LECTURE NOTES IN PHYSICS

E. Ben-Naim
H. Frauenfelder
Z. Toroczkai
(Eds.)

Complex Networks

Springer

E. Ben-Naim H. Frauenfelder Z. Toroczkai (Eds.)

Complex Networks

Springer

Editors

Eli Ben-Naim
Hans Frauenfelder
Zoltan Toroczkai
Los Alamos National Laboratory
Complex Systems Group
Theoretical Division
87545 Los Alamos, NM
USA

E. Ben-Naim H. Frauenfelder Z. Toroczkai (Eds.), *Complex Networks*, Lect. Notes Phys. **650** (Springer, Berlin Heidelberg 2004), DOI 10.1007/b98716

Library of Congress Control Number: 2004107805

ISSN 0075-8450
ISBN 3-540-22354-1 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Data conversion: PTP-Berlin Protago-TeX-Production GmbH
Cover design: *design & production*, Heidelberg

Printed on acid-free paper
54/3141/ts - 5 4 3 2 1 0

Preface

The study of interacting particle systems has traditionally focused on cases where the underlying topology can be described by simple structures such as regular crystalline lattices or by a continuum medium. The emerging science of complex networks addresses complementary situations where the underlying topology is a graph whose structure is complex, irregular, and dynamically evolving. Complex networks are ubiquitous in nature. Natural networks include biological networks (metabolic networks, gene regulatory networks, protein interaction networks, signaling networks, epidemic networks), and ecological networks (food webs). Man-made networks include communications networks (WWW, Internet, phone, wireless), transportation infrastructures (power grid, waterways, natural gas, roadways, airlines), and social interactions (acquaintance networks, scientific collaboration networks, terrorist networks).

Network science dates back to Leonhard Euler who initiated graph theory by his solution in 1736 to the famous Königsberg bridges problem. For the next 200 years graph theory dealt with regular or small structures. Network science was reborn with the introduction of random graph theory, through the seminal works of Ray Solomonoff and A. Rapoport in 1951, and separately, by the works of Pál Erdős and Alfréd Rényi in 1959-1960 who introduced probabilistic methods to graph theory.

Currently, a third revolution is underway. It has been motivated by the emergence of communication networks and the need to characterize biological networks and facilitated by the availability of large data sets and the explosive growth in computing power. Based on characteristics of real-world networks, the small-world network model by Duncan Watts and Steven Strogatz and the preferential attachment model of scale-free networks by Albert-László Barabási and Réka Albert have reshaped the way we think of networks.

These contributions showed that the structure of many real-world large-scale complex networks are far from those of the traditional random graphs, and they opened up many avenues for future research. They demonstrated that complex networks is an intellectually deep and ripe area, relevant to many scientific disciplines including physics, biology, engineering, and social science, far beyond the traditional fields of mathematics and computer science.

Now, the research front turns to networks dynamics. Most networks have the role and function to transport or transfer entities (information, energy, etc.) along the links. Optimizing transport efficiency and quantifying network vulnerabilities and robustness constitute the next open questions. Predicting the dynamical evolution of the network structure and its coupling with the transport processes are the ultimate challenge for complex networks science.

This volume of the Lecture Notes in Physics series focuses on the application of techniques from statistical physics to characterization and modeling of complex networks. There is a deep connection between statistical physics and statistical graph theory as both aim to characterize macroscopic observables based on a probabilistic treatment of all microstates of the system. As a concrete example, the polymerization process proposed by Paul Flory and used by chemical physicists to model gelation is equivalent to the growth of a random graph. This natural connection between statistical mechanics and statistical graph theory is currently being exploited by many physicists and the present volume presents the state-of-the-art in the application of statistical physics methods to complex networks research.

This volume consists of four parts. The first two parts concern theory and modeling of networks while the last two parts involve applications to real-world networks. Part I deals with theoretical characterization of structural properties of networks including spectral and extremal properties and structural robustness. Part II addresses dynamical aspects of networks including evolving networks, dynamical processes and transport on networks, and synchronization of networks. Part III focuses on information and social networks including publication networks, collaboration networks, email communication, and board membership networks. Part IV starts with an overview of networks in biological systems, followed by applications to genetic and neural networks.

The articles in this volume were written by speakers at the conference "Complex Networks: Structure, Dynamics, and Function", the 23rd annual conference of the Center for Nonlinear Studies at Los Alamos National Laboratory, held from May 12–16, 2003 in Santa Fe, New Mexico, USA. The papers in this volume are review articles by experts in network science, many of whom made seminal contributions to the foundations of this novel field. As a collection, this volume covers a large fraction of the state-of-the art of complex network research. The articles are aimed at students, newcomers to the field, as well as experts. All articles have been carefully peer-reviewed not only for scientific content but also for self-consistency and readability.

The editors thank the authors for their contributions and the referees, whose comments improved the articles in a significant way. The editors also wish to thank the conference organizers Benjamin McMahon, Paul Fenimore, and Pieter Swart, as well as the conference coordinator Roderick Garcia.

Los Alamos, New Mexico, USA
February 2004

Eli Ben-Naim
Hans Frauenfelder
Zoltan Toroczkai

Contents

Part I

Network Structure

Tomography and Stability of Complex Networks

Tomer Kalisky[1], Reuven Cohen[1,2], Daniel ben-Avraham[3], and Shlomo Havlin[1]

[1] Minerva Center and Department of Physics, Bar-Ilan University, Ramat-Gan, Israel
[2] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel
[3] Department of Physics, Clarkson University, Potsdam, NY 13699, USA

Abstract. We study the structure of generalized random graphs with a given degree distribution $P(k)$, and review studies on their behavior under both random breakdown of nodes and intentional attack on the most highly connected nodes. We focus on scale free networks, where $P(k) \propto k^{-\lambda}$, for $m < k < K$. We first examine the "Tomography" of these networks, i.e. the structure of layers around a network 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. We then show that scale free networks with $\lambda < 3$ are robust to random breakdown, but vulnerable to intentional attack. We also describe the behavior of the network near the phase transition and show that the critical exponents are influenced by the scale free nature of the network. We show that the critical exponent for the infinite cluster size behaves as $\beta = 1/|\lambda - 3|$, and the exponent for the finite clusters size distribution behaves as $\tau = \frac{2\lambda-3}{\lambda-2}$, for $2 < \lambda < 4$. For $\lambda > 4$ the exponents are $\beta = 1$ and $\tau = 2.5$ as in normal infinite dimensional percolation. It is also shown that for all $\lambda > 3$ the exponent for the correlation length is $\nu = 1$ and formulas for the fractal dimensions are obtained. The size of the largest cluster at the transition point, known to scale as $N^{2/3}$ in regular random graphs, is shown to scale as $N^{(\lambda-2)/(\lambda-1)}$ for $3 < \lambda < 4$ and as $N^{2/3}$ for $\lambda > 4$.

1 Introduction

Much attention has been focused recently on the topic of complex network behavior [1–5]. Most of the interest has been on scale-free networks, which are believed to represent many phenomena in nature. Scale-free degree distributions have been observed in the Internet [6], World Wide Web (WWW) [7], metabolic networks [8] and many others. For recent reviews see [9–13]. In this paper we review the topics of structure [14] and percolation of such networks [1–4]. Understanding network structure can help devise better networks topologies. It may also help design more efficient algorithms for routing and searching in communications networks by taking advantage of the network structure. Percolation is especially important in forecasting and preventing network malfunctions in the Internet, as well as other realistic networks, and may also be important in the understanding of the stability of biological and chemical processes [15].

Percolation theory has been studied for some decades by physicists and mathematicians. In general it deals with the dilution of a fraction p (alternatively, the

T. Kalisky, R. Cohen, D. ben-Avraham, and S. Havlin, Tomography and Stability of Complex Networks, Lect. Notes Phys. **650**, 3–34 (2004)
http://www.springerlink.com/

occupation with a density $q = 1 - p$) of the sites or bonds in a graph [16,17]. It is known that for many graphs a finite threshold p_c exists, such that for dilution of $p < p_c$ a spanning cluster (*i.e.* a cluster of size proportional to that of the entire network) exists. While for $p > p_c$ the graph is fragmented into small clusters. When a spanning cluster exists, its size relative to the graph is denoted $P_\infty(p)$. Near the transition point $P_\infty \sim (p_c - p)^\beta$, where β (as well as other "critical exponents" such as ν, τ and σ) is universal – that is, depends only on the dimension and large scale properties of the graph and not on the local structure. At the transition point the clusters are fractals, while above and below that point the clusters are fractals up to length scale $\xi(p)$ (the correlation length) and have the dimension of the graph above ξ. Near criticality, $\xi \sim |p_c - p|^{-\nu}$.

The number of clusters of size s near criticality also follows a scaling form:

$$n_s \sim s^{-\tau} e^{-s/s^*} . \tag{1}$$

At $p = p_c$, the exponential cutoff $s^* \sim |p - p_c|^{-\sigma}$ diverges and the tail of the distribution behaves as a power law.

The structure of this paper is as follows: In Sect. 2 we discuss general results applicable to generalized random graphs with an arbitrary degree distribution. In Sect. 3 we discuss networks having a scale-free degree distribution, which will be the main concern of this paper. In Sect. 4 we discuss the tomography of scale-free networks, that is, their partition into layers surrounding the maximally connected node at different distances. Section 5 presents the model of random breakdown in scale-free networks and analytical and numerical results for this kind of failure. Section 6 offers a similar approach for an intentional attack on the most highly connected nodes. Section 7 presents an analytical derivation of the critical exponents for the percolation transition on scale-free networks, and finally Sect. 8 presents conclusions and prospects.

2 General Results

2.1 Condition for a Spanning Cluster

For a graph having degree distribution $P(k)$ to have a spanning cluster, a site which is reached by following a link from the giant cluster must have at least one other link in average to allow the cluster to exist[4]. For this to happen the average degree of a site must be at least 2 (one incoming and one outgoing link) given that the site i is connected to j:

$$\langle k_i | i \leftrightarrow j \rangle = \sum_{k_i} k_i P(k_i | i \leftrightarrow j) = 2. \tag{2}$$

[4] If we dilute the graph up to near p_c, the remaining structure resembles a tree, or a branching process. One can show that a branching process with an average branching factor that is less than 1 will die out with probability 1 after a finite number of steps [18].

Using Bayes rule we get

$$P(k_i|i \leftrightarrow j) = P(k_i, i \leftrightarrow j)/P(i \leftrightarrow j) = P(i \leftrightarrow j|k_i)P(k_i)/P(i \leftrightarrow j), \tag{3}$$

where $P(k_i, i \leftrightarrow j)$ is the *joint* probability that node i has degree k_i and that it is connected to node j. For randomly connected networks (neglecting loops) $P(i \leftrightarrow j) = \langle k \rangle/(N-1)$ and $P(i \leftrightarrow j|k_i) = k_i/(N-1)$, where N is the total number of nodes in the network. Using the above criteria (2) reduces to [19,2]:

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} = 2, \tag{4}$$

at the critical point. A spanning cluster exists for graphs with $\kappa > 2$, while graphs with $\kappa < 2$ contain only small clusters whose size is not proportional to that of the entire network. This criterion was derived earlier by Molloy and Reed [19] using somewhat different arguments.

The neglecting of loops can be justified below the threshold since the probability for a bond to form a loop in an s-node cluster is proportional to $(s/N)^2$ (i.e., proportional to the probability of choosing two sites in that cluster). Calculating the fraction of loops P_{loop} in the system yields:

$$P_{loop} \propto \sum_i \frac{s_i^2}{N^2} < \sum_i \frac{s_i S}{N^2} = \frac{S}{N}, \tag{5}$$

where the sum is over all clusters in the system and s_i is the size of the ith cluster. Therefore, the fraction of loops in the system is less than or proportional to S/N, where S is the size of the largest cluster. Below the critical threshold there is no spanning cluster in the system and therefore the fraction of loops is negligible. Hence, until $\kappa = 2$ loops can be neglected. At the threshold the structure of the spanning cluster is almost a tree. Above the threshold loops can no longer be neglected, but since this only happens when a spanning cluster exists the criterion in (4) is valid as a criterion for finding the critical point. A derivation of the exact conditions under which (4) is valid can be found in [19].

2.2 Critical Threshold for Percolation

The above reasoning can be applied to the problem of percolation on a generalized random network. If we randomly remove a fraction p of the sites (or bonds), the degree distribution of the remaining sites will change. For instance, sites with initial degree k_0 will have, after the random removal of nodes, a different number a connections, depending on the number of removed neighbors. The new number of connections will be binomially distributed. If we begin with a distribution of degrees $P_0(k_0)$, the new distribution of degrees of the network will be:

$$P(k) = \sum_{k_0=k}^{\infty} P_0(k_0) \binom{k_0}{k} (1-p)^k p^{k_0-k}. \tag{6}$$

Calculating the first moment for this distribution, given $\langle k_0 \rangle$ and $\langle k_0^2 \rangle$ for the original distribution leads to:

$$\langle k \rangle = \sum_{k=0}^{\infty} P(k)k = (1-p)\langle k_0 \rangle. \tag{7}$$

In the same manner we can calculate the second moment:

$$\langle k^2 \rangle = \sum_{k=0}^{\infty} P(k)k^2 = (1-p)^2 \langle k_0^2 \rangle + p(1-p)\langle k_0 \rangle. \tag{8}$$

Both quantities can be substituted into (4) to find the criterion for criticality. This yields:

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{(1-p)^2 \langle k_0^2 \rangle + p(1-p)\langle k_0 \rangle}{(1-p)\langle k_0 \rangle} = 2. \tag{9}$$

Reorganizing (9), one gets the critical threshold for percolation [2]:

$$1 - p_c = \frac{1}{\kappa_0 - 1}, \tag{10}$$

where $\kappa_0 \equiv \langle k_0^2 \rangle / \langle k_0 \rangle$ is calculated using the original distribution, before the removal of sites.

Equations (4) and (10) are valid for a wide range of generalized random graphs and distributions. For example for a Cayley tree – a graph with a fixed degree z and no loops – the criterion from (10) can be used. This yields the critical concentration $q_c = 1 - p_c = 1/(z-1)$, which is well known [16,17]. Another example is a random Erdös-Rényi (ER) graph. In those graphs edges are distributed randomly and the resulting degree distribution is Poissonian [20]. Applying the criterion from (4) to a Poisson distribution yields:

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\langle k \rangle^2 + \langle k \rangle}{\langle k \rangle} = 2, \tag{11}$$

which reduces to $\langle k \rangle = 1$ as known for ER graphs [20].

2.3 Generating Functions

A general method for studying the size of the infinite cluster and the residual network for a graph with an arbitrary degree distribution was first developed by Molloy and Reed [21]. They suggested viewing the infinite cluster as being explored and used differential equations for the number of un-exposed links and unvisited sites to find the size of the infinite cluster and the degree distribution of the residual graph (the finite clusters).

An alternative and very powerful derivation was given by Newman, Strogatz and Watts [5]. They have used the generating functions method to study the

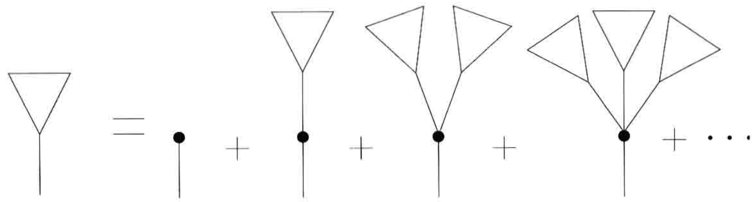

Fig. 1. An illustration of equations 14 and 15 for the probability to reach a branch of a given size by following a link. This is the sum of the probabilities to reach a vertex with zero outgoing links, of reaching a vertex with a single outgoing link connected to another such branch, of reaching a vertex with two outgoing links connected to two such branches etc. After Newman *et al.* [5].

size of the infinite cluster as well as other quantities (such as the diameter and cluster size distribution). They have also applied this method to other types of graphs (directed and bipartite). Here we closely follow their derivation in order to find the size of the infinite cluster and the critical exponents.

In [5] a generating function is built for the degree distribution:

$$G_0(x) = \sum_{k=0}^{\infty} P(k)x^k. \tag{12}$$

If we start from a randomly chosen site and follow each of its links to its nearest neighbors, the sites arrived will have a degree distribution $kP(k)/\langle k\rangle$ [19, 2,5,3]. The generating function describing the probability for k *outgoing* links (excluding the link we arrived along) will be:

$$G_1(x) = \frac{\sum kP(k)x^{k-1}}{\sum kP(k)} = \frac{d}{dx}G_0(x)/\langle k\rangle \; . \tag{13}$$

Let $H_1(x)$ be the generating function for the probability of reaching a branch of a given size by following a link[5]. If we denote the coefficients of $G_1(x)$ by q_k (i.e. the probability for k outgoing links from a site reached by following a random link), then $H_1(x)$ must satisfy the self-consistent equation (see Fig. 1):

$$H_1(x) = xq_0 + xq_1H_1(x) + xq_2[H_1(x)]^2 + \cdots \; . \tag{14}$$

Which can be written as:

$$H_1(x) = xG_1(H_1(x)) \; . \tag{15}$$

If we start from a random site, we have one such branch at the end of each neighboring link. Since $G_0(x)$ is the generating function for the degree of the site, the generating function for the probability of a site to belong to an n-site cluster is:

[5] We assume that the finite clusters have almost no loops and are therefore tree-like structures.