

# SOCIAL AND ECONOMIC NETWORKS

Matthew O. Jackson

A network graph is overlaid on the cover. It consists of numerous orange circular nodes connected by thin black lines. The nodes are scattered across the cover, with a higher density in the lower half. The connections form a complex web of paths and clusters, with some nodes having multiple connections and others being isolated or part of small groups.

**AND**

---

**ECONOMIC  
NETWORKS**

---

---

**Matthew O. Jackson**

PRINCETON UNIVERSITY PRESS  
Princeton and Oxford

Requests for permission to reproduce material from this work  
should be sent to Permissions, Princeton University Press

Published by Princeton University Press, 41 William Street, Princeton,  
New Jersey 08540

In the United Kingdom: Princeton University Press, 6 Oxford Street, Woodstock,  
Oxfordshire OX20 1TW


All Rights Reserved

ISBN-13: 978-0-691-13440-6

Library of Congress Control Number: 2008926162

British Library Cataloging-in-Publication Data is available

This book was composed in Times Roman and Slimbach using ZzTEX  
by Princeton Editorial Associates, Inc., Scottsdale, Arizona.

Printed on acid-free paper. 

[www.press.princeton.edu](http://www.press.princeton.edu)

Printed in the United States of America

1 3 5 7 9 10 8 6 4 2

---

To Sara, Emily, and Lisa

---

---

# Contents

[Preface](#)

## **PART I ■ BACKGROUND AND FUNDAMENTALS OF NETWORK ANALYSIS**

### **Chapter 1 Introduction**

[1.1 Why Model Networks?](#)

[1.2 A Set of Examples](#)

[1.3 Exercises](#)

### **Chapter 2 Representing and Measuring Networks**

[2.1 Representing Networks](#)

[2.2 Some Summary Statistics and Characteristics of Networks](#)

[2.3 Appendix: Basic Graph Theory](#)

[2.4 Appendix: Eigenvectors and Eigenvalues](#)

[2.5 Exercises](#)

### **Chapter 3 Empirical Background on Social and Economic Networks**

[3.1 The Prevalence of Social Networks](#)

[3.2 Observations on the Structure of Networks](#)

## **PART II ■ MODELS OF NETWORK FORMATION**

### **Chapter 4 Random-Graph Models of Networks**

[4.1 Static Random-Graph Models of Random Networks](#)

[4.2 Properties of Random Networks](#)

## [4.3 An Application: Contagion and Diffusion](#)

---

### [4.4 Distribution of Component Sizes](#)

### [4.5 Appendix: Useful Facts, Tools, and Theorems](#)

### [4.6 Exercises](#)

## **[Chapter 5 Growing Random Networks](#)**

### [5.1 Uniform Randomness: An Exponential Degree Distribution](#)

### [5.2 Preferential Attachment](#)

### [5.3 Hybrid Models](#)

### [5.4 Small Worlds, Clustering, and Assortativity](#)

### [5.5 Exercises](#)

## **[Chapter 6 Strategic Network Formation](#)**

### [6.1 Pairwise Stability](#)

### [6.2 Efficient Networks](#)

### [6.3 Distance-Based Utility](#)

### [6.4 A Coauthor Model and Negative Externalities](#)

### [6.5 Small Worlds in an Islands-Connections Model](#)

### [6.6 A General Tension between Stability and Efficiency](#)

### [6.7 Exercises](#)

## **[PART III ■ IMPLICATIONS OF NETWORK STRUCTURE](#)**

## **[Chapter 7 Diffusion through Networks](#)**

### [7.1 Background: The Bass Model](#)

### [7.2 Spread of Information and Disease](#)

## [7.3 Search and Navigation on Networks](#)

---

### [7.4 Exercises](#)

## [\*\*Chapter 8 Learning and Networks\*\*](#)

### [8.1 Early Theory and Opinion Leaders](#)

### [8.2 Bayesian and Observational Learning](#)

### [8.3 Imitation and Social Influence Models: The DeGroot Model](#)

### [8.4 Exercises](#)

## [\*\*Chapter 9 Decisions, Behavior, and Games on Networks\*\*](#)

### [9.1 Decisions and Social Interaction](#)

### [9.2 Graphical Games](#)

### [9.3 Semi-Anonymous Graphical Games](#)

### [9.4 Randomly Chosen Neighbors and Network Games](#)

### [9.5 Richer Action Spaces](#)

### [9.6 Dynamic Behavior and Contagion](#)

### [9.7 Multiple Equilibria and Diffusion in Network Games](#)

### [9.8 Computing Equilibria](#)

### [9.9 Appendix: A Primer on Noncooperative Game Theory](#)

### [9.10 Exercises](#)

## [\*\*Chapter 10 Networked Markets\*\*](#)

### [10.1 Social Embeddedness of Markets and Exchange](#)

### [10.2 Networks in Labor Markets](#)

### [10.3 Models of Networked Markets](#)

## 10.4 Concluding Remarks

---

## 10.5 Exercises

# **PART IV ■ METHODS, TOOLS, AND EMPIRICAL ANALYSES**

## **Chapter 11 Game-Theoretic Modeling of Network Formation**

### 11.1 Defining Stability and Equilibrium

### 11.2 The Existence of Stable Networks

### 11.3 Directed Networks

### 11.4 Stochastic Strategic Models of Network Formation

### 11.5 Farsighted Network Formation

### 11.6 Transfers and Network Formation

### 11.7 Weighted Network Formation

### 11.8 Agent-Based Modeling

### 11.9 Exercises

## **Chapter 12 Allocation Rules, Networks, and Cooperative Games**

### 12.1 Cooperative Game Theory

### 12.2 Communication Games

### 12.3 Networks and Allocation Rules

### 12.4 Allocation Rules When Networks Are Formed

### 12.5 Concluding Remarks

### 12.6 Exercises

## **Chapter 13 Observing and Measuring Social Interaction**

### 13.1 Specification and Identification



## [13.2 Community Structures, Block Models, and Latent Spaces](#)

---

### [13.3 Exercises](#)

[Afterword](#)

[Bibliography](#)

[Index](#)

---

---

# Preface

This book provides an overview and synthesis of models and techniques for analyzing social and economic networks. It is meant to serve as both a resource for researchers and a text on the subject for graduate students. The focus is primarily on the modeling of and theory behind the structure, formation, and implications of social networks. Statistical and experimental analyses of networks are also discussed throughout, especially when they help set the stage for issues to be investigated. The main emphasis is on providing a foundation for analyzing and understanding social and economic networks.

The book is organized into four parts. The first part introduces network analysis and provides some background on various networks, how they are measured, and useful ways of representing them. The second part presents some of the models that have been used to understand how networks are formed. This part draws on two very different perspectives: random-graph models, for which there is some stochastic process that governs the development of the links in a network, and strategic models of network formation, in which the development of links is based on costs and benefits, and game-theoretic techniques are used. These approaches to modeling provide different insights into networks, how they are formed, and why they exhibit certain characteristics. The third part of the book looks at the implications of network structure. Much of the interest in networks is due to the fact that their structure is an important determinant of how societies and economies function. This part examines how network models are used to predict the spread of disease, the dissemination of information, the behavior of people, and how markets function. The final part of the book covers empirical analyses of networks and methods of identifying social interaction.

Although the table of contents represents a categorization of the chapters by subject, the relationship between the chapters is not entirely linear. I have intermingled some subjects to tie together different approaches, and the chapters are cross referenced. There is also a progression through the book, with some of the more technically demanding chapters coming later, as well as those that draw on concepts from earlier chapters.

The modeling of networks requires some mathematical background, but I have made the book as self-contained as possible. I do not presume any knowledge beyond some familiarity with linear algebra (vectors and matrices), calculus, and probability and statistics. The discussions employing graph theory and game theory are self-contained, and there are appendixes with introductions to some of the topics, including various useful results from graph theory, math, game theory, and probability theory. Those sections and exercises that are more mathematical in nature are marked with an asterisk (\*).

I had several reasons for writing this book. First and foremost, networks of relationships play central roles in a wide variety of social, economic, and political interactions. For example, many, if not most, markets do not function as centralized and anonymous institutions, but rather involve a variety of bilateral exchanges or contracts. As a case in point, most jobs are filled by people who were informed about the job through a social contact. This fact has consequences for patterns of employment, inequality in wages across groups, and social mobility. Understanding social network structure and how it influences human interaction is not merely important to science (and the social sciences in particular), it is essential. Second, the topic is timely. Recent technological advances

have made information networks much more prominent (e.g., the world wide web), and people are more conscious of the role of networks in their lives. In addition, the formal modeling of networks has now reached a maturity across fields that permits a book-length treatment. A third motivation for writing this work is the inter- and multidisciplinary nature of research on networks: the pertinent knowledge is diffuse, and there is much to be gained by collecting aspects of it from different fields in a unified treatment. Substantial research on networks has been conducted in sociology, economics, physics, mathematics, and computer science, and these disciplines take different approaches and ask varied questions.<sup>1</sup> Thus it is important to bridge the literatures and produce a text that collects and synthesizes different modeling approaches and techniques and makes them all available to researchers from any discipline who are interested in the study of networks.

At the end of each chapter in this book you will find exercises. The exercises are meant to serve several purposes. They serve the usual purpose of problems in a textbook: to give students a chance to work with concepts and more fully familiarize themselves with the ideas presented in the chapter, thus the exercises can be made part of courses. And the interested researcher can work the exercises as well. But the exercises also introduce *new* material. I have used them to introduce new concepts not covered in the text. These are meant to be closely related to material in the text but complementary to it. Often the ideas were important enough to include in the book, but did not fit easily with the main thread of a chapter without making it longer than I desired or taking us off on a tangent. Thus researchers consulting this book as a reference should not ignore the exercises and in many instances may actually find what they are looking for in them.<sup>2</sup>

As with any such undertaking, there are many acknowledgments due, and they do not adequately represent the scope and depth of the help received. This project would not have been possible without financial support from the Center for Advanced Studies in the Behavioral Sciences, the Guggenheim Foundation, and the Lee Center for Advanced Networking, as well as the National Science Foundation under grants SES-0316493 and SES-0647867. I began this project while I was at the California Institute of Technology and concluded it while at Stanford University, and the support of both institutions is gratefully acknowledged.

As for the content of this monograph, I have been deeply influenced by a number of collaborators. My initial interest in this subject arose through conversations and subsequent research with Ashley Wolinsky. I have continued to learn about networks and have enjoyed interaction with a group of coauthors (in chronological order): Alison Watts, Bhaskar Dutta, Anne van den Nouweland, Toni Calvó-Armengol, Francis Bloch, Gary Charness, Alan Kirman, Jernej Copic, Brian Rogers, Duny Lopez-Pintado, Leeat Yariv, Andrea Galeotti, Sanjeev Goyal, Fernando Vega-Redondo, Ben Golub, Sergio Currarini, and Paolo Pin. Their collaboration and friendship are greatly appreciated.

Although not coauthors on network-related projects, Salvador Barbera, Darrell Duffie, and Hugo Sonnenschein have been great mentors (and friends) and have profoundly shaped my approach to writing. I thank Lada Adamic, Marika Cabral, Toni Calvó-Armengol, Jon Eguia, Marcel Fafchamps, Ben Golub, Carlos Lever, Laurent Mathevet, Tomas Rodriguez Barraquer, Tim Sullivan, and Cy Westmoreland for extensive comments on earlier drafts.

For all the emotional support and enthusiasm needed to keep such a project afloat, I owe profound thanks to my wife, Sara, my daughters, Emily and Lisa, and my parents. Special thanks are due Sara, whose encouragement through the persistent asking of the question “Did you get to work on your book today?” always kept me pointed in the right direction, and whose juggling of many tasks allowed me to answer “yes” more often than “no.”

<sup>1</sup> Social network analysis is a central and well-developed area of study in sociology, with societies, journals, conferences, and decades of research devoted to it. With occasional overlap, a literature on graph theory has matured in mathematics during the

same period. While the literature on networks has been thriving in sociology for more than five decades, it has emerged in economics primarily over the past 10 to 15 years. Its explosion in computer science and statistical physics has been rapid and has taken place mostly during the past decade.

[2](#). This book is obviously not the first to do this, as one sees important results appearing as exercises in many mathematical texts. The usefulness of this technique was made obvious to me through the superb text on axiomatic social choice by Hervé Moulin [492].

# **PART I**

---

---

## **BACKGROUND AND FUNDAMENTALS OF NETWORK ANALYSIS**

## Introduction

This chapter introduces the analysis of networks by presenting several examples of research. The examples provide some idea not only of why the subject is interesting but also of the range of networks studied, approaches taken, and methods used.

### 1.1 ■ Why Model Networks?

Social networks permeate our social and economic lives. They play a central role in the transmission of information about job opportunities and are critical to the trade of many goods and services. They are the basis for the provision of mutual insurance in developing countries. Social networks are also important in determining how diseases spread, which products we buy, which languages we speak, how we vote, as well as whether we become criminals, how much education we obtain, and our likelihood of succeeding professionally. The countless ways in which network structures affect our well-being make it critical to understand (1) how social network structures affect behavior and (2) which network structures are likely to emerge in a society. The purpose of this monograph is to provide a framework for an analysis of social networks, with an eye on these two questions.

As the modeling of networks comes from varied fields and employs a variety of different techniques, before jumping into formal definitions and models, it is useful to start with a few examples that help give some impression of what social networks are and how they have been modeled. The following examples illustrate widely differing perspectives, issues, and approaches, previewing some of the breadth of the range of topics to follow.

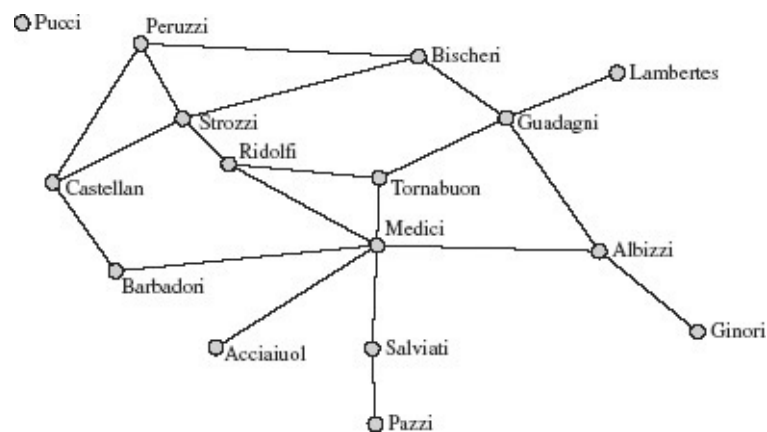


FIGURE 1.1 Network showing fifteenth-century Florentine marriages. Data from Padgett and Ansell [516] (drawn using UCINET).

### 1.2 ■ A Set of Examples

#### 1.2.1 Florentine Marriages

The first example is a detailed look at the role of social networks in the rise of the Medici in Florence during the 1400s. The Medici have been called the “godfathers of the Renaissance.” The accumulation of power in the early fifteenth century in Florence was orchestrated by Cosimo de’ Medici even though his family started with less wealth and political clout than other families in the oligarchy that ruled Florence at the time. Cosimo consolidated political and economic power by leveraging the central position of the Medici in networks of family intermarriages, economic relationships, and political patronage. His understanding of and fortuitous position in these social networks enabled him to build and control an early forerunner to a political party, while other important families of the time floundered in response.

Padgett and Ansell [516] provide powerful evidence for this consolidation by documenting the network of marriages between some key families in Florence in the 1430s. [Figure 1.1](#) shows the links between the key families in Florence at that time, where a link represents a marriage between members of two families.<sup>1</sup>

During this time the Medici (with Cosimo de’ Medici playing the key role) rose in power and largely consolidated control of business and politics in Florence. Previously Florence had been ruled by an oligarchy of elite families. If one examines wealth and political clout, however, the Medici do not stand out at this time and so one has to look at the structure of social relationships to understand why the Medici rose in power. For instance, the Strozzi had both greater wealth and more seats in the local legislature, and yet the Medici rose to eclipse them. The key to understanding the family’s rise, as Padgett and Ansell [516] detail, can be seen in the network structure.

If we do a rough calculation of importance in the network, simply by counting how many families a given family is linked to through marriages, then the Medici do come out on top. However, they only edge out the next highest families, the Strozzi and the Guadagni, by a ratio of 3 to 2. Although suggestive, it is not so dramatic as to be telling. We need to look a bit closer at the network structure to get a better handle on a key to the success of the Medici. In particular, the following measure of betweenness is illuminating.

Let  $P(ij)$  denote the number of shortest paths connecting family  $i$  to family  $j$ .<sup>2</sup> Let  $P_k(ij)$  denote the number of these paths that include family  $k$ . For instance, in [Figure 1.1](#) the shortest path between the Barbadori and Guadagni has three links in it. There are two such paths: Barbadori-Medici-Albizzi-Guadagni and Barbadori-Medici-Tornabuon-Guadagni. If we set  $i = \text{Barbadori}$  and  $j = \text{Guadagni}$ , then  $P(ij) = 2$ . As the Medici lie on both paths,  $P_k(ij) = 2$  when we set  $k = \text{Medici}$ , and  $P_k(ij) = 1$  if  $k = \text{Albizzi}$  or  $k = \text{Tornabuon}$ . In contrast this number is 0 if  $k = \text{Strozzi}$ , and is 1 if  $k = \text{Albizzi}$ . Thus, in a sense, the Medici are the key family in connecting the Barbadori to the Guadagni.

To gain intuition about how central a family is, look at an average of this betweenness calculation. We can ask, for each pair of other families, on what fraction of the total number of shortest paths between the two the given family lies. This number is 1 for the fraction of the shortest paths that the Medici lie on between the Barbadori and Guadagni, and 1/2 for the corresponding fraction that the Albizzi lie on. Averaging across all pairs of other families gives a betweenness or power measure (due to Freeman [255]) for a given family. In particular, we can calculate

$$\sum_{ij \neq j, k \notin \{i, j\}} \frac{P_k(ij)/P(ij)}{(n-1)(n-2)/2} \quad (1.1)$$

for each family  $k$ , where  $\frac{P_k(ij)}{P(ij)} = 0$  if there are no paths connecting  $i$  and  $j$ , and the denominator captures that a given family could lie on paths between as many as  $(n-1)(n-2)/2$  pairs of other families. This measure of betweenness for the Medici is .522. Thus if we look at all the shortest paths

between various families (other than the Medici) in this network, the Medici lie on more than half of them! In contrast, a similar calculation for the Strozzi yields .103, or just over 10 percent. The second-highest family in terms of betweenness after the Medici is the Guadagni with a betweenness of .255. To the extent that marriage relationships were keys to communicating information, brokering business deals, and reaching political decisions, the Medici were much better positioned than other families, at least according to this notion of betweenness.<sup>3</sup> While aided by circumstances (for instance, fiscal problems resulting from wars), it was the Medici and not some other family that ended up consolidating power. As Padgett and Ansell [516, p. 1259] put it, “Medician political control was produced by network disjunctures within the elite, which the Medici alone spanned.”

This analysis shows that network structure can provide important insights beyond those found in other political and economic characteristics. The example also illustrates that the network structure is important for more than a simple count of how many social ties each member has and suggests that different measures of betweenness or centrality will capture different aspects of network structure.

This example also suggests other questions that are addressed throughout this book. For instance, was it simply by chance that the Medici came to have such a special position in the network, or was it by choice and careful planning? As Padgett and Ansell [516, footnote 13] state, “The modern reader may need reminding that all of the elite marriages recorded here were arranged by patriarchy (or their equivalents) in the two families. Intra-elite marriages were conceived of partially in political alliance terms.” With this perspective in mind we then might ask why other families did not form more ties or try to circumvent the central position of the Medici. We could also ask whether the resulting network was optimal from a variety of perspectives: from the Medici’s perspective, from the oligarchs’ perspective, and from the perspective of the functioning of local politics and the economy of fifteenth-century Florence. We can begin to answer these types of questions through explicit models of the costs and benefits of networks, as well as models of how networks form.

### 1.2.2 Friendships and Romances among High School Students

The next example comes from the the National Longitudinal Adolescent Health Data Set, known as “Add Health.”<sup>4</sup> These data provide detailed social network information for more than 90,000 students from U.S. high schools interviewed during the mid-1990s, together with various data on the students’ socioeconomic background, behaviors, and opinions. The data provide insights and illustrate some features of networks that are discussed in more detail in the coming chapters.

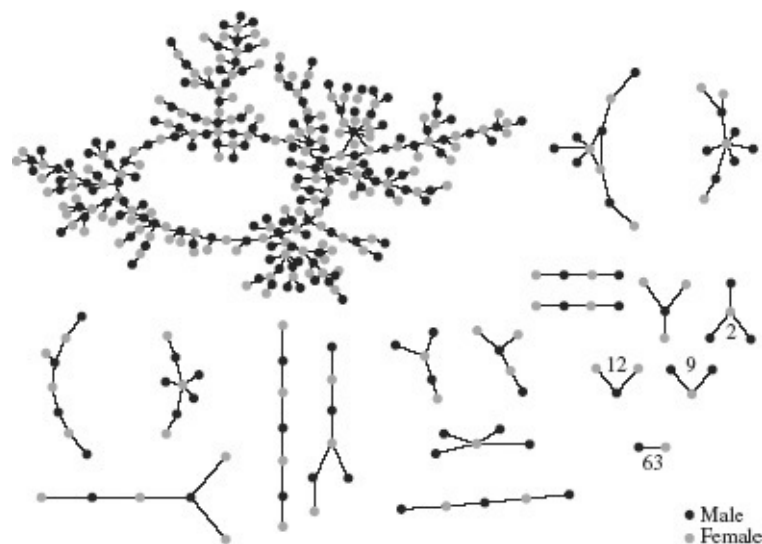




FIGURE 1.2 A network based on the Add Health data set. A link denotes a romantic relationship, and the numbers by some nodes indicate how many such components appear. Figure from Bearman, Moody, and Stovel [51].

Figure 1.2 shows a network of romantic relationships as found through surveys of students in one of the high schools in the study. The students were asked to list the romantic liaisons that they had during the six months previous to the survey.

The network shown in Figure 1.2 is nearly a *bipartite* network, meaning that the nodes can be divided into two groups, male and female, so that links only lie between groups (with a few exceptions). Despite its nearly bipartite nature, the distribution of the degrees of the nodes (number of links each node has) turns out to closely match a network in which links are formed uniformly at random (for details, see Section 3.2.3), and we see a number of features of large random networks. For example, there is a “giant component,” in which more than 100 of the students are connected by sequences of links in the network. The next largest component (the maximal set of students who are each linked to one another by sequences of links) only has 10 students in it. This component structure has important implications for the diffusion of disease, information, and behaviors, as discussed in detail in Chapters 7, 8, and 9, respectively.

In addition, note that the network is quite treelike: there are few loops or cycles in it. There are only a very large cycle in the giant component and a couple of smaller ones. The absence of many cycles means that as one walks along the links of the network until hitting a dead-end, most of the nodes that are met are new ones that have not been encountered before. This feature is important for the navigation of networks. It is found in many random networks in cases for which there are enough links to form a giant component but so few that the network is not fully connected. This treelike structure contrasts with the denser friendship network pictured in Figure 1.3, in which there are many cycles and a shorter distance between nodes.



FIGURE 1.3 Add Health data set friendships among high school students coded by race. Hispanic, black nodes; white, white nodes; black, gray nodes; Asian and other, light gray nodes.

The network pictured in Figure 1.3 is also from the Add Health data set and connects a population of high school students.<sup>5</sup> The nodes are coded by their race rather than sex, and the relationships are friendships rather than romantic relationships. This network is much denser than the romantic network.

A strong feature present in Figure 1.3 is what is known as *homophily*, a term from Lazarsfeld and Merton [425]. That is, there is a bias in friendships toward similar individuals; in this case the homophily concerns the race of the individuals. This bias is greater than what one would expect from the makeup of the population. In this school, 52 percent of the students are white and yet 85 percent

of white students' friendships are with other whites. Similarly, 38 percent of the students are black and yet 85 percent of these students' friendships are with other blacks. Hispanics are more integrated in this school, comprising 5 percent of the population but having only 2 percent of their friendships with other Hispanics.<sup>6</sup> If friendships were formed without race being a factor, then whites would have roughly 52 percent of their friendships with other whites rather than 85 percent.<sup>7</sup> This bias is referred to as "inbreeding homophily" and has strong consequences. As indicated by the figure, the students are somewhat segregated by race, which affects the spread of information, learning, and the speed with which things propagate through the network—themes that are explored in detail in what follows.

### 1.2.3 Random Graphs and Networks

The examples of Florentine marriages and high school friendships suggest the need for models of how and why networks form as they do. The last two examples in this chapter illustrate two complementary approaches to modeling network formation.

The next example of network analysis comes from the graph-theoretic branch of mathematics and has recently been extended in various directions by the literature in computer science, statistical physics, and economics (as will be examined in some of the following chapters). This model is perhaps the most basic one of network formation imaginable: it simply supposes that a completely random process is responsible for the formation of links in a network. The properties of such random networks provide some insight into the properties of social and economic networks. Some of the properties that have been extensively studied are how the distribution of links across different nodes, the connectedness of the network in terms of the presence of paths from one node to another, the average and maximal path lengths, and the number of isolated nodes present. Such random networks serve as a useful benchmark against which we can contrast observed networks; comparisons help identify which elements of social structure are not the result of mere randomness but must be traced to other factors.

Erdős and Rényi [227]-[229] provided seminal studies of purely random networks.<sup>8</sup> To describe one of the key models, fix a set of  $n$  nodes. Each link is formed with a given probability  $p$ , and the formation is independent across links.<sup>9</sup> Let us examine this model in some detail, as it has an intuitive structure and has been a springboard for many recent models.

Consider a set of nodes  $N = \{1, \dots, n\}$ , and let a link between any two nodes,  $i$  and  $j$ , be formed with probability  $p$ , where  $0 < p < 1$ . The formation of links is independent. This is a binomial model of link formation, which gives rise to a manageable set of calculations regarding the resulting network structure.<sup>10</sup> For instance, if  $n = 3$ , then a complete network forms with probability  $p^3$ , and a given network with two links (there are three such networks) forms with probability  $p^2(1 - p)$ , and a given network with one link forms with probability  $p(1 - p)^2$ , and the empty network that has no links forms with probability  $(1 - p)^3$ . More generally, any given network that has  $m$  links on  $n$  nodes has a probability of

$$p^m (1 - p)^{\frac{n(n-1)}{2} - m} \quad (1.2)$$

of forming under this process.<sup>11</sup>

We can calculate some statistics that describe the network. For instance, we can find the degree distribution fairly easily. The *degree* of a node is the number of links that the node has. The degree

distribution of a random network describes the probability that any given node will have a degree  $d$ .<sup>12</sup> The probability that any given node  $i$  has exactly  $d$  links is

$$\binom{n-1}{d} p^d (1-p)^{n-1-d}. \quad (1.3)$$

Note that even though links are formed independently, there is some correlation in the degrees of various nodes, which affects the distribution of nodes that have a given degree. For instance, if  $n = 2$  then it must be that both nodes have the same degree: the network either consists of two nodes of degree 0 or two of degree 1. As  $n$  becomes large, however, the correlation of degree between any two nodes vanishes, as the possibility of a link between them is only 1 out of the  $n - 1$  that each might have. Thus, as  $n$  becomes large, the fraction of nodes that have  $d$  links will approach (1.3). For large  $n$  and small  $p$ , this binomial expression is approximated by a Poisson distribution, so that the fraction of nodes that have  $d$  links is approximately<sup>13</sup>

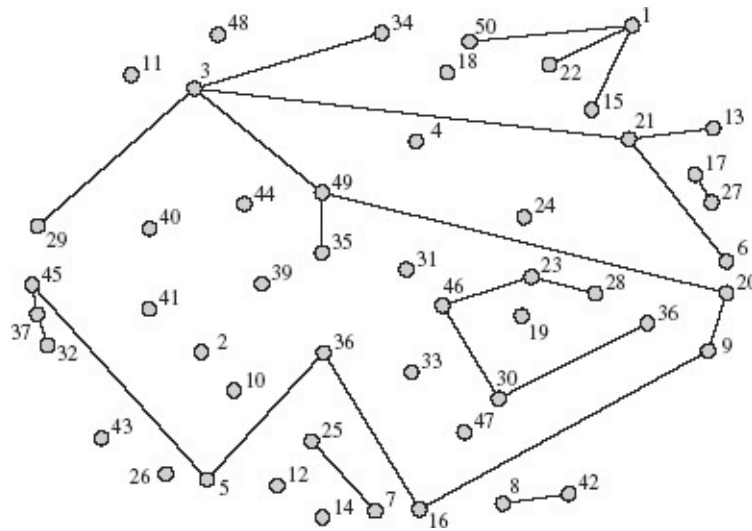


FIGURE 1.4 A randomly generated network with probability .02 on each link.

$$\frac{e^{-(n-1)p} ((n-1)p)^d}{d!}. \quad (1.4)$$

Given the approximation of the degree distribution by a Poisson distribution, the class of random graphs for which each link is formed independently with equal probability is often referred to as the class of *Poisson random networks*. I use this terminology in what follows.

To provide a better feel for the structure of such networks, consider a couple of Poisson random networks for different values of  $p$ . Set  $n = 50$  nodes, as this number produces a network that is easy to visualize. Let us start with an expected degree of 1 for each node, which is equivalent to setting  $p$  roughly .02. [Figure 1.4](#) pictures a network generated with these parameters.<sup>14</sup> This network exhibits a number of features that are common to this range of  $p$  and  $n$ . First, we should expect some isolated nodes. Based on the approximation of a Poisson distribution (1.4) with  $n = 50$  and  $p = .02$ , we should expect about 37.5 percent of the nodes to be isolated (i.e., have  $d = 0$ ), which is roughly 18 or 19 nodes. There happen to be 19 isolated nodes in the network.

[Figure 1.5](#) compares the realized frequency distribution of degrees with the Poisson approximation. The distributions match fairly closely. The network also has some other features common with other random networks with  $p$  and  $n$  in this relative range. In graph theoretical terms, the network is a *forest*, or a collection of trees. That is, there are no cycles in the network (where a *cycle* is a sequence of links that lead from one node back to itself, as described in more detail in Section 2.1.3). The chance of there being a cycle is relatively low with such a small link probability.

In addition, there are six *components* (maximal subnetworks such that every pair of nodes in the subnetwork is connected by a path or sequence of links) that involve more than one node. And one of the components is much larger than the others: it has 16 nodes, while the next largest only has 5 nodes in it. As we shall discuss shortly, this behavior is to be expected.

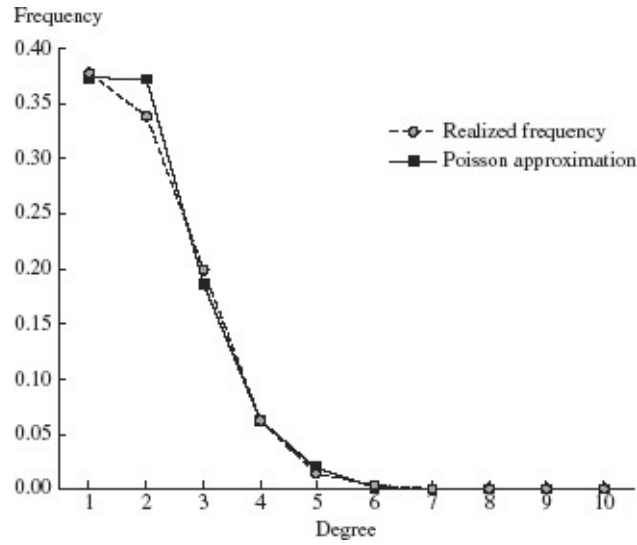


FIGURE 1.5 Frequency distribution of a randomly generated network and the Poisson approximation for a probability of .02 each link.

Let us start with the same number of nodes but increase the probability of a link forming to  $p = \log(50)/50 = .078$ , which is roughly the threshold at which isolated nodes should disappear. (This threshold is discussed in more detail in [Chapter 4](#).) Based on the approximation of a Poisson distribution (1.4) with  $n = 50$  and  $p = .08$ , we should expect about 2 percent of the nodes to be isolated (with degree 0), or roughly 1 node out of 50. This is exactly what occurs in the realized network in [Figure 1.6](#) (again, by chance).

As shown in [Figure 1.7](#), the realized frequency distribution of degrees is again similar to the Poisson approximation, although, as expected at this level of randomness, it is not a perfect match.

The degree distribution tells us a great deal about a network's structure. Let us examine this distribution in more detail, as it provides a first illustration of the concept of a *phase transition*, where the structure of a random network changes as the formation process is modified.

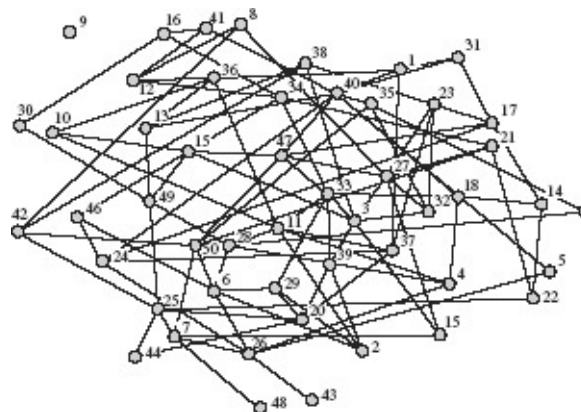


FIGURE 1.6 A randomly generated network with probability .08 on each link.

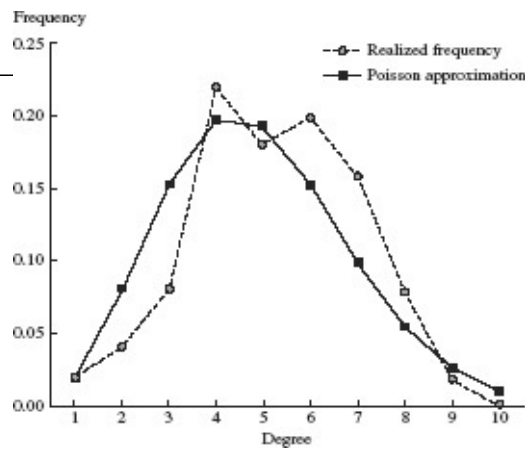


FIGURE 1.7 Frequency distribution of a randomly generated network and the Poisson approximation for a probability of .08 each link.

Consider the fraction of nodes that are completely isolated; that is, what fraction of nodes have degree  $d = 0$ ? From (1.4) this number is approximated by  $e^{-(n-1)p}$  for large networks, provided the average degree  $(n-1)p$  is not too large. To get a more precise expression, examine the threshold which this fraction is just such that we expect to have one isolated node on average, or  $e^{-(n-1)p} = 1/n$ . Solving this equation yields  $p(n-1) = \log(n)$ , where the average degree is  $(n-1)p$ . Indeed, this is the threshold for a phase transition, as shown in Section 4.2.2. If the average degree is substantially greater than  $\log(n)$ , then the probability of having any isolated nodes tends to 0, while if the average degree is substantially less than  $\log(n)$ , then the probability of having at least some isolated nodes tends to 1. In fact, as shown in Theorem 4.1, this threshold is such that if the average degree is significantly above this level, then the network is path-connected with a probability converging to 1 as  $n$  grows (so that any node can be reached from any other by a path in the network); below this level the network consists of multiple components with a probability converging to 1.

Other properties of random networks are examined in much more detail in [Chapter 4](#). While it is clear that completely random networks are not always a good approximation for real social and economic networks, the analysis here (and in [Chapter 4](#)) shows that much can be deduced from such models and there are some basic patterns and structures that emerge more generally. As we build more realistic models, similar analyses can be conducted.

## 1.2.4 The Symmetric Connections Model

Although random network-formation models give some insight into the sorts of characteristics that networks can have and exhibit some of the features seen in the Add Health social network data, they do not provide as much insight into the Florentine marriage network. In that example marriages were carefully arranged. The last example discussed in this chapter is from the game-theoretic economic literature and provides a basis for the analysis of networks that form when links are chosen by the agents in the network. This example addresses questions about which networks would maximize the welfare of a society and which arise if the players have discretion in choosing their links.

This simple model of social connections was developed by Jackson and Wolinsky [361]. In it, links represent social relationships, for instance friendships, between players. These relationships offer benefits in terms of favors, information, and the like, and also involve some costs. Moreover, players also benefit from indirect relationships. Thus having a “friend of a friend” also results in some indirect benefits, although of lesser value than the direct benefits that come from having

friend. The same is true of “friends of a friend of a friend,” and so forth. The benefit deteriorates with the distance of the relationship. This deterioration is represented by a factor  $\delta$  that lies between 0 and 1, which indicates the benefit from a direct relationship and is raised to higher powers for more distant relationships. For instance, in a network in which player 1 is linked to 2, 2 is linked to 3, and 3 is linked to 4, player 1 receives a benefit  $\delta$  from the direct connection with player 2, an indirect benefit  $\delta^2$  from the indirect connection with player 3, and an indirect benefit  $\delta^3$  from the indirect connection with player 4. The payoffs to this network of four players with three links is pictured in [Figure 1.8](#). For  $\delta < 1$  there is a lower benefit from an indirect connection than from a direct one. Players only pay costs, however, for maintaining their direct relationships.<sup>15</sup>



FIGURE 1.8 Utilities to the players in a three-link, four-player network in the symmetric connections model.

Given a network  $g$ ,<sup>16</sup> the net utility or payoff  $u_i(g)$  that player  $i$  receives from a network  $g$  is the sum of benefits that the player gets for his or her direct and indirect connections to other players less the cost of maintaining these links:

$$u_i(g) = \sum_{\substack{j \neq i : i \text{ and } j \text{ are} \\ \text{path-connected in } g}} \delta^{\ell_{ij}(g)} - d_i(g)c,$$

where  $\ell_{ij}(g)$  is the number of links in the shortest path between  $i$  and  $j$ ,  $d_i(g)$  is the number of links that  $i$  has ( $i$ 's degree), and  $c > 0$  is the cost for a player of maintaining a link.

Taking advantage of the highly stylized nature of the connections model, let us now examine which networks are “best” (most “efficient”) from society’s point of view, as well as which networks are likely to form when self-interested players choose their own links.

Let us define a network to be *efficient* if it maximizes the total utility to all players in the society. That is,  $g$  is efficient if it maximizes  $\sum_i u_i(g)$ .<sup>17</sup> It is clear that if costs are very low, it is efficient to include all links in the network. In particular, if  $c < \delta - \delta^2$ , then adding a link between any two agents  $i$  and  $j$  always increases total welfare. This follows because they are each getting at most  $\delta^2$  value from having any sort of indirect connection between them, and since  $\delta^2 < \delta - c$ , the extra value of a direct connection between them increases their utilities (and might also increase, and cannot decrease, the utilities of other agents).

When the cost rises above this level, so that  $c > \delta - \delta^2$  but  $c$  is not too high (see Exercise 1.3), it turns out that the unique efficient network structure is to have all players arranged in a “star” network. That is, there should be some central player who is connected to each other player, so that one player has  $n - 1$  links and each of the other players has 1 link. A star involves the minimum number of links needed to ensure that all pairs of players are path connected, and it places each player within two links of every other player. The intuition behind why this structure dominates other structures for moderate-cost networks is then easy to see. Suppose for instance we have a network with links between 1 and 2, 2 and 3, and 3 and 4. If we change the link between 3 and 4 to be one between 2 and 4, a star network is formed. The star network has the same number of links as the starting network, and thus the same cost and payoffs from direct connections. However, now all agents are within two links of one another, whereas before some of the indirect connections involved paths of length three ([Figure 1.9](#)).

$$\text{Total utility: } 6\delta + 4\delta^2 + 2\delta^3 - 6c$$

$$\text{Total utility: } 6\delta + 6\delta^2 - 6c$$

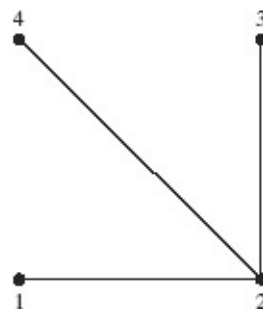
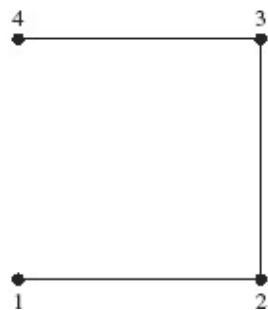


FIGURE 1.9 Gain in total utility from changing a line into a star.

As we shall see, this result is the key to a remarkably simple characterization of the set of efficient networks: (1) costs are so low that it makes sense to add all links; (2) costs are so high that no links make sense; or (3) costs are in a middle range, and the unique efficient architecture is a star network. This characterization of efficient networks being stars, empty, or complete actually holds for a fairly general class of models in which utilities depend on path length and decay with distance, as is shown in detail in Section 6.3.

We can now compare the efficient networks with those that arise if agents form links in a self-interested manner. To capture how agents act, consider a simple equilibrium concept introduced by Jackson and Wolinsky [361]. This concept is called *pairwise stability* and involves two rules about a network: (1) no agent can raise his or her payoff by deleting a link that he or she is directly involved in and (2) no two agents can both benefit (at least one strictly) by adding a link between themselves. This stability notion captures the idea that links are bilateral relationships and require the consent of both individuals. If an individual would benefit by terminating some relationship that he or she is involved in, then that link would be deleted, while if two individuals would each benefit by forming a new relationship, then that link would be added, and in either case the network would fail to be stable.

In the case in which costs are very low ( $c < \delta - \delta^2$ ), the direct benefit to the agents from adding and maintaining a link is positive, even if they are already indirectly connected. Thus in that case the unique pairwise-stable network is complete and is the efficient one. The more interesting case is when  $c > \delta - \delta^2$ , but  $c$  is not too high, so that the star is the efficient network.

If  $\delta > c > \delta - \delta^2$ , then a star network (that involves all agents) will be both pairwise stable and efficient. To see this we need only check that no player wants to delete a link, and no two agents both want to add a link. The marginal benefit to the center player from any given link already in the network is  $\delta - c > 0$ , and the marginal benefit to a peripheral player is  $\delta + (n - 2)\delta^2 - c > 0$ . Thus neither player wants to delete a link. Adding a link between two peripheral players only shortens the distance between them from two links to one and does not shorten any other paths, and since  $c > \delta - \delta^2$  adding such a link would not benefit either of the players. While the star is pairwise stable, in this cost range so are some other networks. For example if  $c < \delta - \delta^3$ , then four players connected in a circle would also be pairwise stable. In fact, as we shall see in Section 6.3, many other (inefficient) networks can be pairwise stable.

If  $c > \delta$ , then the efficient (star) network is not pairwise stable, as the center player gets only a marginal benefit of  $\delta - c < 0$  from any of the links. Thus in this cost range there cannot exist any pairwise-stable networks in which some player has only one link, as the other player involved in that link would benefit by severing it. For various values of  $c > \delta$  there exist nonempty pairwise-stable networks, but they are not star networks: they must be such that each player has at least two links.

This model makes it clear that there are situations in which individual incentives are not aligned

with overall societal benefits. While this connections model is highly stylized, it still captures some basic insights about the payoffs from networked relationships, and it shows that we can model the incentives that underlie network formation and see when the resultant networks are efficient.

This model also raises some interesting questions that are examined in the chapters that follow. How does the network that forms depend on the payoffs to the players for different networks? What are alternative ways of predicting which networks will form? What if players can bargain when they form links, so that the payoffs are endogenous to the network-formation process (as is true in many market and partnership applications)? How does the relationship between the efficient networks and those that form based on individual incentives depend on the underlying application and payoff structure?

### 1.3 ■ Exercises

**1.1 A Weighted Betweenness Measure** Consider the following variation on the betweenness measure in (1.1). Any given shortest path between two families is weighted by the inverse of the number of intermediate nodes on that path. For instance, the shortest path between the Ridolfi and Albizzi involves two links; and the Medici are the only family that lies between them on that path. In contrast, between the Ridolfi and the Ginori the shortest path is three links and there are two families, the Medici and Albizzi, that lie between the Ridolfi and Ginori on that path (see [Figure 1.1](#)).

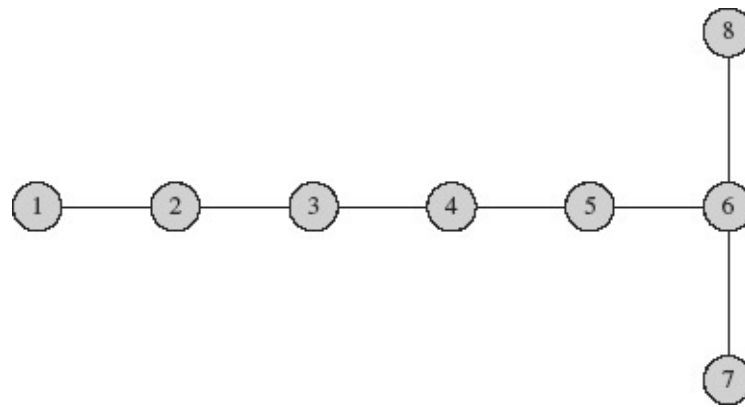


FIGURE 1.10 Differences in betweenness measures.

More specifically, let  $\ell(i, j)$  be the length of the shortest path between nodes  $i$  and  $j$  and let  $W_k(ij) = P_k(ij)/(\ell(i, j) - 1)$  (setting  $\ell(i, j) = \infty$  and  $W_k(ij) = 0$  if  $i$  and  $j$  are not connected). Then the weighted betweenness measure for a given node  $k$  is defined by

$$WB_k = \sum_{ij: i \neq j, k \notin \{i, j\}} \frac{W_k(ij)/P(ij)}{(n-1)(n-2)/2},$$

where the convention is  $\frac{W_k(ij)}{P(ij)} = 0/0 = 0$  if there are no paths connecting  $i$  and  $j$ . Show that

- (a)  $WB_k > 0$  if and only if  $k$  has more than one link in a network and some of  $k$ 's neighbors are not linked to one another;
- (b)  $WB_k = 1$  for the center node in a star network that includes all nodes (with  $n \geq 3$ ); and
- (c)  $WB_k < 1$  unless  $k$  is the center node in a star network that contains all nodes.

Calculate this measure for the network pictured in [Figure 1.10](#) for nodes 4 and 5. Contrast this measure with the betweenness measure in (1.1).

**1.2 Random Networks** Fix the probability of any given link forming in a Poisson random network to be  $p$ , where  $1 > p > 0$ . Fix some arbitrary network  $g$  on  $k$  nodes. Now, consider a sequence of random networks indexed by the number of nodes  $n$ , as  $n \rightarrow \infty$ . Show that the probability that a copy of the  $k$ -node network  $g$  is a subnetwork of the random network on the  $n$  nodes



- [click The Lost Boys: Inside Football's Slave Trade pdf, azw \(kindle\), epub](#)
- **[read online Families and Survivors](#)**
- [Mathematical Analysis II \(Universitext\) pdf, azw \(kindle\)](#)
- [download online Great Sky Woman \(Great Sky Woman, Book 1\)](#)
  
- <http://academialanguagebar.com/?ebooks/SPAD-XII-XIII-Aces-of-World-War-I.pdf>
- <http://anvilpr.com/library/All-the-Birds-in-the-Sky.pdf>
- <http://cavalldecartro.highlandagency.es/library/Mathematical-Analysis-II--Universitext-.pdf>
- <http://bestarthritiscare.com/library/Beggars-in-Spain--Sleepless--Book-1-.pdf>