1

Looking for Social Structure

1.1 Introduction

The social sciences focus on structure: the structure of human groups, communities, organizations, markets, society, or the world system. In this book, we conceptualize social structure as a network of social ties. Social network analysts assume that interpersonal ties matter, as do ties among organizations or countries, because they transmit behavior, attitudes, information, or goods. Social network analysis offers the methodology to analyze social relations; it tells us how to conceptualize social networks and how to analyze them.

In this book, we present the most important methods of exploring social networks, emphasizing visual exploration. Network visualization has been an important tool for researchers from the very beginning of social network analysis. This chapter introduces the basic elements of a social network and shows how to construct and draw a social network.

1.2 Sociometry and Sociogram

The basis of social network visualization was laid by researchers who called themselves sociometrists. Their leader, J. L. Moreno, founded a social science called *sociometry*, which studies interpersonal relations. Society, they argued, is not an aggregate of individuals and their characteristics, as statisticians assume, but a structure of interpersonal ties. Therefore, the individual is not the basic social unit. The social atom consists of an individual and his or her social, economic, or cultural ties. Social atoms are linked into groups, and, ultimately, society consists of interrelated groups.

From their point of view, it is understandable that sociometrists studied the structure of small groups rather than that of of society at large.

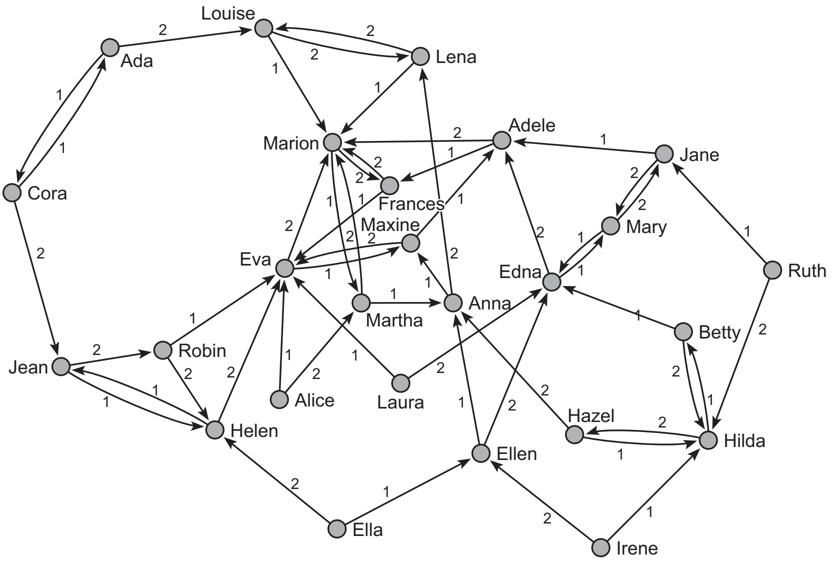

Figure 2. Sociogram of dining-table partners.

In particular, they investigated social choices within a small group. They asked people questions such as, “Whom would you choose as a friend [colleague, advisor, etc.]?” This type of data has since been known as *sociometric choice*. In sociometry, social choices are considered the most important expression of social relations.

Figure 2 presents an example of sociometric research. It depicts the choices of twenty-six girls living in one “cottage” (dormitory) at a training school in New York State. The girls were asked to choose the girls they liked best as their dining-table partners. First and second choices are selected only. (Here and elsewhere, a reference on the source of the data can be found in “Further Reading” at the end of each chapter.)

Figure 2 is an example of a sociogram, which is a graphical representation of group structure. The sociogram is among the most important instruments that originated in sociometry, and it is the basis for the visualization of social networks. You have most likely already “read” and understood the figure without needing the following explanation, which illustrates its visual appeal and conceptual clarity. In this sociogram, each girl in the dormitory is represented by a circle. For the sake of identification, the girls’ names are written next to the circles. Each arc (arrow) represents a choice. The girl who chooses a specific peer as a dining-table companion sends an arc toward her. Irene (in the bottom right of the figure), for instance, chose Hilda as her favorite dining-table partner and Ellen as her second choice, as indicated by the numbers labeling each arrow.

A sociogram depicts the structure of ties within a group. This example shows not only which girls are popular, as indicated by the number of choices they receive, but also whether the choices come from popular or unpopular girls. For example, Hilda receives four choices from Irene, Ruth, Hazel, and Betty, and she reciprocates the last two choices. However, none of these four girls is chosen by any of the other girls. Therefore, Hilda is located at the margin of the sociogram, whereas Frances, who is chosen only twice, is more central because she is chosen by Adele and Marion, two of the "popular" girls. A simple count of choices does not reveal this, whereas a sociogram does.

The sociogram has proved to be an important analytical tool that helped reveal several structural features of social groups. In this book, we make ample use of it.

1.3 Exploratory Social Network Analysis

Sociometry is not the only tradition in the social sciences that focuses on social ties. Without going into historical detail (see the "Further Reading" section for references on the history of social network analysis), we may note that scientists from several social sciences have applied network analysis to different kinds of social relations and social units. Anthropologists study kinship relations, friendship, and gift giving among people rather than sociometric choice; social psychologists focus on affections; political scientists study power relations among people, organizations, or nations; economists investigate trade and organizational ties among firms. In this book, the word *actor* refers to a person, organization, or nation that is involved in a social relation. We may say that social network analysis studies the social ties among actors.

> The *main goal* of social network analysis is detecting and interpreting patterns of social ties among actors.

This book focuses on exploratory social network analysis only. This means that we have no specific hypotheses about the structure of a network beforehand that we can test. For example, a hypothesis on the dining-table partners network could predict a particular rate of mutual choices (e.g., one of five choices will be reciprocated). This hypothesis must be grounded in social theory and prior research experience. The hypothesis can be tested provided that an adequate statistical model is available.

We use no hypothesis testing in this book (except for the last chapter) because we cannot assume prior research experience in an introductory course book, and because the statistical models involved are complicated.

Therefore, we adopt an exploratory approach, which assumes that the structure or pattern of ties in a social network is meaningful to the members of the network and, hence, to the researcher. Instead of testing prespecified structural hypotheses, we explore social networks for meaningful patterns.

For similar reasons, we pay no attention to the estimation of network features from samples. In network analysis, estimation techniques are even more complicated than estimation in statistics, because the structure of a random sample seldom matches the structure of the overall network. It is easy to demonstrate this. For example, select five girls from the dining-table partner network at random and focus on the choices among them. You will find fewer choices per person than the two choices in the overall network for the simple reason that choices directed to girls outside the sample are neglected. Even in this simple respect, a sample is not representative of a network.

We analyze entire networks rather than samples. However, what is the entire network? Sociometry assumes that society consists of interrelated groups, so a network encompasses society at large. Research on the so-called Small World problem suggested that ties of acquaintance connect us to almost every human being on Earth in six or seven steps (i.e., with five or six intermediaries), so our network eventually covers the entire world population, which is clearly too large a network to be studied. Therefore, we must use an artificial criterion to delimit the network we are studying. For example, we may study the girls of one dormitory only. We do not know their preferences for table partners in other dormitories. Perhaps Hilda is the only vegetarian in a group of carnivores, and she prefers to eat with girls of other dormitories. If so, including choices between members of different dormitories will alter Hilda's position in the network tremendously.

Because boundary specification may seriously affect the structure of a network, it is important to consider it carefully. Use substantive arguments to support your decision of whom to include in the network and whom to exclude.

Exploratory social network analysis consists of four parts: the definition of a network, network manipulation, determination of structural features, and visual inspection. In the following subsections, we present an overview of these techniques. This overview serves to introduce basic concepts in network analysis and to help you get started with the software used in this book.

1.3.1 Network Definition

To analyze a network, we must first have one. What is a network? Here, and elsewhere, we use a branch of mathematics called *graph theory* to define concepts. Most characteristics of networks that we introduce in this book originate from graph theory. Although this is not a course in graph theory, you should study the definitions carefully to understand what you are doing when you apply network analysis. Throughout this book, we present definitions in text boxes to highlight them.

> A *graph* is a set of vertices and a set of lines between pairs of vertices.

What is a graph? A graph represents the structure of a network; all it needs for this is a set of vertices (which are also called points or nodes) and a set of lines, with each line connecting two vertices.

A *vertex* (singular of vertices) is the smallest unit in a network. In social network analysis, it represents an actor (e.g., a person, such as a girl in a dormitory; an organization; or a country). A vertex is usually identified by a number.

A *line* is a link between two vertices in a network. In social network analysis, it can be any social relation. A line is defined by its two endpoints, which are the two vertices that are *incident* with the line.

A *loop* is a special kind of line, namely, a line that connects a vertex to itself. In the dining-table partners network, loops do not occur because girls are not allowed to choose themselves as dinner-table partners. However, loops are meaningful in some kinds of networks.

A line is directed or undirected. A directed line is called an *arc*, whereas an undirected line is an *edge*. Sociometric choice is best represented by arcs, because one girl chooses another and choices need not be reciprocated (e.g., Ella and Ellen in Figure 2).

A *directed graph*, or *digraph*, contains one or more arcs. A social relation that is undirected (e.g., cooperation on school projects) is represented by an edge because both individuals are equally involved in the relation. An *undirected graph* contains no arcs: All of its lines are edges.

Formally, an arc is an ordered pair of vertices in which the first vertex is the *sender* (the *tail* of the arc) and the second the *receiver* of the tie (the *head* of the arc). An arc points *from* a sender *to* a receiver. In contrast, an edge, which has no direction, is represented by an unordered pair. It does not matter which vertex is first or second in the pair. We should note, however, that an edge is usually equivalent to a *bidirectional arc*: If Ella and Ellen cooperate (undirected), we can say that Ella cooperates with Ellen and Ellen cooperates with Ella (directed). It is important to note this, as we will see in later chapters.

The dining-table partners network has no *multiple lines* because no girl was allowed to nominate the same girl as both first and second choices. Without this restriction, which was imposed by the researcher, multiple arcs could have occurred, and they actually do occur in other social networks.

In a graph, multiple lines are allowed, but when we say that a graph is *simple*, we indicate that it has no multiple lines. In addition, a simple undirected graph contains no loops, whereas loops are allowed in a simple directed graph. It is important to remember this.

> A *simple undirected graph* contains neither multiple edges nor loops.
>
> A *simple directed graph* contains no multiple arcs.

Now that we have discussed the concept of a graph at some length, it is very easy to define a network. A network consists of a graph and additional information on the vertices or lines of the graph. We should note that the additional information is irrelevant to the structure of the network because the structure depends on the pattern of ties.

> A *network* consists of a graph and additional information on the vertices or the lines of the graph.

In the dining-table partners network, the names of the girls represent additional information on the vertices that turns the graph into a network. Because of this information, we can see which vertex identifies Ella in the sociogram. The numbers printed near the arcs and edges offer additional information on the links between the girls: A 1 indicates a first choice, and a 2 represents a second choice. These are called *line values* and usually indicate the strength of a relation, which is a quantity.

Lines can also have a specific quality; for example, they can be of a particular type. All lines in the dining-table partners network are of the same kind, expressing seating preferences. We can say that they express the seating preferences *relation*. A network can, however, contain more than one relation. Perhaps we also know which of the girls cooperated on class projects. Because this information involves the same set of vertices (the girls), we can add this information as a second set of lines, that is, as a second relation to the network. This creates a *multiple relations network*, which is also called a *multiplex network*.

The dining-table partners network is clearly a network and not a graph. It is a directed simple network because it contains arcs (directed) but not multiple arcs (simple). In addition, we know that it contains just one relation and no loops. Several analytical techniques we discuss assume that loops and multiple lines are absent from a network. However, we do not always spell out these properties of the network but rather indicate whether it is simple. Take care!

Application

In this book, we learn social network analysis by doing it. We use the computer program Pajek – Slovenian for spider – to analyze and draw social networks. The website dedicated to this book (http://mrvar.fdv.uni-lj.si/pajek/) contains the software. We advise you to download and install Pajek on your computer (see Appendix 1 for more details) and all example data sets from this website. Store the software and data sets on the hard disk of your computer following the guidelines provided on the website. When you have done so, carry out the commands that we discuss under "Application" in each chapter. This will familiarize you with the structural concepts and with Pajek. By following the instructions under "Application" step by step, you will be able to produce the figures and results presented in the theoretical sections, unless stated differently. Sometimes, the visualizations on your computer screen will be slightly different from the figures in the book. If the general patterns match, however, you know that you are on the right track.

Network data file

Some concepts from graph theory are the building blocks or *data objects* of Pajek. Of course, a network is the most important data object in Pajek, so let us describe it first. In Pajek, a network is defined in accordance with graph theory: a list of vertices and lists of arcs and edges, where each arc or edge has a value. Take a look at the partial listing of the data file for the dining-table partners network (Figure 3; note that part of the vertices and arcs are replaced by [. . .]). Open the styled `Dormitory.net`, which you have downloaded from the website, in a word processor program to see the entire data file.

First, the data file specifies the number of vertices. Then, each vertex is identified on a separate line by a serial number, a textual label (enclosed in quotation marks [" "]), and three real numbers between 0 and 1, which indicate the position of the vertex in three-dimensional space if the network is drawn. We pay more attention to these coordinates in Chapter 2. For now, it suffices to know that the first number specifies the horizontal position of a vertex (0 is at the left of the screen and 1 at the right), and the second number gives the vertical position of a vertex (0 is the top of the screen and 1 is the bottom). The text label is crucial for identification of vertices, the more so because serial numbers of vertices may change during the analysis.

The list of vertices is followed by a list of arcs. The `*Arcs` statement assigns relation number 1 (the integer after the colon) to the arcs specified in the subsequent lines, and the relation is labeled "Dining-table partner choice." Note that labels should be enclosed in double quotation marks. Each line identifies an arc by the serial number of the sending vertex, followed by the number of the receiving vertex and the value of the arc. Just as in graph theory, Pajek defines a line as a pair of vertices. In Figure 3, the first arc represents Ada's choice (vertex 1) of Louise (vertex 3) as a

```
*Vertices 26
  1 "Ada"                    0.1646    0.2144    0.5000
  2 "Cora"                   0.0481    0.3869    0.5000
  3 "Louise"                 0.3472    0.1913    0.5000
  4 "Jean"                   0.1063    0.5935    0.5000
[…]
 25 "Laura"                  0.5101    0.6133    0.5000
 26 "Irene"                  0.7478    0.8087    0.5000
*Arcs :1 "Dining-table partner choice"
  1   3 2
  1   2 1
  2   1 1
  2   4 2
[…]
*Edges :2 "Cooperation"
  1   2 1 l "Math 2a"
  2   4 1 l "Math 2a"
  1   4 1 l "Math 2a"
[…]
```

Figure 3. Partial listing of a multiple relations network data file for Pajek.

dining-table partner. Louise is Ada's second choice; Cora is her first choice, which is indicated by the second arc. A list of edges is similar to a list of arcs, with the exception that the order of the two vertices that identify an edge is disregarded in computations. In this data file, edges express cooperation among girls, which is coded as relation number 2 with an appropriate label. The edges are labeled with the name of the project the girls cooperated on (see Appendix 2 for directions on adding information to lines and vertices). Note that a Pajek network file can contain several `*Arcs` and `*Edges` statements, and the relation number and label is not mandatory.

It is interesting to note that we can distinguish between the structural data or graph and the additional information on vertices and lines in the network data file. The graph is fully defined by the list of vertex numbers and the list of pairs of vertices, which defines its arcs and edges. This part of the data, which is printed in regular typeface in Figure 3, represents the structure of the network. The vertex labels and coordinates, the relation numbers and labels, line values and labels (in italics) specify the additional properties of vertices and lines that make these data a network. Although this information is extremely useful, it is not required: Pajek will use vertex numbers as default labels and set relation numbers and line values to 1 if they are not specified in the data file. In addition, Pajek can use several other data formats (e.g., the matrix format), which we do not discuss here. They are briefly described in Appendix 1.

Figure 4. Main screen of Pajek.

It is possible to generate ready-to-use network files from spreadsheets and databases by exporting the relevant data in plain text format. For medium or large networks, processing the data as a relational database helps data cleaning and coding. See Appendix 1 for details.

*File>
Network>
Read*

We explain how to create a new network in Section 1.4. Let us first look at the dormitory network containing the dining-table partner choices and the cooperation among girls. First, start Pajek by double-clicking the file `Pajek.exe` on your hard disk. The computer will display the Main screen of Pajek (Figure 4). From this screen, you can open the dormitory network with the *Read* command in the *File* menu or by clicking the button with an icon of a folder under the word *Networks*. In both cases, the usual Windows® file dialog box appears in which you can search and select the file `Dormitory.net` on your hard disk, provided that you have downloaded the example data sets from the book's website.

*Network
drop-down
menu*

When Pajek reads a network, it displays its name in the top Network drop-down menu. This menu is a list of the networks accessible to Pajek. You can open a drop-down menu by left-clicking on the button with the triangle at the right. The network that you select in the list is shown when the list is closed (e.g., the network `Dormitory.net` in Figure 2). Notice that the number of vertices in the network is displayed in parentheses next to the name. The selected network is the *active network*, meaning that any operation you perform on a network will use this particular network. For example, if you use the *Draw* menu now, Pajek draws the dormitory network for you.

The Main screen displays several more drop-down menus beneath the two network drop-down menus. Each of these menus represents a data object in Pajek: partitions (three drop-down menus), vectors (two drop-down menus), permutations (two drop-down menus), clusters, and hierarchies. Later chapters will familiarize you with these data objects. Note that each object can be opened, saved, or edited from the *File* menu or by using the four icons to the left of a drop-down menu (see Section 1.4).

1.3.2 Manipulation

In social network analysis, it is often useful to modify a network. For instance, large networks are too big to be drawn, so we extract a meaningful part of the network that we inspect first. Visualizations work much better for small (some dozens of vertices) to medium-sized (some hundreds of vertices) networks than for large networks with thousands of vertices. When social networks contain different kinds of relations, we may focus on one relation only; for instance, we may want to study dining-table partner choices only in the dormitory network. Finally, some analytical procedures demand that complex networks with loops or multiple lines are reduced to simple graphs first.

Application

Network manipulation is a very powerful tool in social network analysis. In this book, we encounter several techniques for modifying a network or selecting a subnetwork. Network manipulation always results in a new network. In general, many commands in Pajek produce new networks or other data objects, which are stored in the drop-down menus, rather than graphical or tabular output.

Menu structure

The commands for manipulating networks are accessible from menus in the Main screen. The Main screen menus have a clear logic. Manipulations that involve one type of data object are listed under a menu with the object's name; for example, the *Network* menu contains all commands that operate on one network, and the *Networks* menu lists operations on two networks. Manipulations that need different kinds of objects are listed in the *Operations* menu. When you try to locate a command in Pajek, just consider which data objects you want to use. Pajek also groups commands that apply to a special type of network; e.g., all commands that require a multiple relations network are available in the submenu *Multiple Relations Network* of the *Network* Menu. You will learn more about other types of networks in the following chapters.

[Main] Network> Multiple Relations Network> Extract Relation(s) into Separate Network(s)

The following example highlights the use of menus in Pajek and their notation in this book. If we want to restrict the analysis to the dining partners relation (as in Figure 2), we must create a new network containing only the lines that belong to the first relation. Because this operation concerns one network and no other data objects, we must look for it in the *Network* menu. If we left-click on the word *Network* in the upper left