OXFORD

Networks

An Introduction

M. E. J. Newman

Table of Contents

INDEX

Networks

An Introduction

M. E. J. Newman

University of Michigan
and
Santa Fe Institute

OXFORD

UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published in the United States
by Oxford University Press Inc., New York

First printed 2010

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data

Data available

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
CPI Antony Rowe, Chippenham, Wiltshire

ISBN 978-0-19-920665-0 (Hbk.)

1 3 5 7 9 10 8 6 4 2

PREFACE

The scientific study of networks, such as computer networks, biological networks, and social networks, is an interdisciplinary field that combines ideas from mathematics, physics, biology, computer science, the social sciences, and many other areas. The field has benefited enormously from the wide range of viewpoints brought to it by practitioners from so many different disciplines, but it has also suffered because human knowledge about networks is dispersed across the scientific community and researchers in one area often do not have ready access to discoveries made in another. The goal of this book is to bring our knowledge of networks together and present it in consistent language and notation, so that it becomes a coherent whole whose elements complement one another and in combination teach us more than any single element can alone.

The book is divided into five parts. Following a short introductory chapter, Part I describes the basic types of networks studied by present-day science and the empirical techniques used to determine their structure. Part II introduces the fundamental mathematical tools used in the study of networks as well as measures and statistics for quantifying network structure. Part III describes computer algorithms for the efficient analysis of network data, while Part IV describes mathematical models of network structure that can help us predict the behavior of networked systems and understand their formation and growth. Finally, Part V describes theories of processes taking place on networks, such as epidemics on social networks or search processes on computer networks.

The technical level of the presentation varies among the parts, Part I requiring virtually no mathematical knowledge for its comprehension, while Parts II and III require a grasp of linear algebra and calculus at the undergraduate level. Parts IV and V are mathematically more advanced and suitable for advanced undergraduates, postgraduates, and researchers working in the field. The book could thus be used as the basis of a taught course at more than one level. A less technical course suitable for those with moderate mathematical knowledge might cover the material of Chapters 1 to 8, while a more technical course for advanced students might cover the material of Chapters 6 to 14 and selected material thereafter. Each chapter from Part II onward is accompanied by a selection of exercises that can be used to test the reader's understanding of the material.

This book has been some years in the making and many people have helped me with it during that time. I must thank my ever-patient editor Sonke Adlung, with whom I have worked on various book projects for more than 15 years now, and whose constant encouragement and kind words have made working with him and Oxford University Press a real pleasure. Thanks are also due to Melanie Johnstone, Alison Lees, Emma Lonie, and April Warman for their help with the final stages of bringing the book to print.

I have benefited greatly during the writing of this book from the conversation, comments, suggestions, and encouragement of many colleagues and friends. They are, sadly, too numerous to mention exhaustively, but special thanks must go to Steve Borgatti, Duncan Callaway, Aaron Clauset, Betsy Foxman, Linton Freeman, Michelle Girvan, Martin Gould, Mark Handcock, Petter Holme, Jon Kleinberg, Alden Klovdahl, Liza Levina, Lauren Meyers, Cris Moore, Lou Pecora, Mason Porter, Sidney Redner, Puck Rombach, Cosma Shalizi, Steve Strogatz, Duncan Watts, Doug White, Lenka Zdeborova, and Bob Ziff, as well as to the many students, particularly Michelle Adan, Alejandro Balbin, Chris Fink, Ruthi Hortsch, and Jane Wang, whose feedback helped iron out a lot of rough spots. I would also especially like to thank Brian Karrer, who read the entire book in draft form and gave me many pages of thoughtful and thought-provoking comments, as well as spotting a number of mistakes and typos. Responsibility for any remaining mistakes in the book of course rests entirely with myself, and I welcome corrections from readers.

Finally, my profound thanks go to my wife Carrie for her continual encouragement and support during the writing of this book. Without her the book would still have been written but I would have smiled a lot less.

Mark Newman
Ann Arbor, Michigan
February 24, 2010

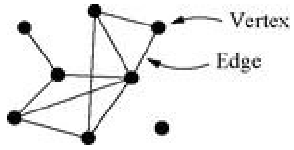
CHAPTER 1

INTRODUCTION

A short introduction to networks and why we study them

A NETWORK is, in its simplest form, a collection of points joined together in pairs by lines. In the jargon of the field the points are referred to as *vertices*[1] or *nodes* and the lines are referred to as *edges*. Many objects of interest in the physical, biological, and social sciences can be thought of as networks and, as this book aims to show, thinking of them in this way can often lead to new and useful insights.

We begin, in this introductory chapter, with a discussion of why we are interested in networks and a brief description of some specific networks of note. All the topics in this chapter are covered in greater depth elsewhere in the book.



A small network composed of eight vertices and ten edges.

WHY ARE WE INTERESTED IN NETWORKS?

There are many systems of interest to scientists that are composed of individual parts or components linked together in some way. Examples include the Internet, a collection of computers linked by data connections, and human societies, which are collections of people linked by acquaintance or social interaction.

Many aspects of these systems are worthy of study. Some people study the nature of the individual components—how a computer works, for instance, or how a human being feels or acts—while others study the nature of the connections or interactions—the communication protocols used on the Internet or the dynamics of human friendships. But there is a third aspect to these interacting systems, sometimes neglected but almost always crucial to the behavior of the system, which is the *pattern* of connections between components.

The pattern of connections in a given system can be represented as a network, the components of the system being the network vertices and the connections the edges. Upon reflection it should come as no surprise (although in some fields it is a relatively recent realization) that the structure of such networks, the particular pattern of interactions, can have a big effect on the behavior of the system. The pattern of connections between computers on the Internet, for instance, affects the routes that data take over the network and the efficiency with which the network transports those data. The connections in a social network affect how people learn, form opinions, and gather news, as well as affecting other less obvious phenomena, such as the spread of disease. Unless we know something about the structure of these networks, we cannot hope to understand fully how the corresponding systems work.

A network is a simplified representation that reduces a system to an abstract structure capturing only the basics of connection patterns and little else. Vertices and edges in a network can be labeled with additional information, such as names or strengths, to capture more details of the system, but even so a lot of information is usually lost in the process of reducing a full system to a network representation. This certainly has its disadvantages but it has advantages as well.

The most common network variants are discussed in detail in Chapter 6.

Scientists in a wide variety of fields have, over the years, developed an extensive set of tools—mathematical, computational, and statistical—for analyzing, modeling, and understanding networks. Many of these tools start from a simple network representation, a set of vertices and edges, and after suitable calculations tell you something about the network that might well be useful to you: which is the best connected vertex, say, or the length of a path from one vertex to another. Other tools take the form of network models that can make mathematical predictions about processes taking place on networks, such as the way traffic will flow over the Internet or the way a disease will spread through a community. Because they work with networks in their abstract form, these tools can in theory be applied to almost any system represented as a network. Thus if there is a system you are interested in, and it can usefully be represented as a network, then there are hundreds of different tools out there, already developed and well understood, that you can immediately apply to the analysis of your system. Certainly not all of them will give useful results—which measurements or calculations are useful for a particular system depends on what the system is and does and on what specific questions you are trying to answer about it. Still, if you have a well-posed question about a networked system there will, in many cases, already be a tool available that will help you address it.

Networks are thus a general yet powerful means of representing patterns of connections or interactions between the parts of a system. In this book, we discuss many examples of specific networks in different fields, along with techniques for their analysis drawn from mathematics, physics, the computer and information sciences, the social sciences, biology, and elsewhere. In doing so, we bring together a wide range of ideas and expertise from many disciplines to give a comprehensive introduction to the science of networks.

SOME EXAMPLES OF NETWORKS

One of the best known and most widely studied examples of a network is the Internet, the computer data network in which the vertices are computers and the edges are physical data connections between them, such as optical fiber cables or telephone lines. Figure 1.1 shows a picture of the structure of the Internet, a snapshot of the network as it was in 2003, reconstructed by observing the paths taken across the network by a large number of Internet data packets traveling between different sources and destinations. It is a curious fact that although the Internet is a man-made and carefully engineered network we don't know exactly what its structure is, since it was built by many different groups of people with only limited knowledge of each other's actions and little centralized control. Our best current data on its structure are derived from experimental studies, such as the one that produced this figure, rather than from any central repository of knowledge or coordinating authority.

We look at the Internet in more detail in Section 2.1.

There are a number of excellent practical reasons why we might want to study the network structure of the Internet. The function of the Internet is to transport data between computers (and other devices) in different parts of the world, which it does by dividing the data into pieces or *packets* and shipping them from vertex to vertex across the network until they reach their intended destination. Certainly the structure of the network will affect how efficiently it accomplishes this function and if we know the network structure we can address many questions of practical relevance. How should we choose the route by which data are transported? Is the shortest route always necessarily the fastest? If not, then what is, and how can we find it? How can we avoid bottlenecks in the traffic flow that might slow things down? What happens when a vertex or an edge fails (which they do with some regularity)? How can we devise schemes to route around such failures? If we have the opportunity to add new capacity to the network, where should it be added?

Knowledge of Internet structure also plays a central role in the development of new communications standards. New standards and protocols are continually being devised for communication over the Internet, and old ones are revised. The parameters of these protocols are tuned for optimal performance with the structure of the Internet in mind. In the early days of the network, rather primitive models of network structure were employed in the tuning process, but as better structural data become available it becomes possible to better understand and improve performance.

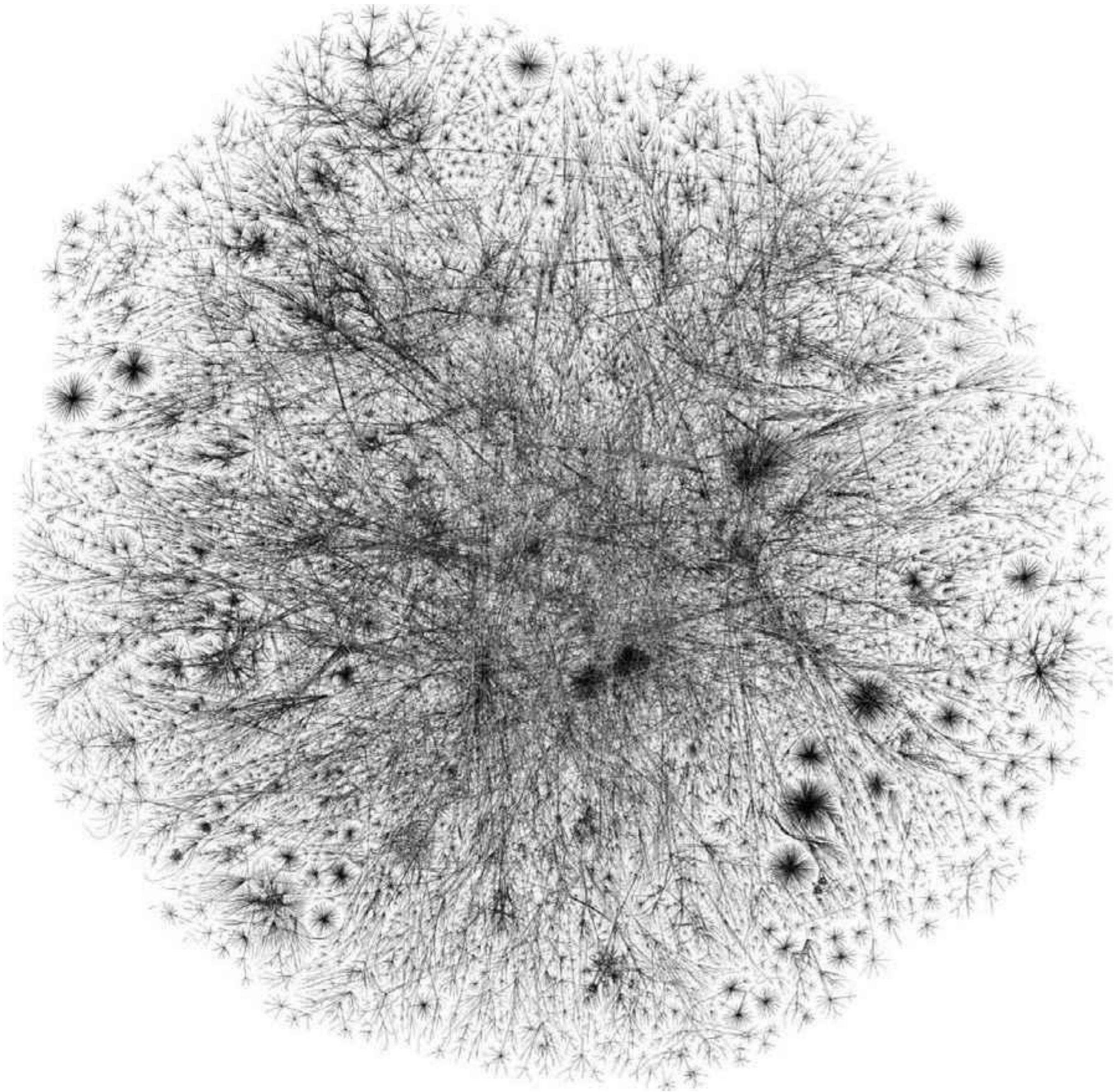


Figure 1.1: The network structure of the Internet. (See Plate I for color version.) The vertices in this representation of the Internet are "class C subnets"—groups of computers with similar Internet addresses that are usually under the management of a single organization—and the connections between them represent the routes taken by Internet data packets as they hop between subnets. The geometric positions of the vertices in the picture have no special meaning; they are chosen simply to give a pleasing layout and are not related, for instance, to geographic position of the vertices. The structure of the Internet is discussed in detail in Section 2.1. Figure created by the Opte Project (www.opte.org). Reproduced with permission.

A more abstract example of a network is the World Wide Web. In common parlance the words "Web" and "Internet" are often used interchangeably, but technically the two are quite distinct. The Internet is a physical network of computers linked by actual cables (or sometimes radio links) running between them. The Web, on the other hand, is a network of information stored on web pages. The vertices of the World Wide Web are web pages and the edges are “hyperlinks,” the highlighted snippets of text or push-buttons on web pages that we click on to navigate from one page to another. A hyperlink is purely a software construct; you can link from your web page to a page that lives on a computer on the other side of the world just as easily as you can link to a friend down the hall. There is no physical structure, like an optical fiber, that needs to be built when you make a new link. The link is merely an address that tells the computer where to look next when you click on it.

The World Wide Web is discussed in more detail in Section 4.1.

Abstract though it may be, the World Wide Web, with its billions of pages and links, has proved enormously useful, not to mention profitable, to many people, and the structure of the network of links is of substantial interest. Since people tend to add hyperlinks between pages with related content, the link structure of the Web reveals something about the content structure. What’s more, people tend to link more often to pages that they find useful than to those they do not, so that the number of links pointing to a page can be used as a measure of its usefulness. A more sophisticated version of this idea lies behind the operation of the popular Web search engine *Google*, as well as some others.

The mechanics of Web search are discussed in Section 19.1.

The Web also illustrates another concept of network theory, the *directed network* . Hyperlinks on the Web run in one specific direction, from one web page to another. Given an appropriate link on page A, you can click and arrive at page B. But there is no requirement that B contains a link back to A again. (It may contain such a link, but there is no law that says that it must and much of the time it will not.) One says that the edges in the World Wide Web are *directed*, running from the linking page to the linked.

Social networks are discussed in more depth in Chapter 3.

Moving away from the technological realm, another type of network of scientific interest is the social network. A social network is, usually, a network of people, although it may sometimes be a network of groups of people, such as companies. The people or groups form the vertices of the network and the edges represent connections of some kind between them, such as friendship between individuals or business relationships between companies. The field of sociology has perhaps the longest and best developed tradition of the empirical study of networks as they occur in the real world, and many of the mathematical and statistical tools that are used in the study of networks are borrowed, directly or indirectly, from sociologists.

Figure 1.2 shows a famous example of a social network from the sociology literature, Wayne Zachary’s “karate club” network. This network represents the pattern of friendships among members of a karate club at a north American university. The network was constructed by direct observation of interactions between the club’s members. As is typical of such studies the network is small, having, in this case, only 34 vertices. Network representations of the Internet or the World Wide Web, by contrast, can have thousands or millions of vertices. In principle there is no reason why social networks cannot be similarly large. The entire population of the world, for example, can be regarded as a very large social network. But in practice social network data are limited to relatively small groups because of the effort involved in compiling them. The network of Fig. 1.2, for instance, was the product of two years of observations by one experimenter. In recent years a few larger social networks have been constructed by dint of enormous effort on the part of large groups of researchers. And online social networking services, such as Facebook or instant message "buddy lists," can provide network data on a previously unreachable scale. Studies are just beginning to emerge of the structure and properties of these larger networks.

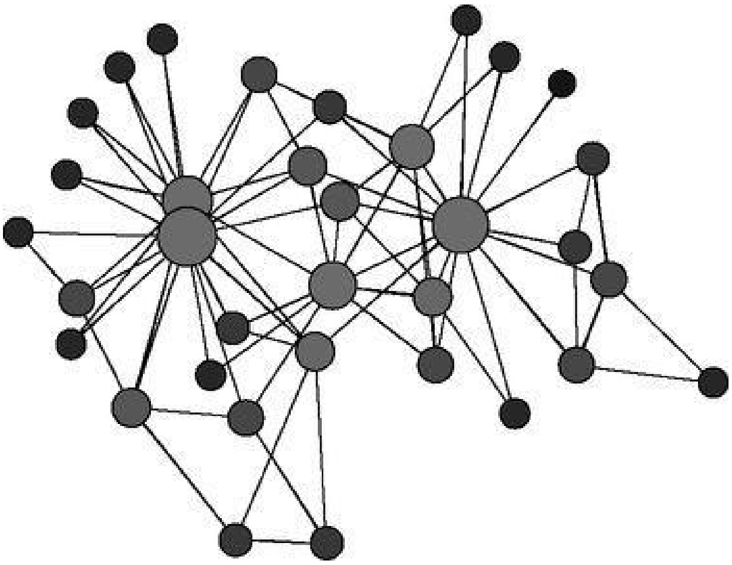


Figure 1.2: Friendship network between members of a club. This social network from a study conducted in the 1970s shows the pattern of friendships between the members of a karate club at an American university. The data were collected and published by Zachary [334].

Neural networks are discussed in Section 5.2 and food webs in Section 5.3

A third realm in which networks have become important in recent years is biology. Networks occur in a number of situations in biology. Some are concrete physical networks like neural networks—the networks of connections between neurons in the brain—while others are more abstract. In Fig. 1.3 we show a picture of a "food web," an ecological network in which the vertices are species in an ecosystem and the edges represent predator-prey relationships between them. That is, pairs of species are connected by edges in this network if one species eats the other. The study of food