multiple links between nodes do not exist. For SF networks that have multiple links, such as in our case (Fig. 5-2(b)), we find a shorter optimal path: $l_{opt} \sim \log N$ for 2<λ<3.

### 5.2.5. The "minimal barrier" path

As explained above, the MST is optimal in two senses: (i) the total weight of all links is minimal (ii) a path between any two nodes on the MST will encounter the smallest maximal barrier (weight) between these nodes. The last property is common to many physical systems. We will also see that the optimal path follows the "minimal-barrier" path in cases where the weights in the network are strongly fluctuating ("strong disorder"). Accordingly, we study the weights encountered when traveling along a typical path on the MST.

We consider all pairs of nodes in the original MST of N nodes (Fig. 5-2(c)) and calculate the typical path length $l_{typ}$, which is the average path length on the MST. For each path of length $l_{typ}$ we rank the weights on its links in descending order. For the largest weights ("rank 1 links"), we calculate the average weight $w_{r=1}$ over all paths. Similarly, for the next largest weights ("rank 2 links") we find the average $w_{r=2}$ over all paths, and so on up to $r = l_{typ}$. Fig. 5-6 shows $w_r$ as a function of rank r for three different network sizes N = 2000, 8000, and 32000. We can distinguish between two types of weights on the minimal-barrier path:

(a) Weights below $p_c$ – these belong to the black links inside the supernodes (i.e. clusters). Their weights are uniformly distributed because the bombing algorithm cannot remove links inside the clusters. The number of black links scales as $l_{black} \sim N^{1/3}$ for ER graphs (see Chapter 2).

(b) Weights above $p_c$ – these belong to the "gray links", which were removed and restored by the bombing algorithm. Therefore they are not uniformly

distributed. As shown above, the number of gray links along the optimal path scales as $l_{gray} \sim \log N$. Notice also that the highest weights (e.g. $w_{r=1}$ and $w_{r=2}$) are independent of the network size N.

We explain these findings as follows: the black links represent a "global" phenomena – an extensive part of the optimal path lies along the giant component (the largest percolation cluster), which is a fractal and whose radius scales as $N^{1/3}$ [18] (see Fig. 5-2(c) and Chapter 2). As opposed to this, the gray links are associated with the finite clusters (or supernodes), which are "local" phenomena. High weights are associated with small clusters which have a small number of external links. The bombing process is limited in removing links connected to small clusters, because removal of such a link is more likely to disconnect the cluster from the graph. Hence, the weights of the highest gray links along the "minimal barrier", being a local phenomena, path do not depend on the network size.

However, the average length of the minimal-barrier path (i.e., the length of the average path on the MST) is dominated by the largest percolation cluster[34] $l_{MST} = l_{black} + l_{gray} = o\left(N^{1/3}\right) + o\left(\log N\right) \sim N^{1/3}$.

[34] The question of exactly which proportion of the average minimal-barrier path follows largest percolation cluster is currently under research. Some discussion may be found in [46].

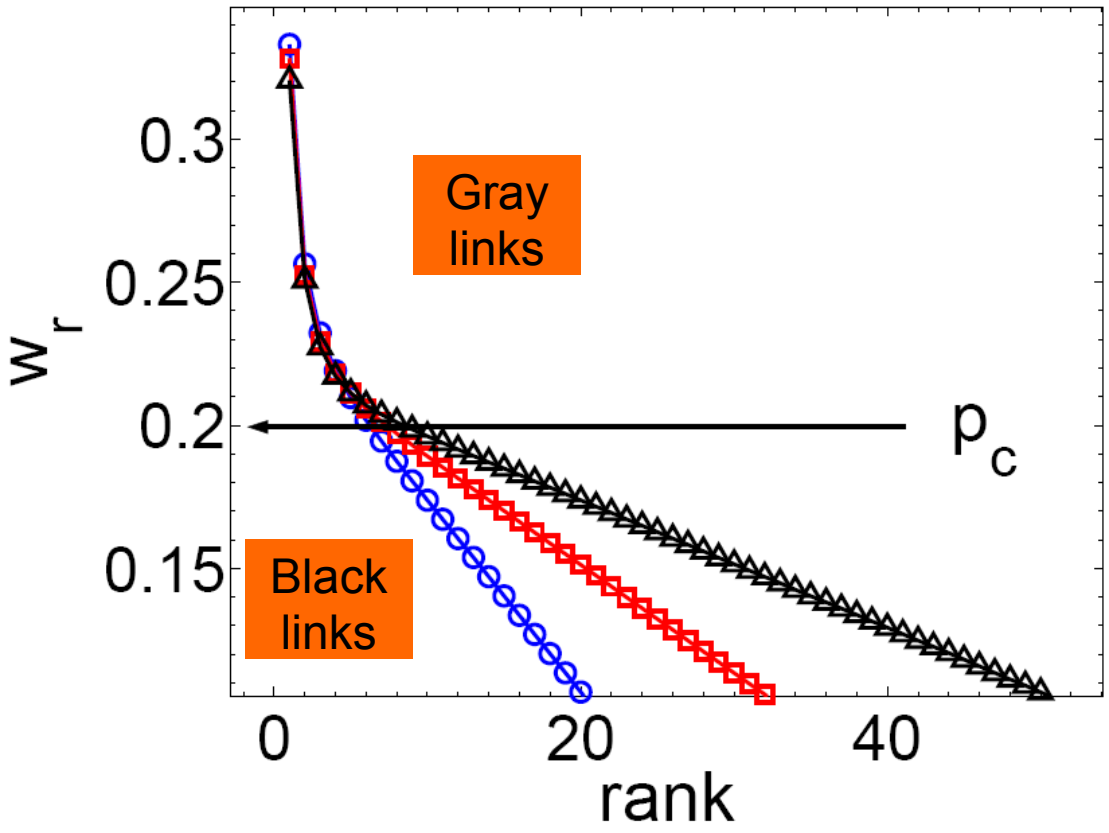



**Figure 5-6:** The average weights $w_r$ along the optimal path of an ER graph with $\langle k \rangle = 5$, sorted according to their rank. Different symbols represent different system sizes: N = 2000 (O), N = 8000 (□) and N = 32000 (Δ). Below $p_c = 0.2$, the weights are uniformly distributed, and their number scales as $N^{1/3}$. Above $p_c$ the number of weights along the minimal-barrier path scales as log(N), and their values are independent of network size N.

## 5.3. The transition from strong to weak disorder

In this subsection we will apply our findings from previous sections to study the behavior of optimal paths on networks with different types of disorder.

### 5.3.1. Strong and weak disorder

Consider a disordered random network, i.e. with each link we associate a "cost" or "weight". A common procedure to implement disorder on a network is as follows [18, 47, 54]: one assigns to each link i of the network a random number $r_i$, uniformly distributed between 0 and 1. The cost associated with link i is then:

$$\tau_i = \exp(a r_i),$$

where the parameter $a$ controls the broadness of the distribution of link costs, or the "strength" of disorder.

When the parameter $a$ is small, all weights in the network are of the same order of magnitude, e.g. $\tau_i = \{1.1, 2.5, 3, 4.7, 5, ...\}$. In this case the network is said to be "weakly disordered". However, when $a \to \infty$, the weights are of different orders of magnitude, e.g. $\tau_i = \{1, 1000, 10^6, 10^9, ...\}$, and the network is in "strong disorder".

It was found by simulations [18] that the length of the optimal path in weak disorder scales as $\log N$, similar to the shortest-hopcount path. However, in the strong disorder regime, the optimal path length scales as $N^{\nu_{opt}}$, where $\nu_{opt} = 1/3$ for ER networks, and $\nu_{opt} = (\lambda - 3)/(\lambda - 1)$ for SF networks (with $3 \le \lambda \le 4$) – see Table 5-1. It can be seen that the length of the optimal path in strong disorder scales the same as the length of the percolation cluster on the network, which is exponentially larger than the shortest-hopcount path.

| | Shortest-hopcount path | Optimal path, weak disorder | Optimal path, strong disorder |
|---|---|---|---|
| **ER graphs** | $\log N$ | $\log N$ | $N^{1/3}$ |
| **SF networks** ($3 \le \lambda \le 4$) | $\log N$ | $\log N$ | $N^{(\lambda-3)/(\lambda-1)}$ |
| **SF networks** ($2 < \lambda < 3$) | $\log\log N$ | - | $\log N$ |

**Table 5-1:** Scaling (with N) of the average length of the optimal path in disordered networks [18]. It can be seen that for strong disorder the optimal path length is exponentially larger than the shortest-hopcount path, i.e., the optimal path in strong disorder scales as a power of N rather than a logarithm. In weak disorder the optimal path is larger only by a multiplication factor. The behavior for scale-free networks in the anomalous regime $2 < \lambda < 3$ is similar, with the optimal path in strong disorder being exponentially larger than the shortest-hopcount path. The behavior in weak disorder is yet unclear due to numerical difficulties.

These results may be explained as follows: take a typical optimal path between two nodes A and B on the network. This path is chosen such that its total cost $\sum_i \tau_i$ is minimal. When the system is in strong disorder, all weights along the optimal path are of different orders of magnitude. Thus the maximal weight $\tau_{\max}$ is of higher order of magnitude than all other weights along the path, and it dominates the sum, i.e., $\sum_i \tau_i \approx \tau_{\max}$. Hence, in strong disorder the optimal path is chosen such as to minimize the maximal weight $\tau_{\max}$. In other words, in strong disorder the *optimal path* is equivalent to the *minimal-barrier path*.

We have seen in previous sections that the minimal-barrier path lies on the MST, which is composed of percolation clusters and gray links. Hence, the

optimal path in strong disorder ($a \to \infty$) follows (on average) the minimal-barrier path, whose length is dominated by the length of the percolation cluster $N^{\nu_{opt}}$.

We now ask the following question: how does the optimal path behave for intermediate values of disorder strength $a$ - between strong and weak disorder ?

### 5.3.2. Transition in average length of the optimal path

We have seen that the optimal path length scales differently for strong and weak disorder:

$$l(a) \sim \begin{cases} \log N & a \ll 1 \\ N^{1/3} & a \gg 1 \end{cases}$$

We propose the following scaling form:

$$l(a) = l_\infty \cdot F\left( \frac{1}{p_c} \frac{l_\infty}{a} \right) \tag{5.1}$$

Where $l_\infty = l(\infty) \sim N^{1/3}$ is the optimal path length for $a \to \infty$, and:

$$F(u) \sim \begin{cases} \dfrac{\log u}{u} & u \gg 1 \\ const. & u \ll 1 \end{cases}$$

The motivation for this scaling form is as follows: the weights along the optimal path for $a \to \infty$ (which lies on the MST) are $\tau_i = \exp(a r_i)$, where the values $r_i$ along the optimal path may be sorted according to their rank as shown in Fig. 5-7 (because of the universality of the MST, the monotonous transformation $\{\tau_i\} \to \{r_i\}$ leaves the MST unchanged). We now take any two

consecutive-rank "black" links $\tau_1 = \exp(ar_1)$ and $\tau_2 = \exp(ar_2)$, such that $r_1 < r_2$ (and[35] $r_1, r_2 < p_c$).

From the figure it can be seen that[36] $\Delta r = r_2 - r_1 = \frac{p_c}{l_\infty}$, and thus $\frac{\tau_2}{\tau_1} = \exp(a\Delta r) = \exp\left(a \cdot \frac{p_c}{l_\infty}\right)$. The criterion for strong disorder is that all consecutive rank weights will be of different orders of magnitude, i.e. $\frac{ap_c}{l_\infty} \gg 1$. This suggests that the *control parameter* of the scaling is $Z \equiv \frac{1}{p_c}\frac{l_\infty}{a}$. When $Z \ll 1$ we are in the strong disorder regime, and when $Z \gg 1$ we are in weak disorder.

The scaling form of Eq. (5.1) is confirmed by simulations, as shown in Fig. 5-8. The optimal path for infinite disorder strength, $l_\infty$, was found using Prim's algorithm for finding the minimal-barrier path, and the optimal path $l(a)$ was found using Dijkstra's algorithm. The collapse of all curves confirms that $Z \equiv \frac{1}{p_c}\frac{l_\infty}{a}$ is indeed the control parameter of the transition from strong to weak disorder. Similar results can be obtained on SF networks [47] and finite-dimensional lattices [55].

---

[35] We do not take weights above $p_c$ because the "gray" links are a local property and their effect on the optimal path length is much smaller than that of the "black" links.

[36] The number of gray links along the optimal path in strong disorder is much smaller than the number of black links. Thus, the number of black links may be well approximated by $l_\infty$.

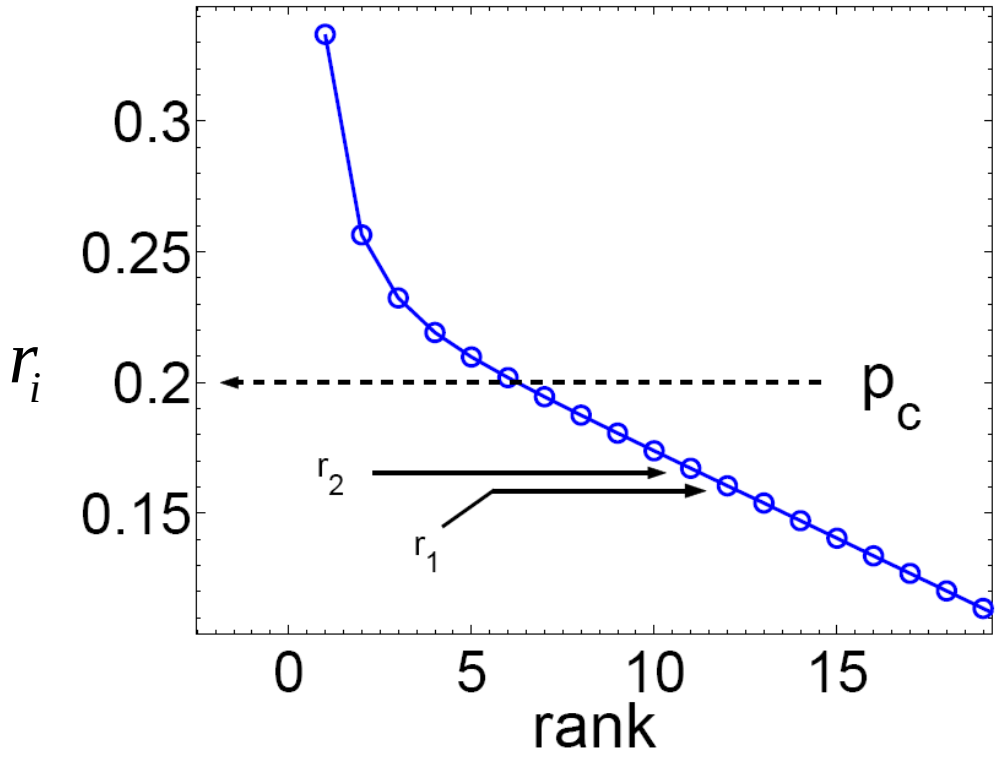



**Figure 5-7:** Weights $r_i$ along the typical optimal path (in strong disorder), sorted according to their ranks, for a network of N=2,000 nodes. There are approximately $l_\infty$ "black" links whose weights are uniformly distributed in the range $[0, p_c)$. The difference between two consecutive "black" weights is $\Delta r = r_2 - r_1 = \frac{p_c}{l_\infty}$.

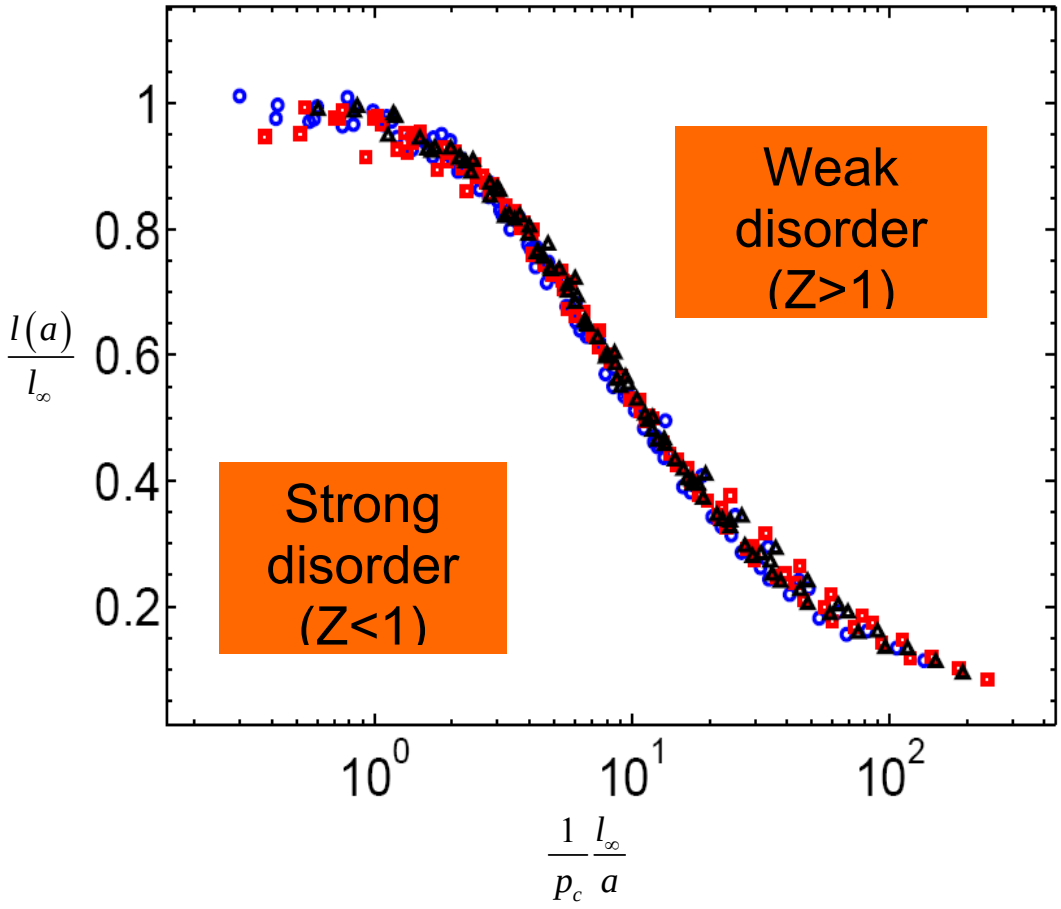



**Figure 5-8:** The order-disorder transition. Shown is $l(a)/l_\infty$ vs. $Z \equiv \frac{1}{p_c}\frac{l_\infty}{a}$ for networks different size N (which determines $l_\infty$), average degree $\langle k \rangle$ (which determines $p_c$), and disorder strength $a$. Different symbols represent different values of $\langle k \rangle$: $\langle k \rangle = 3$ (○), $\langle k \rangle = 5$ (□), and $\langle k \rangle = 8$ (Δ). The collapse of all curves on one function $F(Z)$ shows that a single control parameter $Z \equiv \frac{1}{p_c}\frac{l_\infty}{a}$ controls the transition.

### 5.3.3. Scaling of the optimal-path-lengths distribution

We now ask the following question: How are the different optimal paths in a network distributed? The distribution of the optimal paths lengths is especially important in communication networks, in which the overall network performance depends on the different path lengths between all nodes of the network, and not only on the average.

Recent works have studied the distribution form of the shortest paths lengths on minimum spanning trees [21, 42], which correspond to optimal paths on networks with large variations in link weights (i.e., strong disorder $a \to \infty$). The exact analytical form of the distribution is not known, but a good fit was found for a Maxwellian function [42].

However, the results in the previous section suggest that the optimal-paths-lengths distribution function obeys the following scaling form:

$$P(l,N,a) = \frac{1}{l_\infty} G\left( \frac{l}{l_\infty}, \frac{1}{p_c}\frac{l_\infty}{a} \right) \tag{5.2}$$

Where $P(l,N,a)$ is the probability to have an optimal path of length $l$ in a network of size N and disorder strength $a$. Notice that the parameter $Z \equiv \frac{1}{p_c}\frac{l_\infty}{a}$ determines the functional form of the distribution.

Relation (5.2) is supported by simulations for both ER and SF graphs, including SF graphs with $2 \le \lambda \le 3$, for which $p_c \to 0$ with system size N. We simulate ER graphs with weights on the links for different values of graph size N, control parameter $a$, and average degree $\langle k \rangle$ (which determines $p_c = 1/\langle k \rangle$; see

Table 5-2). We then generate the shortest path tree (SPT) using Dijkstra's algorithm [19] from some randomly chosen root node. Next, we calculate the probability distribution function of the optimal paths lengths from this node to all nodes in the graph. In Fig. 5-9 we plot $l_\infty P(l,N,a)$ vs. $l/l_\infty$ for different values of $N$, $a$, and $\langle k \rangle$. A collapse of the curves is seen for all graphs with the same value of $Z=(1/p_c)(l_\infty/a)$.

Figure 5-10 shows similar plots for SF graphs - with a degree distribution of the form $P(k)\sim k^{-\lambda}$ and with a minimal degree[37] m. A collapse is obtained for different values of $N$, $a$, $\lambda$ and m, with $\lambda \geq 3$ (see Table 5-3).

Next, we study SF networks with $2<\lambda<3$. In this regime the second moment of the degree distribution $\langle k^2 \rangle$ diverges, leading to several anomalous properties [12-14]. For example, the percolation threshold approaches zero with system size: $p_c \sim N^{-(3-\lambda)/(\lambda-1)} \to 0$, and the optimal path length $l_\infty$ was found numerically to scale logarithmically (rather than polynomially) with N [18]. Nevertheless, as can be seen from Fig. 5-11 and Table 5-4, the optimal paths lengths probability distribution for SF networks with $2<\lambda<3$ exhibits the same collapse for different values of N and a (although its functional form is different from that $\lambda>3$).

---

[37] Note that the minimal degree is m=2, thus ensuring that there exists a giant component for any $\lambda$, and thus $0<p_c<1$. For the case of m=1 there is almost surely no infinite cluster for $\lambda>\lambda_c\approx 4$ (or for a slightly different model, $\lambda_c$=3.47875 [56], resulting in an effective percolation threshold $p_c=\langle k \rangle / \langle k(k-1) \rangle > 1$. See [28, 56] for details.

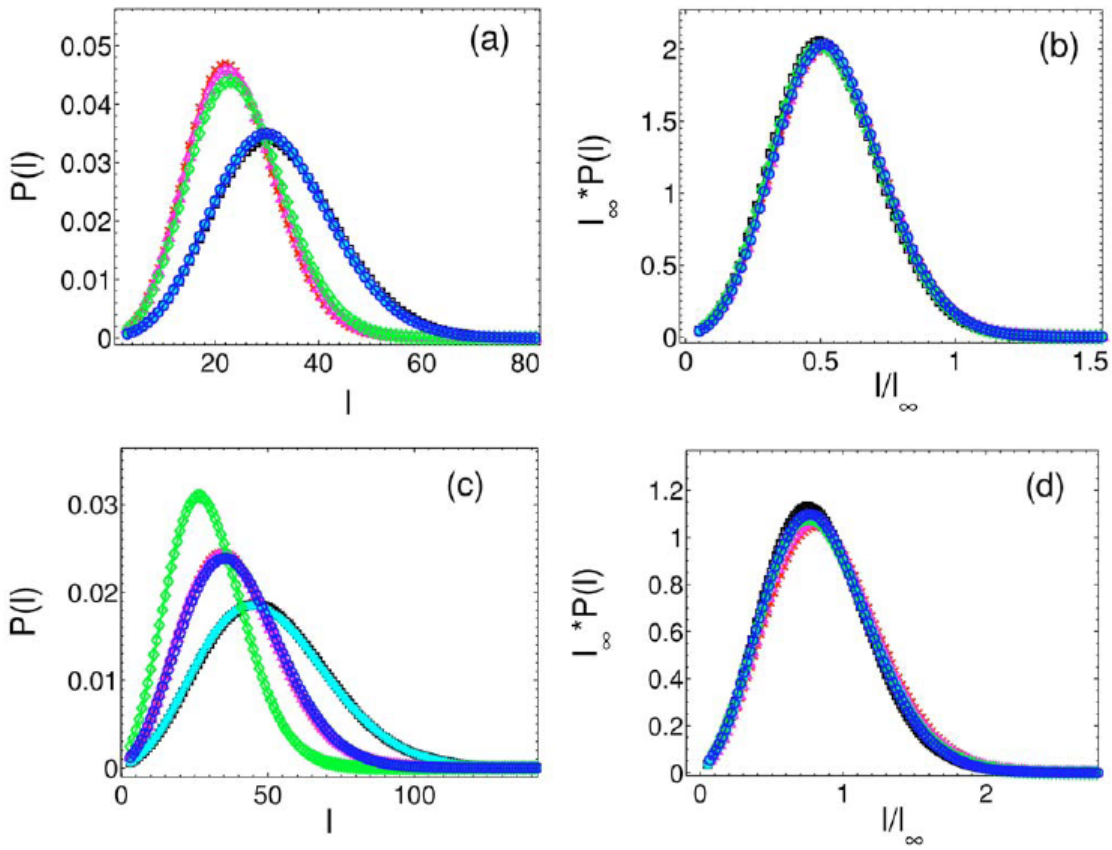



**Figure 5-9:** Optimal paths lengths distribution, $P(l)$, for ER networks with (a), (b) $Z \equiv (1/p_c)(l_\infty / a) = 10$ and (c), (d) $Z = 3$. (a) and (c) represent the un-scaled distributions for Z=10 and Z=3, respectively, while (b) and (d) are the scaled distributions. Different symbols represent networks with different characteristics such as size N (which determines $l_\infty \sim N^{1/3}$), average degree $\langle k \rangle$ (which determines $p_c = 1/\langle k \rangle$), and disorder strength $a$ (see Table 5-2 for details). Results were averaged over 1500 realizations.

| $N$ | $\langle k \rangle$ | $\ell_\infty$ | $p_c$ | $a$ | $Z = \frac{1}{p_c}\frac{\ell_\infty}{a}$ | Symbol |
|---|---|---|---|---|---|---|
| 4000 | 3 | 42.48 | 1/3 | 12.73 | 10 | × |
| 8000 | 3 | 60.59 | 1/3 | 18.16 | 10 | □ |
| 4000 | 5 | 44.01 | 1/5 | 22.00 | 10 | △ |
| 8000 | 5 | 58.42 | 1/5 | 29.19 | 10 | * |
| 4000 | 8 | 45.99 | 1/8 | 36.78 | 10 | ◇ |
| 8000 | 8 | 58.25 | 1/8 | 46.60 | 10 | ○ |
| 4000 | 3 | 42.48 | 1/3 | 42.45 | 3 | × |
| 8000 | 3 | 60.59 | 1/3 | 60.55 | 3 | □ |
| 4000 | 5 | 44.01 | 1/5 | 73.33 | 3 | △ |
| 8000 | 5 | 58.42 | 1/5 | 97.31 | 3 | * |
| 2000 | 8 | 34.94 | 1/8 | 93.15 | 3 | ◇ |
| 4000 | 8 | 45.99 | 1/8 | 122.62 | 3 | ○ |

**Table 5-2:** Different disordered ER graphs with same value of $Z \equiv (1/p_c)(l_\infty / a)$. The symbols refer to Fig. 5-9.

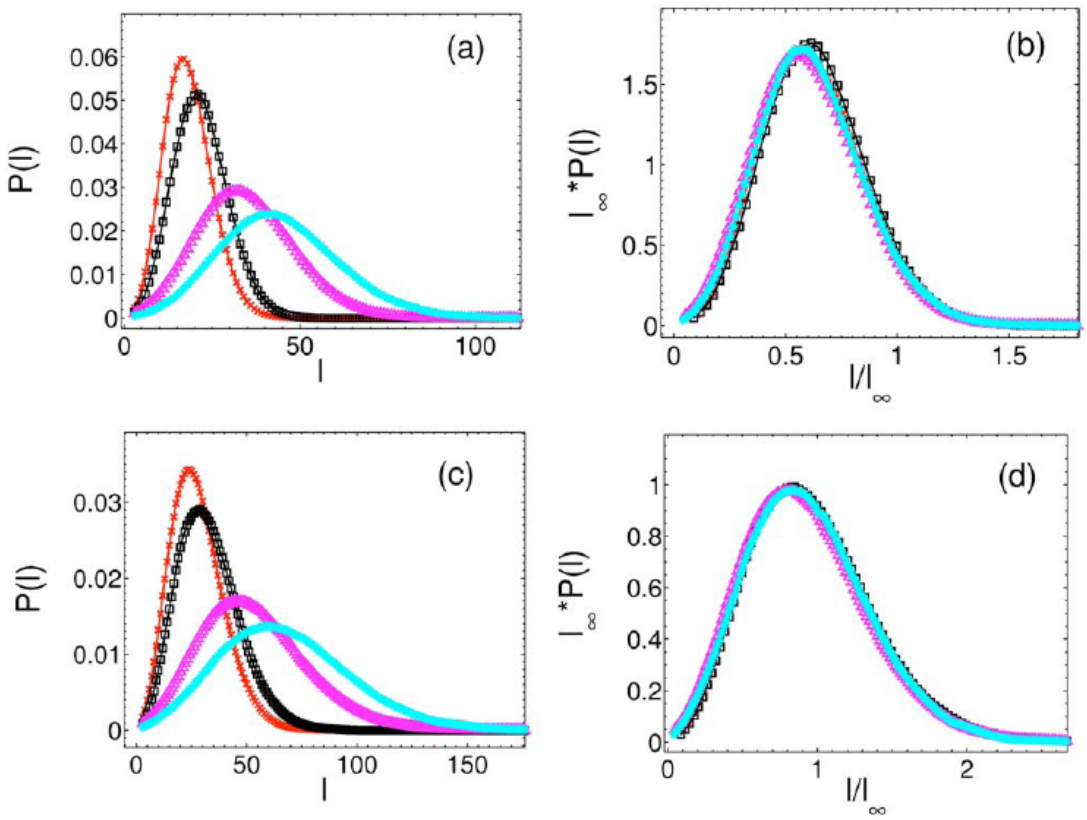



**Figure 5-10:** Optimal paths lengths distribution, $P(l)$, for SF networks with (a), (b) $Z \equiv (1/p_c)(l_\infty / a) = 10$ and (c), (d) $Z = 2$. (a) and (c) represent the un-scaled distributions for Z=10 and Z=2, respectively, while (b) and (d) are the scaled distributions. Different symbols represent networks with different characteristics such as size N (which determines $l_\infty \sim N^{\nu_{opt}}$), $\lambda$ and m (which determine $p_c$), and disorder strength $a$ (see Table 5-3). Results were averaged over 250 realizations.

| $N$ | λ | $m$ | $\ell_\infty$ | $p_c$ | $a$ | $Z=\frac{1}{p_c}\frac{\ell_\infty}{a}$ | Symbol |
|---|---|---|---|---|---|---|---|
| 4000 | 3.5 | 2 | 29.02 | 0.27 | 10.51 | 10 | × |
| 8000 | 3.5 | 2 | 34.13 | 0.26 | 12.88 | 10 | □ |
| 4000 | 5 | 2 | 57.70 | 0.5 | 11.54 | 10 | △ |
| 8000 | 5 | 2 | 72.03 | 0.5 | 14.40 | 10 | * |
| 4000 | 3.5 | 2 | 29.02 | 0.27 | 52.56 | 2 | × |
| 8000 | 3.5 | 2 | 34.13 | 0.26 | 64.44 | 2 | □ |
| 4000 | 5 | 2 | 57.70 | 0.5 | 57.70 | 2 | △ |
| 8000 | 5 | 2 | 72.03 | 0.5 | 72.03 | 2 | * |

**Table 5-3:** Different disordered SF graphs with same value of $Z \equiv (1/p_c)(l_\infty / a)$. The percolation threshold was calculated according to: $p_c = \langle k \rangle / \langle k(k-1) \rangle$. The symbols refer to Fig. 5-10.

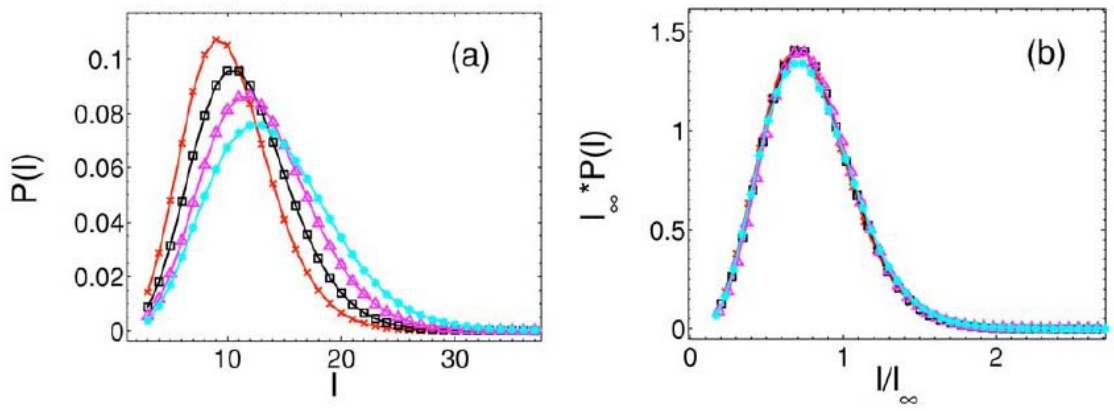



**Figure 5-11:** Optimal path lengths distribution function for SF graphs with $\lambda = 2.5$, and with $Z \equiv (1/p_c)(l_\infty / a) = 10$. (a) represents the un-scaled distribution for Z=10, while (b) shows the scaled distribution. Different symbols represent graphs with different characteristics such as size N (which determines $l_\infty \sim \ln N$ and $p_c \sim N^{-1/3}$), and disorder strength $a$ (see Table 5-4). Results were averaged over 1500 realizations.

| $N$ | λ | $m$ | $\ell_\infty$ | $p_c$ | $a$ | $Z=\frac{1}{p_c}\frac{\ell_\infty}{a}$ | Symbol |
|---|---|---|---|---|---|---|---|
| 2000 | 2.5 | 2 | 13.19 | 0.048 | 27.01 | 10 | × |
| 4000 | 2.5 | 2 | 14.66 | 0.037 | 38.70 | 10 | □ |
| 8000 | 2.5 | 2 | 16.14 | 0.029 | 54.50 | 10 | △ |
| 16000 | 2.5 | 2 | 17.69 | 0.022 | 77.48 | 10 | * |

**Table 5-4:** Different disordered SF graphs with same value of $Z \equiv (1/p_c)(l_\infty / a)$. Notice that $p_c \sim N^{-1/3} \to 0$ for $N \to \infty$. The symbols refer to Fig. 5-11.

### 5.3.4. Discussion

We have seen that both the average optimal path length and the optimal-paths-lengths distribution follow the simple scaling relations (5.1) and (5.2), and are controlled by a single parameter: $Z \equiv \frac{1}{p_c}\frac{l_\infty}{a}$.

We suggest the following explanation for this phenomenon: At strong disorder ($a \to \infty$) the optimal path follows the percolation cluster[38], whose radius scales as $N^{1/3}$. Because $a$ is very large, this optimal path takes long detours and twists in order to avoid links with large weights, resulting in an effective "dilution" of these links. For smaller values of $a$, the percolation cluster is followed only up to a characteristic length $\xi = ap_c$. Then, the optimal path takes a "shortcut" outside the percolation cluster. These shortcuts shorten[39] the optimal path from $N^{1/3}$ (strong disorder) to $\log N$ (weak disorder); see sketch in Fig. 5-12. The control parameter $Z \equiv \frac{l_\infty}{ap_c}$ is actually the number of "shortcuts" taken. Hence it determines the length of the average optimal path and also the functional form of the optimal-paths-lengths distribution.

In order to prove this argument, consider some portion of the optimal path that follows the percolation cluster at strong disorder. Assume that the length of this portion is $l_b$, and the values of $r_i$ are uniformly distributed in the range $[0, p_c)$ (see Fig. 5-7). The sum of weights S along this portion is:

---

[38] More accurately, the optimal path, which is equivalent to the minimal-barrier path, follows the MST, and an extensive portion of it (consisting only of "black" links) follows the largest percolation cluster.

[39] This resembles the shortcuts in the "small-world" model [57].

$$S=\sum_{i=1}^{l_b}\tau_i=\sum_{i=1}^{l_b}\exp\left(ar_i\right)\approx\frac{1}{\Delta r}\int_{r=0}^{p_c}\exp\left(ar\right)\cdot dr=\frac{l_b}{p_c}\int_{r=0}^{p_c}\exp\left(ar\right)\cdot dr=$$
$$=\frac{l_b}{p_c}\frac{1}{a}\left[\exp\left(ap_c\right)-1\right]\equiv\exp\left(ar^*\right)$$

Where:

$$r^*=\frac{1}{a}\ln\left\{\frac{l_b}{p_c}\frac{1}{a}\left[\exp\left(ap_c\right)-1\right]\right\}=$$
$$=\frac{1}{a}\ln\left(\frac{l_b}{ap_c}\right)+\frac{1}{a}\ln\left\{\left[\exp\left(ap_c\right)-1\right]\right\}\approx$$
$$\approx\frac{1}{a}\ln\left(\frac{l_b}{ap_c}\right)+p_c,$$

and we have assumed that: $\exp\left(ap_c\right)\gg 1$. Hence we get: $S=\sum_{i=1}^{l_b}\tau_i\equiv\exp\left(ar^*\right)$

where $r^*\approx p_c+\frac{1}{a}\ln\left(\frac{l_b}{ap_c}\right)$. Thus, taking a shortcut link outside the percolation cluster instead of following all these $l_b$ links (inside the percolation cluster) will reduce the cost of the optimal path, but only if the weight $r_{short-cut}$ of this link will obey:

$$\exp\left(ar_{short-cut}\right)<\exp\left(ar^*\right),$$

or:

$$p_c<r_{short-cut}<r^*\approx p_c+\frac{1}{a}\ln\left(\frac{l_b}{ap_c}\right).$$

If $l_b \ll ap_c$, such link cannot exist. However, if $l_b > ap_c$, there starts to be a finite probability for such a link to exist. Therefore the optimal path will follow the percolation cluster up to a characteristic length[40] $\xi = ap_c$.

The above results are consistent with results found for finite dimensional systems [23, 58, 59]: In a finite dimension, the parameter controlling the transition is $\frac{L^{1/\nu}}{ap_c}$, where L is the system length and $\nu$ is the correlation length critical exponent (for random graphs $\nu_l = 1$ when calculated in the chemical path metric). The expression $L^{1/\nu}$ is proportional to the number of "red bonds"[41] - bonds that, if cut, would disconnect the percolation cluster [15, 39]. The length of the optimal path changes most considerably when the optimal path deviates from the percolation cluster instead of passing through one of its "red bonds". Thus, in finite dimensional systems, the “red bonds” control the order-disorder transition.

## 5.4. Summary and conclusions

In this Chapter we have studied optimization problems in networks. We have shown that the minimum spanning tree is composed of percolation clusters, interconnected by a scale-free tree. We then used this to study the average and distribution of the optimal paths lengths in networks with different types of disorder. The analysis presented in this Chapter demonstrates how statistical

---

[40] Note that in the case of SF networks with $2 < \lambda < 3$, $p_c$ approaches zero (with increasing N) and consequently $\xi = ap_c \to 0$. This suggests that for any finite value of disorder strength $a$, a SF network with $2 < \lambda < 3$ is in the weak disorder regime.

[41] The "red bonds" are also termed "cutting bonds".

physics and percolation theory can be used to solve optimization problems in networks, which are of practical importance for the design of efficient routing and searching algorithms.

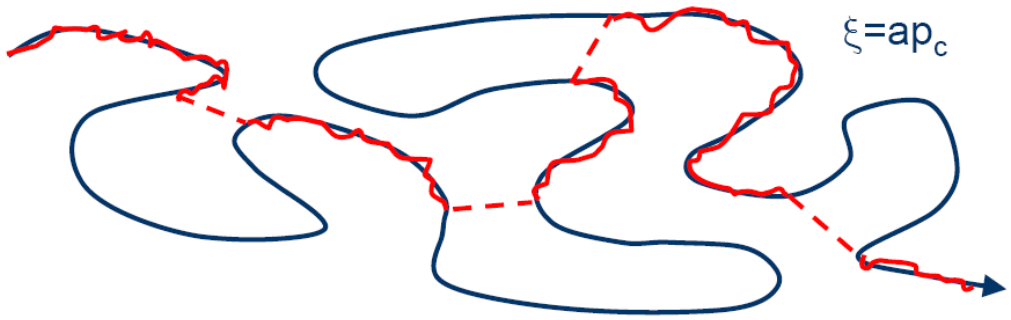



**Figure 5-12:** A sketch of the optimal path for a finite value of disorder strength $a$. The optimal path (red curve) follows the percolation cluster up to a characteristic length $\xi = ap_c$, after which it becomes beneficial to take a shortcut (dotted red lines) outside the percolation cluster. The optimal path in the case of strong disorder (blue curve) follows the percolation cluster, and it is much longer than the shortest-hopcount path.

# Bibliography


[1] A. L. Barabasi, *"Linked: The New Science of Networks"* (Perseus Books Group, Cambridge, MA, 2002).

[2] S. N. Dorogovtsev and J. F. F. Mendes, *"Evolution of Networks: From Biological Nets to the Internet and WWW"* (Oxford University Press, Oxford, 2003).

[3] R. Pastor-Satorras and A. Vespignani, *"Evolution and Structure of the Internet: A Statistical Physics Approach"* (Cambridge University Press, Cambridge, 2004).

[4] P. Erdos and A. Renyi, *"On the evolution of random graphs"*, Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17 (1960).

[5] R. Albert and A. L. Barabasi, *"Statistical mechanics of complex networks"*, Rev. Mod. Phys. **74**, 47 (2002).

[6] B. Bollob'as, *"Random Graphs"* (Cambridge University Press, Cambridge, 2001).

[7] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *"On Power-Law Relationships of the Internet Topology"*, Proceedings of the ACM SIGCOMM 1999 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, 251 (1999).

[8] J. Balthrop, S. Forrest, M. E. J. Newman, and M. M. Williamson, *"Technological networks and the spread of computer viruses"*, Science **304**, 527 (2004).

[9] E. Almaas, B. Kovacs, T. Vicsek, Z. N. Oltvai, and A. L. Barabasi, *"Global organization of metabolic fluxes in the bacterium Escherichia coli"*, Nature **427**, 839 (2004).

[10] A. L. Barabasi and Z. N. Oltvai, *"Network biology: Understanding the cell's functional organization"*, Nat. Rev. Genet. **5**, 101 (2004).

[11] M. E. J. Newman, *"Coauthorship networks and patterns of scientific collaboration"*, Proc. Natl. Acad. Sci. U. S. A. **101**, 5200 (2004).

[12] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *"Resilience of the Internet to random breakdowns"*, Phys. Rev. Lett. **85**, 4626 (2000).

[13] R. Cohen and S. Havlin, *"Scale-free networks are ultrasmall"*, Phys. Rev. Lett. **90** (2003).

[14] R. Cohen, D. ben-Avraham, and S. Havlin, *"Percolation critical exponents in scale-free networks"*, Phys. Rev. E **66** (2002).

[15] A. Bunde and S. Havlin, *"Fractals and Disordered Systems"* (Springer, Berlin, 1996).

[16] D. Stauffer and A. Aharony, *"Introduction to Percolation Theory"* (Taylor and Francis, London, 1994).

[17] R. Albert, H. Jeong, and A. L. Barabasi, *"Error and attack tolerance of complex networks"*, Nature **406**, 378 (2000).

[18] L. A. Braunstein, S. V. Buldyrev, R. Cohen, S. Havlin, and H. E. Stanley, *"Optimal paths in disordered complex networks"*, Phys. Rev. Lett. **91** (2003).

[19] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *"Introduction to Algorithms "* (MIT Press, Cambridge, MA, 2001).

[20] R. Dobrin and P. M. Duxbury, *"Minimum spanning trees on random networks"*, Phys. Rev. Lett. **86**, 5076 (2001).

[21] P. Van Mieghem and S. van Langen, *"Influence of the link weight structure on the shortest path"*, Phys. Rev. E **71**, 056113 (2005).

[22] D. Dolev, O. Mokryn, and Y. Shavitt, *"On Multicast Trees: Structure and Size Estimation "*. IEEE INFO-COM'03, San-Francisco, CA, USA, 2003,

[23] M. Porto, N. Schwartz, S. Havlin, and A. Bunde, *"Optimal paths in disordered media: Scaling of the crossover from self-similar to self-affine behavior"*, Phys. Rev. E **60**, R2448 (1999).

[24] M. Molloy and B. Reed, *"The size of the giant component of a random graph with a given degree sequence"*, Comb. Probab. Comput. **7**, 295 (1998).

[25] M. Molloy and B. Reed, *"A Critical-Point for Random Graphs with a Given Degree Sequence"*, Random Struct. Algorithms **6**, 161 (1995).

[26] H. S. Wilf, *"Generating functionology"* (Academic Press Professional, Inc., San Diego, CA, USA 1990).

[27] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *"Random graphs with arbitrary degree distributions and their applications"*, Phys. Rev. E **6402** (2001).

[28] T. Kalisky, R. Cohen, D. ben-Avraham, and S. Havlin, *"Tomography and stability of complex networks", Lecture Notes in Physics* **650**, 3. (Springer, Berlin, 2004), Proceedings of the 23rd LANL-CNLS Conference on Complex Networks, Santa-Fe, 2003, Eds. E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai.

[29] R. Cohen, D. Dolev, S. Havlin, T. Kalisky, O. Mokryn, and Y. Shavitt, *"On the tomography of networks and trees"*, cond-mat/0305582 (submitted to Phys. Rev. E).

[30] R. Xulvi-Brunet, W. Pietsch, and I. M. Sokolov, *"Correlations in scale-free networks: Tomography and percolation"*, Phys. Rev. E **68** (2003).

[31] S. Maslov and K. Sneppen, *"Specificity and stability in topology of protein networks"*, Science **296**, 910 (2002).

[32] M. E. J. Newman, *"Assortative mixing in networks"*, Phys. Rev. Lett. **89** (2002).

[33] H. Burch and B. Cheswick, *"Mapping the Internet"*, Computer **32**, 97 (1999).

[34] A. Vazquez, R. Pastor-Satorras, and A. Vespignani, *"Large-scale topological and dynamical properties of the Internet"*, Phys. Rev. E **65** (2002).

[35] T. Kalisky and R. Cohen, *"width of percolation threshold in complex networks"*, Phys. Rev. E (RC) (in press).

[36] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *"Network robustness and fragility: Percolation on random graphs"*, Phys. Rev. Lett. **85**, 5468 (2000).

[37] R. Cohen, S. Havlin, and D. ben-Avraham, *"Efficient immunization strategies for computer networks and populations"*, Phys. Rev. Lett. **91** (2003).

[38] M. E. J. Newman, *"Threshold effects for two pathogens spreading on a network"*, Phys. Rev. Lett. **95** (2005).

[39] A. Coniglio, *"Cluster structure near the percolation threshold"*, Journal of Physics A **15** 3829 (1982).

[40] F. Tzschichholz, A. Bunde, and S. Havlin, *"Loopless Percolation Clusters"*, Phys. Rev. A **39**, 5470 (1989).

[41] P. Grassberger, *"Directed percolation in 2+1 dimensions"*, Journal of Physics A: Mathematical and General **22**, 3673 (1989).

[42] L. A. Braunstein, S. V. Buldyrev, S. Sreenivasan, R. Cohen, S. Havlin, and H. E. Stanley, *"The optimal path in a random network", Lecture Notes in Physics*, **650**, 127 (Springer, Berlin, 2004), Proceedings of the 23rd LANL-CNLS Conference on Complex Networks, Santa-Fe, 2003, Eds. E. Ben-Naim, H. Frauenfelder and Z. Toroczkai.

[43] R. Cohen and S. Havlin, *"Fractal dimensions of percolating networks"*, Physica A **336**, 6 (2004).

[44] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *"Network motifs: Simple building blocks of complex networks"*, Science **298**, 824 (2002).

[45] T. Kalisky, L. A. Braunstein, S. V. Buldyrev, S. Havlin, and H. E. Stanley, *"Scaling of optimal-path-lengths distribution in complex networks"*, Phys. Rev. E **72** (2005).

[46] T. Kalisky, S. Sreenivasan, L. A. Braunstein, S. V. Buldyrev, S. Havlin, and H. E. Stanley, *"Scale-Free properties of weighted random graphs: Minimum Spanning Trees and Percolation", AIP conference proceedings*, **776**. Proceedings of the conference on “Science of Complex Networks:

From Biology to the Internet and WWW" (CNET 2004), Aveiro, Portugal, 2004, Eds. J.F.F. Mendes, S.N. Dorogovtsev, A. Povolotsky, F.V. Abreu and J.G. Oliveira.

[47] S. Sreenivasan, T. Kalisky, L. A. Braunstein, S. V. Buldyrev, S. Havlin, and H. E. Stanley, *"Effect of disorder strength on optimal paths in complex networks"*, Phys. Rev. E **70** (2004).

[48] E. Dekel and U. Alon, *"Optimality and evolutionary tuning of the expression level of a protein"*, Nature **436**, 588 (2005).

[49] E. Perlsman and S. Havlin, *"Directed polymer - directed percolation transition: the strong disorder case"*, Eur. Phys. J. B **43**, 517 (2005).

[50] A. S. Ioselevich and D. S. Lyubshin, *"Phase transition in a self-repairing random network"*, JETP **79**, 286 (2004).

[51] P. J. Macdonald, E. Almaas, and A. L. Barabási, *"Minimum spanning trees of weighted scale-free networks"*, Europhys. Lett **72 (2)**, 308 (2005).

[52] G. J. Szabo, M. Alava, and J. Kertesz, *"Geometry of minimum spanning trees on scale-free networks"*, Physica A **330**, 31 (2003).

[53] D. H. Kim, J. D. Noh, and H. Jeong, *"Scale-free trees: The skeletons of complex networks"*, Phys. Rev. E **70** (2004).

[54] S. Sreenivasan, T. Kalisky, L. A. Braunstein, S. V. Buldyrev, S. Havlin, and H. E. Stanley, *"Transition between strong and weak disorder regimes for the optimal path"*, Physica A **346**, 174 (2005).

[55] Y. M. Strelniker, R. Berkovits, A. Frydman, and S. Havlin, *"Percolation transition in a two-dimensional system of Ni granular ferromagnets"*, Phys. Rev. E **69** (2004).

[56] 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, Portland, Oregon, USA, 2000, p171–180. (ACM Press, New York)

[57] D. J. Watts and S. H. Strogatz, *"Collective dynamics of 'small-world' networks"*, Nature **393**, 440 (1998).

[58] Y. M. Strelniker, S. Havlin, R. Berkovits, and A. Frydman, *"Resistance distribution in the hopping percolation model"*, Phys. Rev. E **72** (2005).

[59] Z. H. Wu, E. Lopez, S. V. Buldyrev, L. A. Braunstein, S. Havlin, and H. E. Stanley, *"Current flow in random resistor networks: The role of percolation in weak and strong disorder"*, Phys. Rev. E **71** (2005).

# Supplement: List of publications for Tomer Kalisky

## Refereed journals

1. M. Sade, T. Kalisky and R. Berkovits, "*Study of localization transition on complex networks via spectral statistics*" (Phys. Rev. E, in press).
2. T. Kalisky and R. Cohen, "*Width of percolation transition in complex networks*", (Phys. Rev. E (RC), in press).
3. T. Kalisky, L. A. Braunstein, S. V. Buldyrev, S. Havlin and H. E. Stanley, "*Scaling of optimal-path-lengths distribution in complex networks*", Phys. Rev. E (RC) 72, R025102 (2005).
4. T. Kalisky, S. Sreenivasan, L. A. Braunstein, S. V. Buldyrev, S. Havlin and H. E. Stanley, "*Scale-free networks emerging from weighted random graphs*", (submitted to Phys. Rev. E).
5. S. Sreenivasan, T. Kalisky, L. A. Braunstein, S. V. Buldyrev, S. Havlin and H. E. Stanley, "*The Transition from Strong to Weak Disorder in Complex Networks*", Phys. Rev. E 70, 046133 (2004).
6. R. Cohen, D. Dolev, S. Havlin, T. Kalisky, O. Mokryn, and Y. Shavitt, "*On the tomography of networks and trees*", Hebrew University technical report TR49 (2002) and cond-mat/0305582, (submitted to Phys. Rev. E).
7. T. Kalisky, Y. Ashkenazy and S. Havlin, "*Volatility of Linear and Nonlinear Time Series*", Phys. Rev. E 72, 011913 (2005).

   Selected for publication in the virtual journal for biological physics research (August 2005).
8. Itzhack Dana and T. Kalisky, "*Symbolic dynamics for strong chaos on stochastic webs: General quasisymmetry*", Phys. Rev. E (RC) 53, R2025–R2028 (1996).

## Conference Proceedings:

9. T. Kalisky, S. Sreenivasan, L. A. Braunstein, S. V. Buldyrev, S. Havlin and H. E. Stanley , "*Scale-Free properties of weighted random graphs: Minimum Spanning Trees and Percolation*" [Proceeding of the conference on "Science of Complex Networks: From Biology to the Internet and WWW"

(CNET 2004), Aveiro, Portugal, 2004, Eds. J.F.F. Mendes, S.N. Dorogovtsev, A. Povolotsky, F.V. Abreu and J.G. Oliveira (AIP conference proceedings vol. 776)].

10. S. Sreenivasan, T. Kalisky, L. A. Braunstein, S. Havlin and H. E. Stanley, "*Transition between strong and weak disorder regimes for the optimal path*", [Proceedings of the Kolkata Conference on Networks], Physica A 346, 174-182 (2005).

11. S. Havlin , L. A. Braunstein, S. V. Buldyrev, R. Cohen, T. Kalisky, S. Sreenivasan, and H. E. Stanley, "*Optimal Path in Random Networks with Disorder: A mini review*", [Proceedings of the Kolkata Conference on Networks], Physica A 346, 82-92 (2005).

12. T. Kalisky, R. Cohen, D. ben-Avraham, and S. Havlin, "*Tomography and stability of complex networks*", Lecture Notes in Physics 650, 3 (2004). Proceedings of the 23rd LANL-CNLS Conference, "Complex Networks", Santa-Fe, 2003, Eds. E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai (Springer, Berlin, 2004).

13. N. Madar, T. Kalisky, R. Cohen, D. ben-Avraham, and S. Havlin, "*Immunization and epidemic dynamics*", Eur. Phys. Jour. B 38, 269 (2004). Proceedings of the Conference on "Growing Networks and Graphs in Statistical Physics, Finance, Biology and Social Systems", Rome, 2003, Eds. G. Caldarelli, and A. Vespignani.

# רשתות מורכבות: מבנה, חלחול, ואופטימיזציה

חיבור לשם קבלת התואר "דוקטור לפילוסופיה"
מאת
תומר קליסקי
המחלקה לפיסיקה

הוגש לסנט של אוניברסיטת בר-אילן
רמת גן א' טבת תשס"ו

עבודה זו נעשתה בהדרכתו של פרופ' שלמה הבלין מן המחלקה לפיסיקה של אוניברסיטת בר-אילן.

רשימת תודות בשער באנגלית.

# תוכן עניינים

# תקציר

בשנים האחרונות התעורר עניין רב בחקר רשתות מורכבות. התברר כי מערכות רבות בעולמנו ניתנות לתיאור על ידי "רשת" של קודקודים וקשתות המקשרות ביניהם. אחת הדוגמאות הבולטות ביותר היא רשת האינטרנט, המכילה מליוני מחשבים המקושרים ביניהם. דוגמאות נוספות ניתן למצוא ברשתות חברתיות בהן הקודקודים הם אנשים והקישורים הם היחסים החברתיים ביניהם, ורשתות ביולוגיות המתארות את יחסי הגומלין הכימיים בין חלבונים וגנים בתא.

לאחרונה נמצא כי רוב הרשתות בעולם אינן מתוארות היטב על ידי המודל המקובל של גראפים אקראיים. מודל זה, שהוצע על ידי ארדש ורניי (Erdös and Rényi) בשנות ה – 60, מניח כי התפלגות הדרגות ("דרגה" של קודקוד היא מספר הקשתות היוצאות ממנו) בגרף אקראי הינה התפלגות פואסונית (Poisson). אולם, התברר כי התפלגות הדרגות ברוב הרשתות האמיתיות הינה לפי חוק חזקה, כלומר: מספר הקודקודים היחסי בעלי דרגה $k$ נתון על ידי הביטוי $P(k) \sim k^{-\lambda}$, כאשר $2 < \lambda < 3$. לרשתות אלו ניתן השם "רשתות חסרות סקאלה" (scale-free networks) מכיוון שהן חסרות דרגה אפיינית: קיימים ברשת קודקודים עם דרגות מאוד גבוהות ומאוד נמוכות. נמצא כי לרשתות אלו תכונות ייחודיות, למשל, המרחק הממוצע בין זוג קודקודים ברשת קטן בהרבה מהמרחק הממוצע במודל של ארדש ורניי. כמו כן נמצא כי רשתות אלו עמידות יחסית להריסה אקראית של קודקודים או קשתות.

אף על פי שהתכונות המבניות של רשתות נחקרו, מתברר כי ברוב הרשתות בעולם יש "מחיר" או "משקל" לכל קשת. רשתות עם משקלים על הקשתות נקראות "רשתות ממושקלות" או "רשתות לא מסודרות", ותכונותיהן טרם נחקרו. באינטרנט, למשל, לכל קישור בין שני נתבים יש רוחב פס מוגדר, ולכן קצב האינפורמציה שניתן לשדר בין כל שני הנתבים הוא מוגבל. כאשר אנו שולחים אינפורמציה בין שני מחשבים מרוחקים ברשת, אנו מעוניינים לנתב אותה לאורך

המסלול בעל המחיר הכולל הנמוך ביותר. מסלול זה נקרא "המסלול האופטימלי". בעבודה זו נחקרו השאלות הבאות: מהו המבנה של המסלול האופטימלי? כיצד מתנהג המסלול האופטימלי ברשתות עם סוגים שונים של אי סדר? ומהי התפלגות האורכים של המסלולים האופטימליים השונים ברשת? הרעיון המרכזי בעבודת מחקר זו הוא שניתן לפתור בעיות אופטימיזציה ברשתות ממושקלות על ידי שימוש בכלים של פיסיקה סטטיסטית ותורת החילחול (פרקולציה).

העבודה מאורגנת כדלהלן: בפרקים הראשון והשני אנו מציגים סקירה של תוצאות קודמות ושל שיטות מקובלות בחקר רשתות מורכבות. הפרק השני כולל גם פורמאליזם חדש ופשוט יותר של שיטת הפונקציות היוצרות, כפי שהיא מיושמת בבעיות חילחול ברשתות.

בפרק השלישי אנו מציגים איפיון מבני חדש לרשתות חסרות סקאלה: ה"טומוגרפיה" של הרשת. לצורך איפיון זה אנו בוחרים את הקודקוד המקושר ביותר ברשת, ובוחנים את תכונות ה"שכבות" (כלומר, הקליפות הכימיות) סביבו. בכל שכבה כזו אנו סופרים את מספר הקודקודים ובודקים את התפלגות הדרגות של כל הקודקודים בשכבה. אנו מראים שלפונקצית התפלגות המרחקים של קודקודי הרשת מהקודקוד המקושר ביותר יש שני תחומים: בתחום ראשון יש עליה חזקה במספר הקודקודים, ואילו בתחום השני יש דעיכה אקספוננציאלית. בנוסף לכך אנו מראים כי דרגות הקודקודים בכל שכבה מתפלגות לפי חוק חזקה עם חיתוך אקספוננציאלי. בסיום הפרק אנו משווים את תוצאות המודל עם ממצאים מרשת האינטרנט.

בפרק הרביעי אנו חוקרים את רוחב סף החילחול ברשתות. ידוע מתורת החילחול בסריגים כי סף החילחול אינו סף חד, כי אם תחום בעל רוחב סופי התלוי בגודל המערכת. בתוך תחום זה מתקיימות כל התכונות הקריטיות המאפיינות את נקודת מעבר הפאזה. אנו מראים כי ברשתות מורכבות רוחב סף החילחול מקיים את הקשר $\Delta p_c \sim p_c / l$, כאשר $l \sim N^{\nu_{opt}}$ הוא האורך הממוצע של צביר החילחול, ו – N הוא מספר הקודקודים ברשת. עבור רשתות ארדש-רניי מתקיים $\nu_{opt} = 1/3$, ואילו עבור רשתות חסרות סקאלה (כלומר, רשתות בעלות התפלגות דרגות מהצורה $P(k) \sim k^{-\lambda}$ עבור $3 < \lambda < 4$) מתקיים $\nu_{opt} = (\lambda - 3)/(\lambda - 1)$. על מנת להראות

זאת אנו מוכיחים אנליטית ומראים נומרית שה"שרידות" (survivability) $S(p,l)$, כלומר, הסיכוי שצביר החילחול ישרוד $l$ קליפות כימיות עבור הסתברות הולכה p, מקיימת את הקשר הבא בקרבת הנקודה הקריטית: $S(p,l)=S(p_c,l)\cdot\exp\left[\frac{1}{p_c}(p-p_c)l\right]$. לכן, עבור הסתברויות הולכה בתחום $|p-p_c|<\frac{p_c}{l}$ התנהגות המערכת היא בלתי ניתנת להבחנה מהתנהגות המערכת בקריטיות.

בפרק החמישי אנו מיישמים את תורת החילחול לבעיות אופטימיזציה ברשתות. בתחילת הפרק אנו מראים כי כאשר לכל קשת יש משקל אקראי, גם ברשתות הקלאסיות של ארדש ורניי מתקיימות תכונות חסרות סקאלה. אנו מראים כי על ידי איחוד של קודקודים המקושרים על ידי קשתות בעלות משקלים מתחת לסף החילחול לתוך "קודקודי-על" (supernodes), מתקבלת רשת חסרת סקאלה עם התפלגות דרגות $P(k)\sim k^{-\lambda}$ ו - $\lambda=2.5$. בנוסף לכך, אנו מראים כי העץ הפורש המינימלי ברשתות ארדש-רניי (עם N קודקודים) מורכב מ - $N/2$ צבירי חלחול המקושרים ביניהם על ידי אוסף קשתות היוצרות עץ חסר סקאלה עם $\lambda=2.5$.

בהמשך הפרק אנו משתמשים בתוצאות אלו על מנת לחקור את התנהגות המסלול האופטימלי ברשתות ממשוקלות. לכל קשת $i$ אנו נותנים משקל $\tau_i=\exp(ar_i)$, כאשר $r_i$ הוא מספר אקראי בין 0 ל – 1, והפרמטר $a$ קובע את מידת השונות (או ה"חוזק") של אי הסדר. ברשתות אלו נמצא כי אורך המסלול האופטימלי שקול ל - $N^{\nu_{opt}}$ באי סדר חזק, ול - $\log N$ באי סדר חלש. אנו מראים כי האורך הממוצע של המסלול האופטימלי, עבור ערכי ביניים של הפרמטר $a$, מקיים את יחס הכיול הבא: $l(a)=l_\infty\cdot F\left(\frac{1}{p_c}\frac{l_\infty}{a}\right)$, כאשר $l_\infty$ הוא אורך המסלול האופטימלי באי סדר חזק ($a\to\infty$), ו - $p_c$ הוא סף החלחול. ההתפלגות של ארכי המסלולים

האופטימליים ברשת הנה בעלת צורה פונקציונלית דומה: $P(l,N,a)=\frac{1}{l_\infty}G\left(\frac{l}{l_\infty},\frac{1}{p_c}\frac{l_\infty}{a}\right)$.

היחסים הנ"ל מודגמים על רשתות ארדש-רניי וגם על רשתות חסרות סקאלה. המיוחד ביחסים אלו הוא שפרמטר בודד $Z\equiv\frac{1}{p_c}\frac{l_\infty}{a}$ קובע את תכונות הממוצע וההתפלגות של המסלולים האופטימליים.

כדי להסביר זאת אנו מראים כי מסלולים האופטימליים ברשת עוקבים (בממוצע) אחר צביר החלחול עד האורך אפייני $\xi=ap_c$. לאחר מכן הם עושים "קיצור דרך" ועוברים דרך קשת הנמצאת מחוץ לצביר. הפרמטר $Z\equiv\frac{1}{p_c}\frac{l_\infty}{a}$ מייצג את מספר קיצורי הדרך הללו. מכך אנו למדים כי התנהגותן של רשתות ממושקלות דומה למערכות חלחול הנמצאות בקרבת הנקודה הקריטית, כך שהתנהגותן הנה פראקטלית עד אורך קורלציה $\xi$.

לסיכום, בעבודה זו חקרנו בעיות מבנה, חלחול, ואופטימיזציה ברשתות מורכבות עם אי סדר. ניתחנו בין היתר את מבנה העץ הפורש המינימלי, ואת הממוצע וההתפלגות של ארכי המסלולים האופטימליים. עבודה זו מדגימה כיצד ניתן לפתור בעיות שימושיות של אופטימיזציה ברשתות באמצעות שיטות שמקורן בפיסיקה סטטיסטית ותורת הפרקולציה.