C Related Documents And Translation
相关文献和译文

文章来源于：Chapter 1, *Social and Economic Network*, by Jackson (2009).

C.1 Why Model Networks? / 为什么对网络建模？

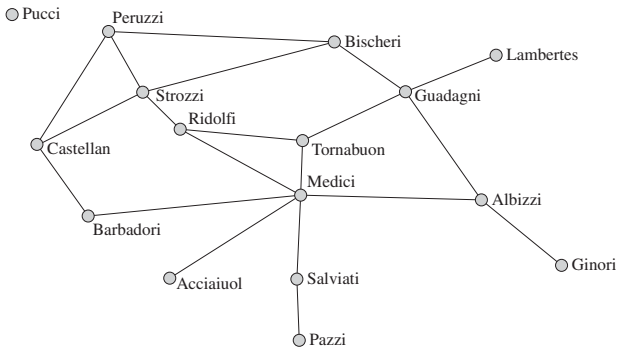
Social networks permeate our social and economic lives. They play a central role in the transmission of information about job opportunities and are critical to the trade of many goods and services. They are the basis for the provision of mutual insurance in developing countries. Social networks are also 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.

社会网络在我们的经济社会生活中无处不在。它在求职信息的传递过程中扮演了极其重要的角色、许多商品和服务的交易过程也离不开它。它还是发展中国家提供相互保险的基础。社会网络在疾病传播方式、商品购买决策、语言交流、投票行为，以及犯罪、接受教育、个人成功等等方面都有着决定性的重要作用。正因为社会网络结构从方方面面影响着我们的生活水平，所以我们理解它有助于(1) 社会网络结构是怎么影响人们的行为的(2)社会网络结构是怎么在一个社会中形成的。本书的目的就是对上述两个问题提供一个社会网络的分析框架。

在许多领域中都有不同网络建模方法，对应不同的技术，所以在介绍正式的相关定义和模型之前，有必要先通过一些例子形成对社会网络及各种模型的直观印象。我们在下面的例子中涵盖了不同的研究视角、话题和方法，以力求提供一个宽广的视野。

图 15: Network showing fifteenth-century florentine marriages. Data from Padgett and Ansell [516] (drawn using UCINET).
十五世纪佛罗伦萨人的婚姻网络，数据来源于Padgett and Ansell [516]，图像使用UCINET绘制。



C.2 A Set of Examples / **实例**

C.2.1 Florentine Marriages / 佛罗伦萨的家族联姻

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." Their accumulation of power in the early fifteenth century in florencewas 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 inter-marriages, 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 floun-

在第一个例子中我们首先来关注社会网络在意大利佛罗伦萨的银行家族美第奇（Medici）十五世纪兴起的过程中所扮演的角色。美第奇家族被称为"文艺复兴教父"。当时美第奇家族在通知弗洛伦斯的少数几个家族中并不占优势，他们拥有的财富相比而言较少，政治地位也较低，但科西莫·德·美第奇一手导演了该家族的权利聚集。科西莫通过平衡美第奇家族在家族联姻、经济交往和政治支持网络中的中心地位，巩固了该家族的政治和经济实力。他对于社会网络中不规则地位的理解帮助他建立并且控制一个政党先驱，与此同时其他家

dered 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[1].

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 did 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.

族还在各种事务中挣扎。Padgett and Ansell [516]绘制了十五世纪三十年代佛罗伦萨几个重要家族婚姻网络的图形，为上述观点提供了有力的证据。图15展示了当时佛罗伦萨几个重要家族之间的联系，每条线代表两个家族尘缘之间的婚姻[2]。

美第奇家族（科西莫·德·美第奇作为关键角色）在其此间兴起并逐渐掌控了佛罗伦萨的政治和商业。以前佛罗伦萨处于被几个大家族所统治所形成的寡头政治。如果我们关注各个家族的财富和政治势力情况，可以发现美第奇家族在其中并不突出，故而我们需借助观察社会关系的结构来理解为什么美第奇家族会异军突起。比如，斯特罗齐（Strozzi）家族有着更多的财富和当地立法机构更多的席位，然而美第奇家族的兴起让其黯然失色。正如Padgett and Ansell [516]所阐述的，理解家族兴起的关键正在于社会网络结构。

如果我们粗略计算当时网络中各个家族的重要性，只是单单的对家族婚姻连接数进行加总，就会发现美第奇家族位居榜首。然而，这种优势并不明显，他们仅仅以3/2险胜居于次席的斯特罗齐和瓜达尼（Guadagni）。虽然这对我们有所启发，但还不足以用作证据。我们需要更仔细的观察社会网络的结构，从而更好地挖掘美第奇家族成功的秘密。尤

Let $P(ij)$ denote the number of shortest paths connecting family i to family j [3]. 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 $i = Barbadori$ and $j = Guadagni$, then $P(ij) = 2$. As the Medici lie on both paths, $P_k(ij) = 2$ when we set $k = Medici$, and $i = Barbadori$ and $j = Guadagni$. In contrast this number is 0 if $k = Strozzi$, and is 1 if $k = Albizzi$. 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 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

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

其值得一提的就是对于中间度（betweenness）的衡量。

$P(ij)$ 代表家族i与j之间最短通路的数量[4]，$P_k(ij)$ 代表了其中包含家族k的通路的数量。比如，在图15中，Barbadori和Guadagni家族之间最短通路长度为3，有两条：Barbadori - Medici - Albizzi - Guadagni和Barbadori - Medici - Tornabuon - Guadagni。如果我们定义$i = Barbadori$、$j = Guadagni$，那么$P(ij) = 2$。因为美第奇家族在两条通路上，所以$P_k(ij) = 2$。如果我们定义$k = Medici$，、$i = Barbadori$ 、$j = Guadagni$。同样的，如果让$k = Strozzi$，那么该值为0；如果$k = Albizzi$则该值为1。正因如此，美第奇家族在Barbadori 和Guadagni家族的联系中居于关键地位。

为了获取一个家族何种程度上居于中心地位的直觉，我们可以计算中心度平均值。我们可以计算其在每一对家族之间最短通路的总数量所占的比例。美第奇家族在Barbadori和Guadagni的最短通路中所占的比例为1，而Albizzi为1/2。一个家族在其他各家族两两组合中的平均值代表了中心度，用来衡量其权利（依据freeman [255]）。具体而言，我们可以对人一个家族k

for each family k, where $Pk(ij)/P(ij) = 0$ 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 $(n \quad 1)(n \quad 2)/2$ 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[5]. While aided by circumstance (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.

计算

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

其中$Pk(ij)/P(ij) = 0$如果i和j之间没有通路。分母表明其他家族最多可能有$(n \quad 1)(n \quad 2)/2$种组合。通过这样的方法计算可知，美第奇家族的中心度为0.522。故我们如果观察其他所有家族之间的最短通路，其中一半以上都需通过美第奇家族。相比而言，Strozzi为0.103，约10%。紧接美第奇家族位于次席的是Guadagni，其中心度为0.255。在某种程度上，婚姻关系正是交流信息、进行商贸往来、制定政治决策的关键。正因如此，美第奇家族相比于其他家族而言居于更有利的地位，至少从中心度上来看。当继续考虑当时社会环境的时候（比如战争后的财政问题），除了美第奇家族外别无旁人可以巩固势力。正如Padgett and Ansell [516, p. 1259]所指出的，"美第奇的政治权利起源于其他家族的在社会网络中的分隔，只有美第奇一个家族沟通了他们"。

这里的分析表明，社会网络结构是如何超越其他政治和经济特征所进行的分析，来提供一个有洞察力的视角。这个例子也说明社会网络结构的重要性并不是仅仅计算社会之间联系数就可得知的，对中间度或者中心性（centrality）的计算可以从不同角度来捕捉社会网络结构的特征。

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 patriarchs (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.

这个例子亦提及了贯穿于本书的其他问题。比如，美第奇家族是偶然获得了社会网络中的如此特殊的地位，还是通过周密计划的？正如Padgett and Ansell [516, 脚注13]所述，"现代的读者须知，当时所有家族的婚姻都是其族长（或同等地位的人）一手主导的。家族之间的联姻一定程度上正是一种政治联盟的体现。"在这种观点之下，我们接下来会问为什么其他家族相互之间没有形成更多联系，或者试图削弱美第奇的中心地位。我们亦会好奇，最终的社会网络是不是对于各家族来说都是最优的：对美第奇来说、对寡头通知来说、对十五世纪佛罗伦萨当地的政治经济运行情况来说。我们可以通过建立社会网络的成本和收益分析模型、以及网络形成原理的模型，来开始回答上述类型的问题。

1 These data were were originally collected by Kent [387], but were first coded by Padgett and Ansell [516], who discuss the network relationships in more detail. The analysis provided here is just a teaser that offers a glimpse of the importance of the network structure. The interested reader should consult Padgett and Ansell [516] for a much richer analysis.

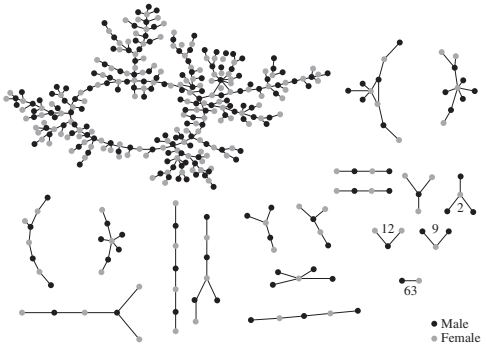
2 数据最初由Kent [387]搜集，但最早被Padgett and Ansell [516]编码并分析，讨论了更为具体的社会网络联系情况。这里的分析仅仅是对社会网络结构重要性的概览。对更详细的分析感兴趣的读者可以参见Padgett and Ansell [516]。

3 Formal definitions of path and some other terms used in this chapter appear in Chapter 2. The ideas should generally be clear, but the unsure reader can skip forward if helpful. Paths represent the obvious thing: a series of links connecting one node to another.

4 关于路径和其他术语更为正式的定义将在第二章给出。下述定义一般说来比较清楚，不过对此不确定的读者也可以直接跳过。通路（path）代表了一个节点到另一个结点之间的边的集合。

5 The calculations here are conducted on a subset of key families (a data set fromWasserman and faust [650]), rather than the entire data set, which consists of hundreds of families. As such, the numbers differ slightly from those reported in footnote 31 of Padgett and Ansell [516]. Padgett and Ansell also find similar results for centrality between the Medici and other families in terms of a network of business ties.

图 16: A network based on the Add Health data set. A link denotes a romantic relationship, and the numbers by some components indicate how many such components appear. figure from Bearman, Moody, and Stovel [51].
来源于增加健康数据集的一个网络。一条边代表一个恋爱关系，某些分支的上面的数字代表了它们出现的次数。图来源于Bearman, Moody, and Stovel [51]。



C.2.2 friendships and Romances among High School Students / 高中生的朋友和恋爱网络

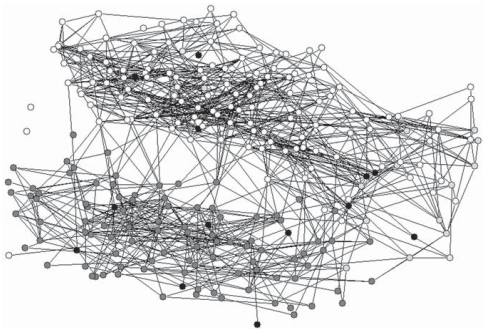
The next example comes from the the National Longitudinal Adolescent Health Data Set, known as "Add Health." [1] These data provide detailed social network information for more than 90,000 students from U.S. high schools interviewed during the mid-1990s, together with various data on the students' socioeconomic 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 shows a network of romantic relationships as found through surveys of students in one of the high schools in the study. The students

下面的例子来源于全国青少年健康纵向调查数据（National Longitudinal Adolescent Health Data Set），被称作"增加健康" [2]。这些数据提供了多于9万个美国高中生的详细的社会网络信息，调查于二十世纪九十年代中期，且涵盖了学生们的社会经济背景、行为和观点。该数据提供且指出了网络的一些特征，本书后面章节将更为详细的讨论它们。

图16展现了该调查中某高中的学生之间恋爱关系的网络。这些学生被要求列出在调查前六个

图 17: 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.
增加健康数据集中高中生之间的朋友关系，由种族区分：西班牙裔为黑色节点，白人为白色节点，黑人为灰色节点，亚裔及其他为浅灰色节点。



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 divide 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

月内恋爱的对象。

图16近似为一个二分网络（*bipartite* network），即节点可以被分成两组——男生和女生，所以只有两组之间有联系（有少数例外）。除了它的近乎二分特性，节点度的分布（与每个节点相连的边的数量）很符合边一致排列的随机图（具体见3.2.3），并且我们看到大型随机网络的很多特征。例如，有一个"巨大的分支（giant component）"，其中多余100个学生由一系列边所连接，而次大的分支只有10个学生。分支的结构对于疾病、信息和行为的传播有着重要的影响，将在

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.

第7、8、9章分别讨论。

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 in 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.

此外，这个网络很类似于树：只有很少的循环或回路。在巨大的分支里面只有一个非常大的回路和一些小回路。回路的缺乏说明当沿着网络的边走直到死节点的时候，经过的大多数节点都是以前没有碰到过的。这种特征在网络导航中尤为重要。在许多随机网络中都发现，有足够的边来构造一个巨大的分支，但只有极少数情况为全连通的。这种类树结构与图17中的密集的朋友关系形成了鲜明的对比，后者中回路更多而且节点之间的距离更短。

The network pictured in figure 1.3 is also from the Add Health data set and connects a population of high school students[3]. 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 romance network.

图17也是来源于增加健康数据集，涵盖了一群高中生[4]。节点用种族区分，而且他们之间的关系仅仅是朋友关系而不是恋人关系。该网络比恋爱网络更为稠密。

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 make up of the population. In this school, 52 percent of the students are white

图17中呈现的一个重要特征是Lazarsfeld and Merton [425]所提出的一个术语——同质性（homophily），即物以类聚的倾向。而且这种倾向性比我们从人口组成中所预期的更高。在该学校，52%的学生都为白人，且其中85%的朋友均为白人。类似的，

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[5]. 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. 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.

38%的学生为黑人，且其中85%的朋友也为黑人。西班牙裔在学校中融合的更好，只占5%的人口而其中只有2%的朋友也是西班牙裔[6]。如果种族并不应该决定友谊的因素，那么白人的朋友中应该有约52%的也是白人，而不是85%。这种倾向被称作"血缘同质性"，而且又很想的连续性。正如该图所指出的，学生们多多少少被种族所分隔，因而影响了网络中信息、学习的传播范围和速度——下文中将进一步探讨各种模式。

1 Add Health is a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris and funded by grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from AddHealth should contact AddHealth, Carolina Population Center, 123 West franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). The network data that I present in this example were extracted by James Moody from the Add Health data set.

2 "增加健康"是由J. Richard Udry、Peter S. Bearman和Kathleen Mullan Harris设计并发起、由美国国家儿童健康和人类发展研究所和17个其他组织共同资助的项目（P01-HD31921）。特别感谢Ronald R. Rindfuss 和Barbara Entwisle在最初设计中的帮助。对增加健康数据感兴趣的读者请直接联系：AddHealth, Carolina Population Center, 123 West franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu)。我这里使用的网络数据是James Moody从该数据集中抽取形成的。

3 A link indicates that at least one of the two students named the other as a friend in the survey. Not all friendships were reported by both students. for more detailed discussion of these particular data see Currarini, Jackson, and Pin [182].

4 一条边表明至少两个当中的一个学生在调查中把另一个学生视作朋友。并不是所有的朋友关系都被双方所报告。更具体的关于此点的讨论参见Currarini, Jackson, and Pin [182]。

5 The hispanics in this school are exceptional compared to what is generally observed in the larger data set of 84 high schools. Most racial groups (including hispanics) tend to have a greater percentage of own-race friendships than the percentage of their race in the population, regardless of their fraction of the population. See Currarini, Jackson, and Pin [182] for details.

6 该学校中西班牙裔的表现在所有84个高中中表现的最为突出。大多数种族群体（包括西班牙裔）倾向于有更高比例的同同种族友谊，相对于该种族所占的人口比例而言。详情可见Currarini, Jackson, and Pin [182]。