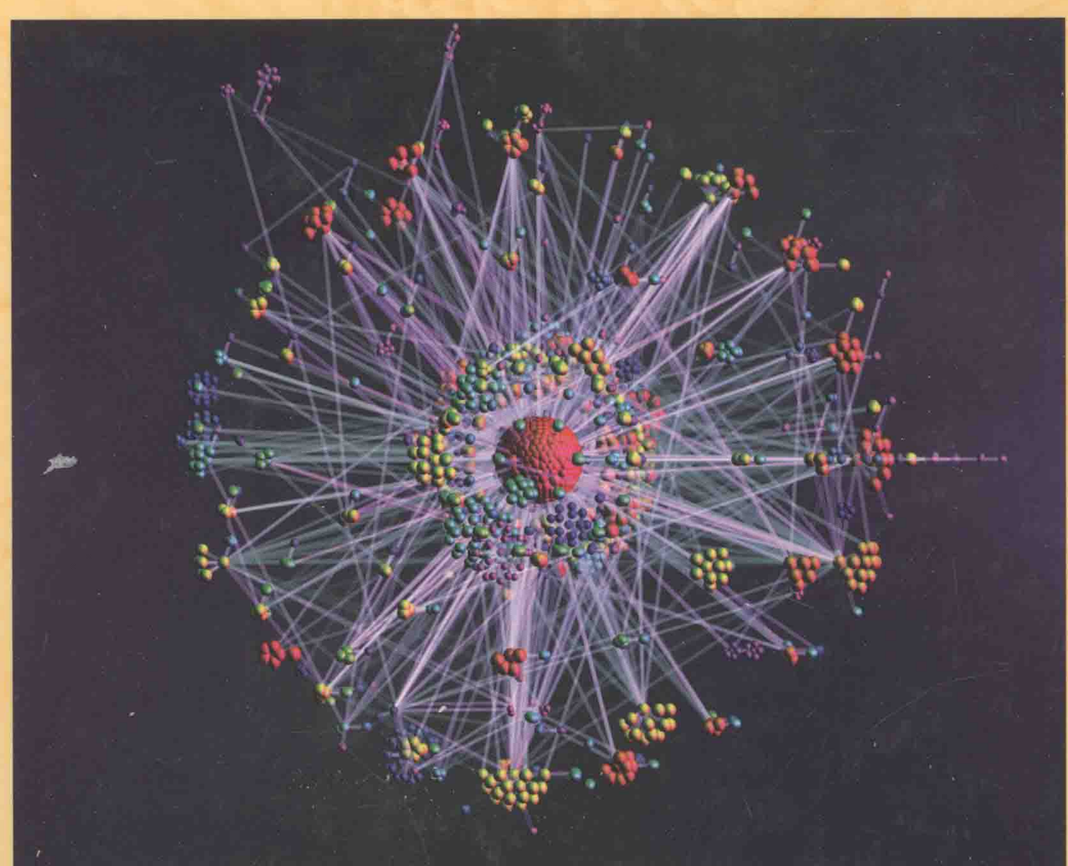
COMPLEX NETWORKS

Structure, Robustness and Function



Reuven Cohen and Shlomo Havlin

Complex Networks

Structure, Robustness and Function

REUVEN COHEN
Bar-Ilan University

SHLOMO HAVLIN
Bar-Ilan University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521841566

First published 2010

Printed in the United Kingdom at the University Press, Cambridge

A catalog record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Cohen, Reuven, 1974–
Complex networks : structure, robustness, and function / Reuven Cohen, Shlomo Havlin.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-521-84156-6 (hardback)
1. System analysis. I. Havlin, Shlomo. II. Title.
T57.6.C638 2010
003 – dc22 2010007255

ISBN 978-0-521-84156-6 Hardback

Complex Networks

Structure, Robustness and Function

Examining important results and analytical techniques, this graduate-level textbook is a step-by-step presentation of the structure and function of complex networks.

From the stability of the Internet to efficient methods of immunizing populations, from epidemic spreading to how to efficiently search for individuals, this textbook explains the theoretical methods used, and the experimental and analytical results obtained. Ideal for graduate students and researchers entering this field, it gives detailed derivations of many results in complex networks theory. End-of-chapter review questions help students monitor their understanding of the materials presented.

Reuven Cohen is a Senior Lecturer in the Department of Mathematics at Bar-Ilan University, Israel. He has written many papers in the fields of complex networks, robot swarms, algorithms and communication networks, and has won several national and international prizes for his work.

Shlomo Havlin is a Professor in the Department of Physics at Bar-Ilan University, Israel. He is an Editor of several physics journals, has published over 600 articles in international journals, co-authored and co-edited 11 books, and won numerous awards for his work including the Weizmann Prize (2009) and the APS Lilienfeld Prize (2010).

Contents

1 Introduction

Networks are present in almost every aspect of our life. The technological world surrounding us is full of networks. Communication networks consisting of telephones and cellular phones, the electrical power grid, computer communication networks, airline networks and, in particular, the world-wide Internet network are an important part of everyday life. The symbolic network of HTML pages and links – the World Wide Web (WWW) – is a virtual network that many of us use every day, and the list is long. Society is also networked. The network of friendship between individuals, working relations, or common hobbies, and the network of business relations between people and firms are examples of social and economic networks. Cities and countries are connected by road or airline networks. Epidemics spread in population networks. A great deal of interest has recently focused on biological networks representing the interactions between genes and proteins in our body. Ecological networks such as predator–prey networks are also under intensive study today. The physical world is also rich in network phenomena such as interactions between atoms in matter, between monomers in polymers, between grains in granular media, and the network of relations between similar configurations of proteins (i.e. between configurations that are in reach of each other by a simple move). Recently, studies have shown that polymer networks in real space can actually have a wide distribution of the branching factor, which is also similar to other real-world networks [ZKM$^+$03].

Graphs are used for describing mathematical concepts in networks. Graphs represent the essential topological properties of a network by treating the network as a collection of nodes and edges. For example, in computer networks, such as the Internet, computers can be represented by nodes, and the cables between them are represented by the edges. In the WWW the nodes are the HTML pages, and the edges represent the links between pages. This is a simple, yet powerful concept. Because of its simplicity it considers different complex systems, such as those described above, using the same mathematical tools and methods and, in many cases, the properties of the networks are similar.

Graph theory is rooted in the eighteenth century, beginning with the work of Euler, who is the father of the field of topology as well as many other fields in mathematics. The theory began with the famous problem of the bridges of Kőnigsberg, where people had been wondering for years whether all seven bridges connecting the different parts

Figure 1.1 The bridges of Kőnigsberg (after Wikipedia).

of the town could be traversed, without passing any of them twice (see Figure 1.1). The genius of Euler led him to the understanding that the only important factor in this problem is the topological network structure, and therefore it can be simplified into a graph traversal problem, containing nodes (parts of the city) and links (bridges). He then proceeded to solve the problem by concluding that to fulfill the requirement every node in the graph, except possibly the first and last nodes visited, should be connected by an even number of bridges (since it is entered and left the same number of times). In Kőnigsberg more than two nodes have an odd number of links, and therefore the bridges cannot be traversed by such a path, known thereafter as an Eulerian path.

This simple yet powerful argument shows the strength of graph theory, enabling deduction of properties of real-world systems using simplification in order to construct a very basic model. Studies of graph theory usually focus on the properties of special graphs or on extremal properties (finding graphs with extreme properties). However, the networks mentioned above are hardly appropriate for such research. They change over time, social links are created and broken, technological networks are changed daily by the addition of new nodes, as are the links between them. Biological networks change by evolutionary processes and by environmental processes. Even at a given time point, one cannot usually find the complete data for the network structure.

In the 1960s, two mathematicians, Paul Erdős and Alfred Rényi (ER), introduced a new concept that allows the treatment of such networks – random graph theory [ER59, ER60].[1] Their ingenious idea was to combine the concepts of graph theory with tools from probability theory and to consider families of graphs rather than specific graphs. Random graph theory is to graph theory what statistical mechanics is to Newtonian physics. The microscopic theory underlies the small-scale behavior, but when the entire ensemble is considered, new statistical concepts and collective behavior emerge.

The study of random graphs has led to ideas very similar to those of statistical physics. Since statistical physics deals with a system of many interacting atoms and molecules it is natural to assume that methods from this field will be useful for network study. Indeed, percolation, scaling, order parameters, renormalization, self-similarity, phase transitions, and critical exponents from statistical physics are all present in the field of random graphs, and are used in studying such networks.

At the end of the twentieth century, with the advancement of computers and the availability of large-scale data and the tools to analyze them, it became clear that the classical theory of random graphs fails to describe many real-world networks. The works of Barabási and Albert on the WWW [BA99] and of Faloutsos, Faloutsos, and Faloutsos on the Internet router network [FFF99] made clear that the link distribution of these and many other networks is not completely random, and it cannot be described by ER graph theory. These findings and others have led to a new, generalized form of random graph theory, taking into account some less trivial correlations found in real-world networks. These results explain several long-standing puzzles, for example, why viruses and worms are able to survive in the Internet for a very long time. Moreover, studying these new types of networks leads to novel physical laws, which arise owing to the new topology. If materials such as polymers can be constructed with a similar topology, it is expected that they will obey new and anomalous physical laws such as phase transitions, elasticity, and transport.

This book will focus on this modern theory of complex networks. Since thousands of papers, as well as several popular science [Bar03, Wat03] and scientific books [BBV08, DM03, PV03] have appeared on this subject in the last few years, it will be impossible to cover all existing works. In this book we have tried to focus on results concerning the structure of these networks and also partially cover works regarding the dynamics and applications. Since this is also intended as a textbook for students and scientists aiming to enter this growing field, we will attempt to present a detailed and clear description of the methods used in analyzing complex networks. This, we believe, will allow the reader to obtain further results in this growing field and to comprehend further literature on this subject.

[1] In fact, some of these ideas had been raised before, in particular, by Rapoport [Rap57]. However, only with the systematic works of Erdős and Rényi was much attention given to this subject.

The rest of this introduction will present some basic definitions and concepts from physics and mathematics. The main body of this book is divided into three parts. Part I will present results based on measurements in real-world networks, and will present several ensembles and growth models studied in this field. Part II will discuss the structural and robustness properties of complex networks. It will focus mainly on scale-free networks, which are thought to be most relevant for real-world systems, but in most cases this approach is also suitable for other types of random networks. Part III will discuss some dynamics regarding complex networks, and applications of the knowledge gained to real-world problems. The appendices will provide more technical details regarding probability theory, as well as algorithmic and simulation aspects.

1.1 Graph theory

A **graph** according to its mathematical definition is a pair of sets (V, E), where V is a set of vertices (the nodes of the graph), and E is a set of edges, denoting the links between the vertices. Each edge consists of a pair of vertices and can be regarded as similar to "bonds" in physical systems.

In a **directed graph** (also termed "digraph"), the edges are taken as ordered pairs, i.e., each edge is directed from the first to the second vertex of the pair.

A "**multigraph**" is a graph in which more than one edge is allowed between a pair of vertices and edges are also allowed to connect a vertex to itself. This is less restrictive than the notion of a graph, and therefore many of the networks studied in this work will actually be multigraphs.

A graph is represented frequently by an **adjacency matrix**, A_{ij}, which is a matrix in which every row and column represents a vertex of the graph. The A_{ij} entry is 1 if a link exists between the ith and jth vertices, or 0 otherwise. In a directed graph, the matrix will, in general, be asymmetrical. In a multigraph the entries can also be integers larger than 1, and the diagonal entries are not necessarily 0.

1.2 Scale-free processes and fractal structures

In statistical physics, it is well known that systems approaching a critical point in a phase transition develop a behavior that spans all length-scales of the system. Close to criticality, the correlations between physically remote regimes change from decaying exponentially with the distance, to a slow, power-law, decaying behavior.

This power-law phenomenon has no characteristic length-scale, and is therefore often termed "scale free." The reaction to external disturbances, for example, the susceptibility of the system, also diverges as a power law when approaching the critical point. Another situation where power laws and scale-free behavior appear is in self-organized criticality (SOC) [BTW87], where events such as earthquakes and forest fires tend to drive themselves into a criticality-like power-law behavior.

Power-law distributions have been studied in physics, particularly in the context of fractals and Lévy flights. Fractals are objects having no characteristic length-scale and appear similar (at least in a statistical sense) at every length-scale [BBV08, BH94, BH96, bH00, BLW94, Fed88, Man82]. Many natural objects, such as mountains, clouds, coastlines and rivers, as well as the cardiovascular and nervous systems are known to be fractals and are self-similar. This is why we find it hard to distinguish between a photograph of a mountain and part of the mountain; neither can we ascertain the altitude from which a picture of a coastline was taken. Diverse phenomena, such as the distribution of earthquakes, biological rhythms, and rates of transport of data packets in communication networks, are also known to possess a power-law distribution. They come in all sizes and rhythms, spanning many orders of magnitude [BH96].

Lévy flights were suggested by Paul Lévy [Lév25], who was studying what is now known as Lévy stable distributions. The question he asked was, when is the length distribution of a single step in a random walk similar to that of the entire walk? Besides the known result, that of the Gaussian distribution, Lévy found an entire new family – essentially that of scale-free distributions. Stable distributions do not obey the central limit theorem (stating that the sum of a large number of steps, having *finite variance*, tends to a Gaussian distribution [Fel68]), owing to the divergence of the variance of individual steps. Lévy walks have numerous applications [GHB08, HBG06, KSZ96, SK85, SZK93]. An interesting observation is that animal foraging patterns that follow Lévy stable distributions have been shown to be the most efficient strategy [Kle00a, VBH+99]. For recent reviews and books on complex networks and, in particular, scale-free networks, see [AB02, BBV08, BLM+06, DG08, DM02, DMS03, New02b, PV03].

PART I

RANDOM NETWORK MODELS

2 The Erdős–Rényi models

Before 1960, graph theory mainly dealt with the properties of specific individual graphs. In the 1960s, Paul Erdős and Alfred Rényi initiated a systematic study of random graphs [ER59, ER60, ER61]. Some results regarding random graphs were reported even earlier by Rapoport [Rap57]. Random graph theory is, in fact, not the study of individual graphs, but the study of a statistical ensemble of graphs (or, as mathematicians prefer to call it, a *probability space* of graphs). The ensemble is a class consisting of many different graphs, where each graph has a probability attached to it. A property studied is said to exist with probability P if the total probability of a graph in the ensemble possessing that property is P (or the total fraction of graphs in the ensemble that has this property is P). This approach allows the use of probability theory in conjunction with discrete mathematics for studying graph ensembles. A property is said to exist for a class of graphs if the fraction of graphs in the ensemble which does not have this property is of zero measure. This is usually termed as a property of **almost every (a.e.)** graph. Sometimes the terms "almost surely" or "with high probability" are also used (with the former usually taken to mean that the residual probability vanishes exponentially with the system size).

2.1 Erdős–Rényi graphs

Two well-studied graph ensembles are $G_{N,M}$ – the ensemble of all graphs having N vertices and M edges, and $G_{N,p}$ – the ensemble consisting of graphs with N vertices, where each possible edge is realized with probability p. These two families, initially studied by Erdős and Rényi, are known to be similar if $M = \binom{N}{2}p$, so long as p is not too close to 0 or 1 [Bol85]; they are referred to as ER graphs. These families are quite similar to the microcanonical and canonical ensembles studied in statistical physics.

Examples of other well-studied ensembles are the family of **regular graphs**, where all nodes have the same number of edges, $P(k) = \delta_{k,k_0}$, and the family of *unlabeled* graphs, where graphs that are isomorphic under permutations of their nodes are considered to be the same object.