Part I - Fundamentals

Social network analysis focuses on relations between people, groups of people, organizations, countries, et cetera. These relations combine into networks which we will learn to analyze.

In the first part of the book, we introduce the concept of a social network. We discuss several types of networks and the ways in which we can analyze them numerically and visually with the computer software (program Pajek) which is used throughout this book.

After studying Chapters 1 and 2, you should understand the concept of a social network and you should be able to create, manipulate, and visualize a social network with the software provided with this book.

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 relations. Social network analysts assume that interpersonal relations matter, as do relations between 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 will present the most important methods of exploring social networks, emphasizing visual exploration of social networks. 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 studied interpersonal relations. Society, it was argued, is not an aggregate of individuals and their characteristics, as statisticians assume, but a structure of interpersonal relations. Therefore, the individual is not the basic social unit. The social atom consists of an individual and his or her social, economic, or cultural relations. 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 the structure of society at large. In particular, they investigated social choices within a small group. They asked people questions like: 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 to be the most important expression of social relations. Figure 1 presents an example of sociometric research. It depicts the choices of 26 girls living in one 'cottage' (dormitory) at a New York State Training School. The girls were asked to choose the girls they liked best as their dining-table partners. First and second choices are selected only. The data can be found in the file `Dining-table_partners.net` on the CD ROM.
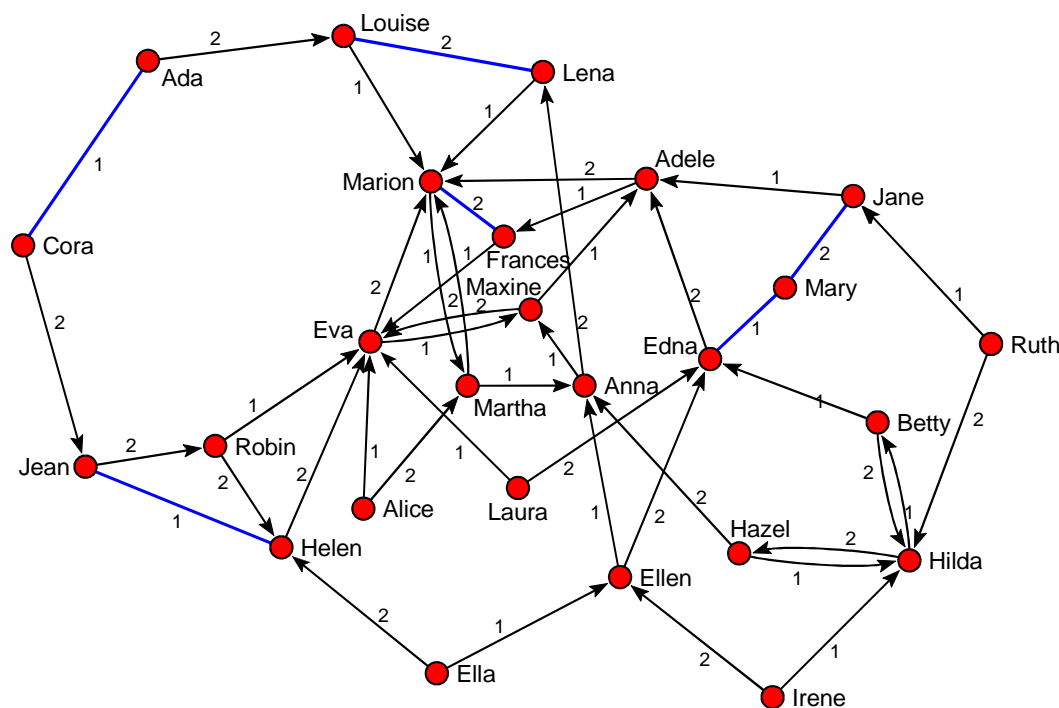
Figure 1 - Sociogram of dining-table partners.

Figure 1 is an example of a sociogram, which is a graphical representation of group structure. The sociogram is among the most important instruments originated in sociometry and it is the basis for the visualization of social networks. Probably, you have already 'read' and understood the figure without the explanation that follows now, which illustrates its visual appeal and conceptual clarity. In the sociogram, each girl of the dormitory is represented by a circle. In this network, girls are the social actors. For the sake of identification, names are written next to the circles.

Each arc represents a choice. The girl that chooses a peer as a dining-table companion sends an arc towards her. In this sociogram, some mutual choices are represented by blue lines because a single line is visually more attractive than a pair of arcs. Cora and Ada (top left), for instance, choose each other as their favorite dining-table partner, so their choices are replaced by a blue line with value one. In contrast, Hilda and Hazel (bottom right) are not linked by equivalent choices: Hilda is Hazel's first choice but Hazel is Hilda's second choice. Unequal mutual choices are represented by two arcs, whereas equivalent mutual choices are represented by a line. As a rule, this will be done throughout the book.

A sociogram depicts the structure of relations within a group. It shows 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. But none of these four girls receives any choices. 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 'popular' girls like Adele and Marion. A simple count of choices does not reveal this, but a sociogram does.

The sociogram has proved to be an important analytic tool that helped to reveal several structural features of social groups. In this book, we will 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 relations. Actually, the concept of a social network is attributed to the anthropologist J. Barnes in the 1950s. Anthropologists study kinship relations, friendship, and gift-giving among people rather than sociometric choice. Furthermore, social psychologists focus on affections, political scientists study power relations among people, organizations, or nations, and economists investigate trade and organizational relations between firms. In this book, the word **actor** refers to a person, organization, or nation which is involved in a social relation. We may say that social network analysis studies the social relations between actors.

The **main goal** of social network analysis is detecting and interpreting patterns of social relations between actors.

This book deals with exploratory social network analysis only. This means that we do not have specific hypotheses about the structure of a network beforehand, which we can test. For example, a hypothesis on the dining-table partners network could predict a particular rate of mutual choices, e.g., one out 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 will not use hypothesis testing here, because we cannot assume prior research experience in a course book and because the statistical models involved are complicated. Therefore, we adopt an exploratory approach which assumes that the structure or pattern of relations in a social network is meaningful to the members of the network and, hence, to the researcher. Instead of testing pre-specified structural hypotheses, we will explore social networks for meaningful patterns.

For similar reasons, we will not pay attention to the estimation of network features from samples. In network analysis, estimation techniques are even more complicated than estimation in statistics, since 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 partners network at random and focus on the choices among them . You will find less choices per person than the two choices in the overall network for the simple reason that choices to girls outside the sample are neglected. Even in this simple respect, a sample is not representative of a network.

Instead of samples, we will analyze entire networks. However, what is the entire network? Sociometry assumed that society consists of interrelated groups, so a network encompasses society at large. Research on the so-called Small

World problem showed that ties of acquaintanceship connect us to almost every human being on the 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 which we will study. For example, we may study the girls of one dormitory only. We do not know their preferences for table partners in other dormitories. Maybe 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.

Since boundary specification may seriously affect the structure of a network, it is important to consider it carefully. Use substantive arguments to support your decision on 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 getting started with the software which accompanies this book.

1.3.1   Network definition

In order to analyze a network, we must first have one. What is a network? Here, and elsewhere, we will use a branch of mathematics called graph theory to define concepts. Most characteristics of networks that we will introduce in this book originate from graph theory. Although this is not a course in graph theory, you should study the definitions carefully in order to understand what you are doing when you apply network analysis. 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 and all it needs for this is a set of vertices (which are also called points or nodes) and a set of lines where each line connects 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 girl in the dormitory, an organization or a country. A vertex is usually identified by a number.

A **line** is a relation 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 which connects a vertex to itself. In the dining-table partners network, loops do not occur because girls are not allowed to choose themselves as a dinner-table partner. We will find out, however, that 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 1.

A **directed graph** or **digraph** contains one or more arcs. A social relation which is undirected, e.g., *is sister of*, is represented by an edge because both individuals are equally involved in the relation. An **undirected graph** does not contain 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 relation (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 may usually be replaced by a **bi-directional arc**: if Ella and Ellen are sisters (undirected), we may say that Ella is the sister of Ellen and Ellen is the sister of Ella (directed). It is important to note this, as we will see in later chapters.

The dining-table partners network has no **multiple lines** since a girl was not allowed to nominate the same girl as first and second choice. Without this restriction, which was imposed by the researcher, multiple arcs could have occurred and they actually do occur in several 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 does not contain 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** does not contain 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 relations.

In the dining-table partners network, the names of the girls represent additional information on the vertices which turns the graph into a network. Thanks to this information, we can tell Ella from Ellen in the sociogram. The numbers which are printed near the arcs and edges offer additional information on the links between the girls: a one indicates a first choice and a two represents a second choice. They are called **line values** and they usually indicate the strength of a relation.

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

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 does not contain loops. Several analytic techniques which we will discuss assume that loops and multiple lines are absent

from a network. However, we will not spell out these properties of each network but we will simply indicate whether it is simple or not. Take care!

Application

In this book, we will learn social network analysis by doing it. We will use the computer program Pajek - Slovenian for spider - to analyze and draw social networks. The software can be installed on your computer from the CD ROM which accompanies this book (see Appendix 1).

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 2, note that part of the vertices and arcs are replaced by […]). Open the file `Dining-table_partners.net` in a word processor 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 "") and three real numbers between 0 and 1, which indicate the position of the vertex in 3-dimensional space if the network is drawn. We will 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, 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.

```
*Vertices   26
   1 "Ada"              0.1646 0.1077 0.5000
   2 "Cora"             0.0481 0.3446 0.5000
   3 "Louise"           0.3472 0.0759 0.5000
   4 "Jean"             0.1063 0.6284 0.5000
[…]
  25 "Laura"            0.5101 0.6557 0.5000
  26 "Irene"            0.7478 0.9241 0.5000
 *Arcs
   1    3     2
   1    2     1
   2    1     1
   2    4     2
   3    9     1
   3   11     2
[…]
  25   15     1
  25   17     2
  26   13     1
  26   24     2
 *Edges
```

Figure 2 - Partial listing of a network data file for Pajek.

The list of vertices is followed by a list of arcs. 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 like graph theory, Pajek defines a line as a

pair of vertices. In Figure 2, the first arc represents Ada's choice of Louise as a dining-table partner. This is her 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 which identify an edge is disregarded in computations. In this data file, no edges are listed.

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 2, represents the structure of the network. The vertex labels, coordinates, and line values (in italics) specify the additional properties of vertices and lines which 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 line values to one if they are not specified in the data file. In addition, Pajek can use several other data formats, which we will not discuss. They are listed in Appendix 1.

Figure 3 - Pajek Main screen.

File>Read We will explain how to create a new network later. Let us first have a look at the network of the dining-table partners. First, start Pajek by double-clicking the file `Pajek.exe` on your hard disk. The computer will display the Main screen of Pajek (Figure 3). From this screen, you can open the dining-table partners network with the *Read* command in the *File* menu or by clicking the button with an icon of a folder under the word *Network*. In both cases, the usual Windows file dialog box appears with which you can search and select the file `Dining-table partners.net` from the `\Chapter1` directory on the CD ROM.

Network drop list   When a network is read by Pajek, its name is displayed in the Network drop list. This drop list is a list of the networks that are accessible to Pajek. You can open a drop list 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 `Dining-table partners.net` in Figure 3. 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 dining-partners network for you.

The Main screen displays five more drop lists beneath the Network drop list. Each of these drop lists represents a data object in Pajek: partitions, permutations, clusters, hierarchies, and vectors. Later chapters will familiarize you with these data objects. Just note that each object can be opened, saved or edited from the *File* menu or using the three icons left of a drop list (see section 1.4).

We advise you to open the network in Pajek and carry out the commands which we will discuss in subsequent sections. This will familiarize you with the concepts discussed as well as with Pajek.

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 first choices only in the dining-table partners network. Finally, some analytic 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 will encounter several techniques to modify a network or select a subnetwork. Network manipulation always results in a new network. In general, many commands in Pajek produce new networks or other data objects rather than graphical or tabular output. This motivates the central place of the data object drop lists in the main menu of Pajek.

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, e.g., the *Net menu* contains all commands that operate on one network and the *Nets* 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.

An example may illustrate this. Suppose we want to change reciprocated choices in the dining-table partners network into edges. Because this operation concerns one network and no other data objects, we must look for it in the *Net* menu. If we left-click on the word *Net* in the upper left of the Main screen, a drop down menu is displayed. Position the cursor on the word *Transform* in the drop down menu and a new submenu is opened with a command to change arcs into edges (*Arcs→Edges).* Finally, we reach the option allowing us to change bi-directional arcs into edges and to assign the smallest line value of the two arcs to the new edge which will replace them (see Figure 4). In this book, we will abbreviate this sequence of commands to:

[Main]Net>Transform>Arcs→Edges>Bidirected only>Min Value.
The screen or window that contains the menu is presented between square brackets and transition to a submenu is indicated by > . The screen name is specified only if the context is ambiguous. The abbreviated command is also displayed in the margin (see above) for the purpose of quick reference.
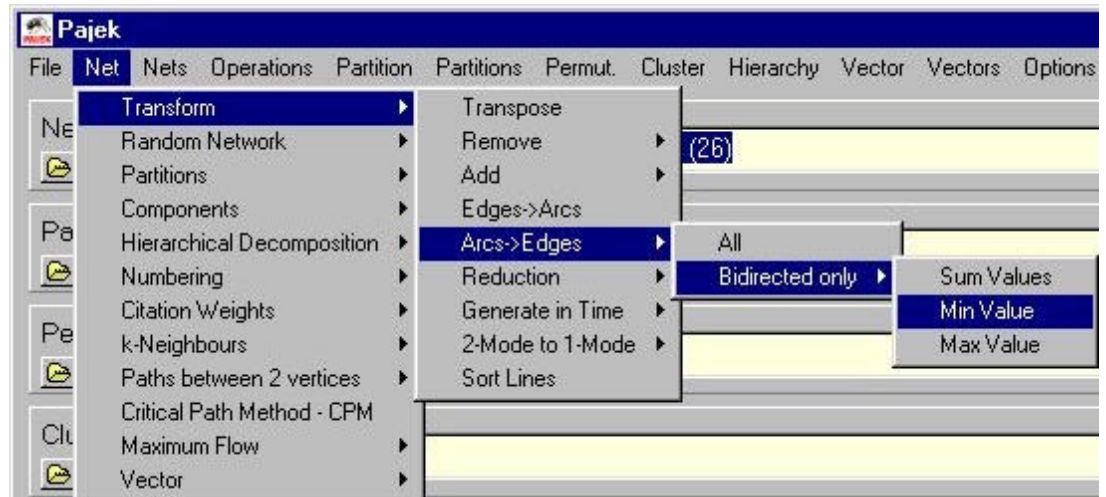
Figure 4 - Menu structure in Pajek.

When the command to change arcs into edges is executed, an information box appears asking whether a new network must be made (Figure 5). If the answer is yes, a new network named `Bidirected Arcs to Edges of N1 (26)` is added to the Network drop list with serial number 2. On the other hand, answering no to the question in the information box causes Pajek to change the original network. This example highlights the use of menus in Pajek and their notation in this book.

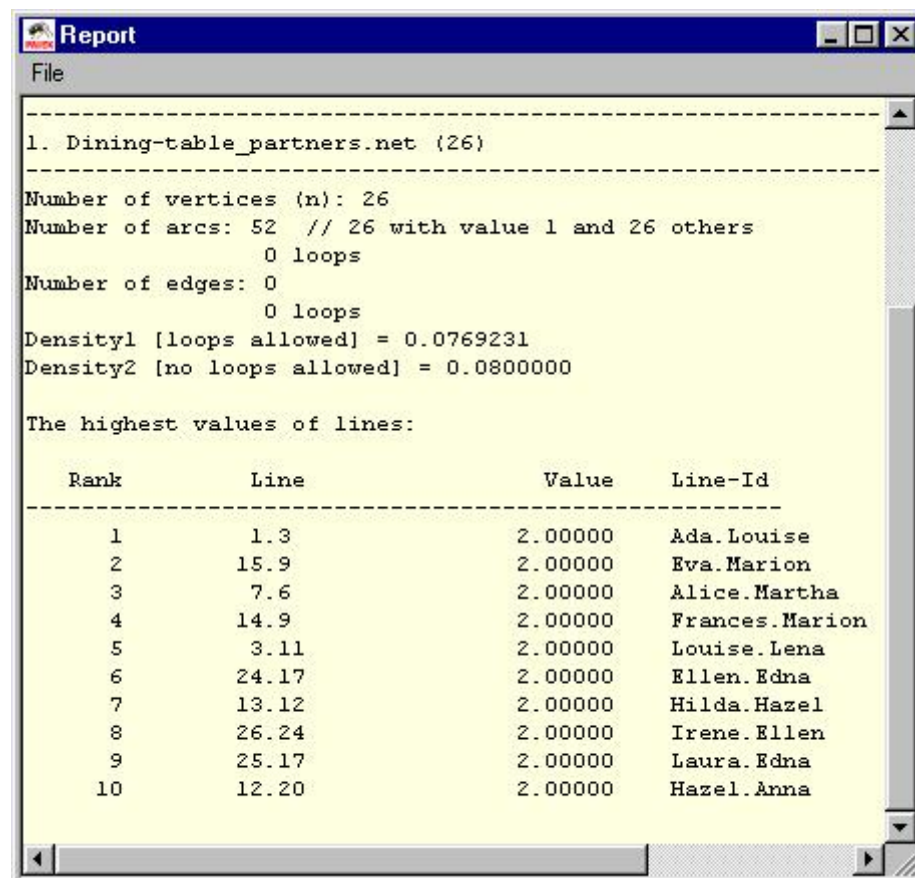**Figure 5** - An information box in Pajek.

11

1.3.3   Calculation

In social network analysis, several structural features have been quantified, e.g. an index that measures the centrality of a vertex. Some measures pertain to the entire network, whereas others summarize the structural position of a subnetwork or a single vertex. Calculation outputs a single number in the case of a network characteristic and a series of numbers in the case of subnetworks and vertices.

Exploring network structure by calculation is much more concise and precise than visual inspection. However, structural indices are sometimes abstract and difficult to interpret. Therefore, we will use both visual inspection of a network and calculation of structural indices to analyze network structure.

Application

Report screen   In Pajek, results of calculations are reported in a separate window which we will call the Report screen. This screen displays numeric results that summarize structural features as a single number, a frequency distribution, or a cross-tabulation. Calculations that assign a value to each vertex are not reported in this screen. They are stored as data objects in Pajek, notably partitions and vectors (see Chapter 2). The Report screen displays text but no network drawings. The contents of the Report screen can be saved as a text file from its *File* menu.


```
----------------------------------------------------------
1. Dining-table_partners.net {26}
----------------------------------------------------------
Number of vertices (n): 26
Number of arcs: 52   // 26 with value 1 and 26 others
                0 loops
Number of edges: 0
                0 loops
Density1 [loops allowed] = 0.0769231
Density2 [no loops allowed] = 0.0800000

The highest values of lines:

   Rank          Line              Value      Line-Id
   ----------------------------------------------------
      1           1.3             2.00000     Ada.Louise
      2          15.9             2.00000     Eva.Marion
      3           7.6             2.00000     Alice.Martha
      4          14.9             2.00000     Frances.Marion
      5           3.11            2.00000     Louise.Lena
      6          24.17           2.00000     Ellen.Edna
      7          13.12           2.00000     Hilda.Hazel
      8          26.24           2.00000     Irene.Ellen
      9          25.17           2.00000     Laura.Edna
     10          12.20           2.00000     Hazel.Anna
```

Figure 6 - Report screen in Pajek.

Info>Network>General   The Report screen depicted in Figure 6 shows the number of vertices, edges, and arcs in the original dining-table partners network. This is general information on

the network that is provided by the command *Info>Network>General* (as you know now, this means the *General* command within the *Network* submenu of the *Info* menu). Besides the number of vertices, edges, and arcs, the screen also shows the number of loops and two indices of network density that will be explained in Chapter 3. Also, this command displays the number of lines requested in a dialog box which opens when the *Info* command is selected (Figure 7).
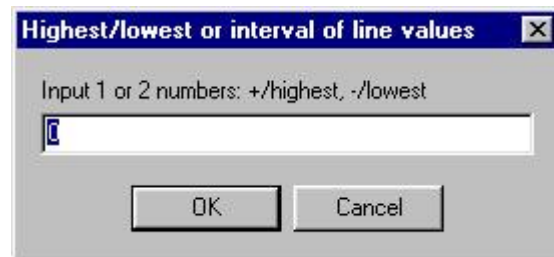
Figure 7 - Dialog box of *Info>Network>General* command.

In this example, we typed 10 in the dialog box, so the report screen shows 10 lines with the highest line values, which are second choices in the dining-table partners network. In the report screen, each line is described by its rank according to its line value, a pair of vertex numbers, the line value, and a pair of vertex labels. Hence, the first line in Figure 6 represents the arc from Ada (vertex number 1) to Louise (vertex number 3), which is Ada's second choice. To get a list of all first choices in the dining-table partners network, we should have typed -26 or the range 27 52, which contains the arcs on ranks 27 up to and including 52, in the dialog box.

1.3.4   Visualization

The human eye is trained in pattern recognition. Therefore, network visualizations help to trace and present patterns of relations. In section 1.2, we presented the sociogram as the first systematic visualization of a social network. It was the sociometrists' main tool to explore and understand the structure of relations in human groups. In books on graph theory, visualizations are used to illustrate concepts and proofs. Visualizations facilitate an intuitive understanding of network concepts, so we will use them a lot.

Our eyes are easily fooled, however. A network can be drawn in many ways and each drawing stresses other structural features. Therefore, the analyst should be careful and rely on systematic rather than ad hoc principles for network drawing. In general, we should use automatic procedures which generate an optimal layout of the network when we want to explore network structure. Subsequently, we may edit the automatically generated layout manually if we want to present it.

Some basic principles of network drawing should be observed. The most important principle states that the distance between vertices should express the strength or number of their relations as closely as possible. In a map, the distance between cities matches their geographical distance. In psychological charts, spatial proximity of objects usually expresses perceived similarity. Since social

network analysis focuses on relations, a drawing should position vertices according to their relations: vertices which are connected should be drawn closer together than vertices which are not related. A good drawing minimizes the variation in the length of lines. In the case of lines with unequal values, line length should be proportional to line value.

The legibility of a drawing poses additional demands, which are known as graph drawing esthetics. Vertices or lines should not be drawn too closely together and small angles between lines which are incident with the same vertex should be avoided. Distinct vertices and distinct lines should not merge into one lump. Vertices should not be drawn on top of a line which belongs to other vertices. The number of crossing lines should be minimized because the eye tends to see crossings as vertices.

Application

Draw screen   Pajek offers many ways to draw a network. It has a separate window for drawing, which is accessed from the *Draw>Draw* menu in the Main screen. The Draw screen has an elaborate menu of choices (see Figure 8), some of which will be presented in later chapters. Use the index of Pajek commands in this book to find them. We will discuss the most important commands now to help you draw your first network.
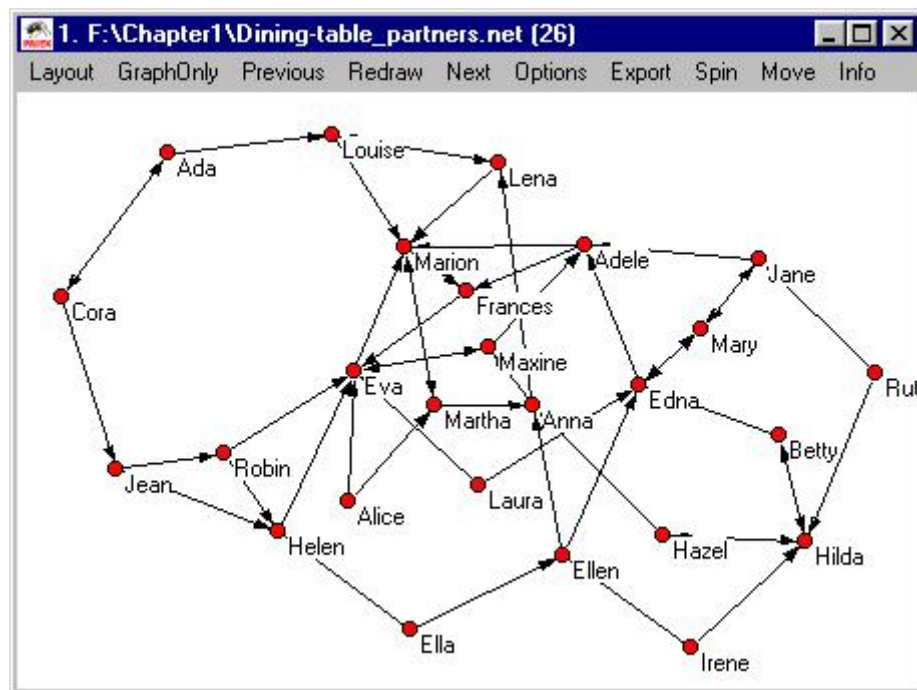
Figure 8 - Draw screen in Pajek.

1.3.4.1   Automatic drawing

Automated procedures for finding an optimal layout are a better way to obtain a basic layout than manual drawing, because the resulting picture depends less on the preconceptions and misconceptions of the investigator. Besides, automated

drawing is much faster and quite spectacular because the drawing evolves before your eyes.

In Pajek, several commands for automatic layout are implemented. Two commands are accessible from the *Layout>Energy* menu and we will refer to them as **energy commands**. Both commands move vertices to locations that minimize the variation in line length. Imagine that the lines are springs which pull vertices together. The energy commands 'pull' vertices to better positions until they are in a state of equilibrium. Therefore, these procedures are known as spring embedders.

The relocation technique used in both automatic layout commands has some limitations. First, the results depend on the starting positions of vertices. Different starting positions may yield different results. Most results will be quite similar, but results can differ markedly. The *Starting positions* submenu on the *Layout>Energy* menu gives you control over the starting positions. You can choose random and circular starting positions, or you can use the present positions of vertices as their starting positions: option *Given xy*, where *x* and *y* refer to the (horizontal and vertical) coordinates of vertices in the plane of the Draw screen. The fourth option (*Given z*) refers to the third dimension, which we will present in Chapter 5.

Figure 9 - Continue dialog box.

The second limitation of the relocation technique is that it stops if improvement is relatively small or if the user so requests in a dialog box (Figure 9). This means that automatic layout generation outputs a drawing that is very good but not perfect. Manual improvements can be made. However, make small adjustments only to reduce the risk of discovering a pattern which you introduced to the drawing yourself. It is usually worth the effort to repeat an energy command a couple of times with given starting positions to improve a drawing.

Because of these limitations, take the following advice to heart: *never rely on one run of an energy command*. Experiment with both commands and manual adjustment until you obtain the same layout time and again.

The first energy command is named Kamada-Kawai after its authors. This command produces regularly spaced results, especially for connected networks which are not very large. It seems to produce more stable results than the other energy command (Fruchterman Reingold) but it is much slower and it should not be applied to networks which contain more than 500 vertices. With Kamada-Kawai, commands *Fix first and last* and *Fix one in the middle* enable you to fix vertices that should not be relocated. The first command fixes the two vertices with lowest and highest serial numbers. This is very useful if these vertices

represent the sender and target in an information network. The second command allows you to specify one vertex that must be placed in the middle of the drawing.

*Layout>Energy
>Kamada-Kawai
>Selected group only*

The Kamada-Kawai command can also be applied to a section of a network. First, use your right mouse button (see above) to zoom in on the section which you want to energize. Second, select the only *Kamada-Kawai* submenu command visible after zooming in, which reads *Selected group only*. After relocation of the selected vertices, Pajek asks you whether it should resize the optimized section to the original network. *Yes* is usually the right answer to this question.

*Layout>Energy
>Fruchterman Reingold
>2D, 3D*

The second energy command, which is called Fruchterman Reingold, is faster and it also works with larger networks. This command separates unconnected parts of the network nicely, whereas Kamada-Kawai draws them on top of one another. It can generate two and three dimensional layouts as indicated by commands *2D* and *3D* in the submenu. We will postpone discussion of three dimensional visualizations until Chapter 5.

*Layout>Energy
>Fruchterman Reingold
>Factor*

The third command in the submenu, which is labeled *Factor*, allows the user to specify the optimal distance between vertices in a drawing which is being energized with Fruchterman Reingold. This command displays a dialog box asking for a positive number. A low number yields small distances between vertices and many vertices are placed in the center of the plane. A high number pushes vertices out of the center towards positions on a circle. An optimal distance of 1 is a good starting point. Try a smaller distance if the center of the drawing is quite empty, but try a higher distance if the center is too crowded.

Each energy command has strong and weak points. We advise starting with Fruchterman Reingold and using several optimal distances until a stable result appears. The drawing can then be improved using Kamada-Kawai with the actual location of vertices as starting positions. Finally, improve the drawing by manual editing, which we will discuss in the following section.

1.3.4.2   Manual drawing

Pajek supports manual drawing of a network. Use the mouse to drag individual vertices from one position to another. Place the cursor on a vertex, hold down the left mouse button, and move the mouse to drag a vertex. As you will see, the lines which are incident with the vertex move too. You can drop the vertex any place you want, unless you constrain the movement of vertices with the options in the *Move* menu.

*Move>Fix,
Grid,
Circles*

The *Move* menu in the Draw screen has (at least) three options: *Fix*, *Grid*, and *Circles*. The *Fix* option allows you to restrict the movement of a vertex. It cannot be moved horizontally if you select *x* in the drop down menu, or vertically (select *y*). Select *Radius* to restrict the movement of a vertex to a circle. The options *Grid* and *Circles* allow you to specify a limited number of positions to which a vertex can be moved. In the case of small networks, this generates esthetically pleasing results. In Pajek, when you select an option, it is marked by a dot and it is effective until it is selected again (Figure 10).

Figure 10 - A selected option in the Draw screen.

Redraw

Previous

Next

GraphOnly

You can zoom in on the drawing by pressing the right mouse button and by dragging a rectangle over the area which you want to enlarge. Within the enlargement, you can zoom in again. To return to the entire network, you must select the *Redraw* command on the Draw screen menu. It is not possible to zoom out in steps. The *GraphOnly* menu is similar to the *Redraw* menu, except that *GraphOnly* removes all labels as well as the heads of the arcs. It shows vertices and lines only. This option speeds up automatic layout of a network, which is particularly helpful for drawing large networks. The *Previous* and *Next* commands on the menu allow you to display the network before or after the current network in the *Network* drop list, so you need not return to the Main screen first to select another network. Note, however, that this is true only if the option *Network* is selected in the *Previous/Next>Apply to* submenu of the *Options* menu.
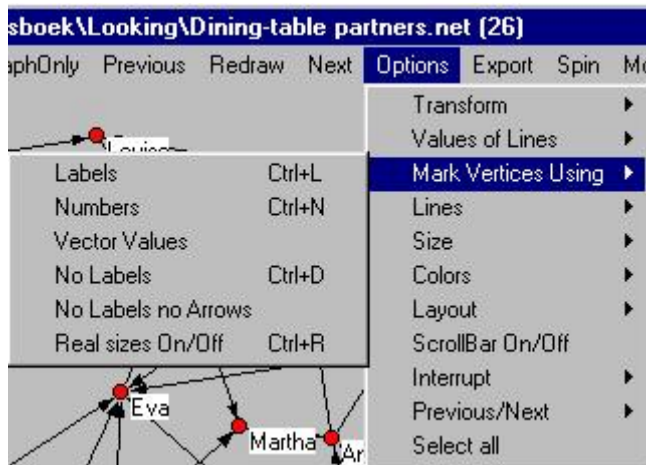
Figure 11 - *Options* menu of the Draw screen.

Options menu:

Transform,

Layout,

Mark vertices using

The *Options* menu offers a variety of choices for changing the appearance of a network in the Draw screen and for setting options to commands in other Draw screen menus (Figure 11). Many drawing options are self-explanatory. Options for changing the shape of the network are listed in the *Transform* and *Layout* submenu, whereas options for the size, color, and labeling of vertices and lines are found in several other submenus. Figure 11 shows the *Options>Mark vertices using* submenu and its options for changing the type of vertex label displayed. Note the keyboard shortcuts in this submenu: pressing the control key along with the L, N, or D key has the same effect as selecting the accompanying item in the submenu.

[Draw]Info

To improve your drawing, Pajek can evaluate a series of esthetic properties, such as closest vertices which are not linked, or the number of crossing lines.

This allows you to find the worst aspects of a drawing and to improve them manually. The *Info* menu in the Draw screen allows you to select one or all esthetic properties. If you select one of these commands, Pajek identifies the vertices which perform worst: they are identified by a color (other than light blue) in the drawing and they are listed in the Report screen. When you select all esthetic properties, each property will be associated with a color (see Figure 12 for esthetic information on the dining-table partners network as drawn in Figure 8). Sometimes, vertices violate several esthetic indices, so they ought to be marked by several colors. Since this is not possible, some colors may not show up in the drawing.

```
-------------
 Layout Info
-------------
YELLOW: The closest vertices: 6 and 15. Distance: 0.07873
GREEN: The smallest angle: 2.1.2. Angle: 0.00000
RED: The shortest line: 9.14. Length: 0.09809
BLUE: The longest line: 11.20. Length: 0.40417
MAGENTA: Number of crossings: 13
WHITE: Closest vertex to line: 6 to 15.25. Distance: 0.02139
```

Figure 12 - Textual output from *Info>Check*.

1.3.4.3   Saving a drawing

In order to present pictures of networks to an audience, we have to save our visualizations. This subsection sketches the options for exporting network drawings from Pajek.

[Main]File>Network >Save   If a researcher wants to save a layout for future use, the easiest thing to do is to save the network itself. Remember that a network consists of lists of vertices, arcs, and edges. The list of vertices specifies a serial number, a label, and coordinates in the plane for each vertex. Relocation of a vertex in the Draw screen changes its coordinates and the latest coordinates are written to the network file on execution of the *Save* command. When the network is reopened in Pajek, the researcher obtains the layout s/he saved, that is to say, the general pattern because the size and colors of vertices and lines are usually not specified in the network data file.

Export menu   Pajek offers several options to save the drawing as a picture for presentation to an audience. The options are listed in the *Export* menu of the Draw screen. Three options produce two-dimensional output (EPS/PS, SVG, and Bitmap) and another three options yield three-dimensional output (VRML, MDL MOLfile, and Kinemages). Although three-dimensional representations can be quite spectacular (Figure 13), we will not discuss them now. We will postpone the third dimension until Chapter 5. We will merely sketch the commands for two-dimensional output and we refer to Appendix 2 for more details and additional viewers.

Export>Bitmap   The *Bitmap* export option produces an image of the Draw screen: each screen pixel is represented by a point in a raster, which is called a bitmap. You get exactly what you see, even the size of the picture matches the size of the Draw

screen. Unfortunately, this means loops are invisible and that reciprocal choices are represented by one line with two arrowheads rather than two separate arcs. Every word processor and presentation program operating under Windows can load and display bitmaps. Bitmaps, however, are cumbersome to edit and they lose their sharpness if they are enlarged or reduced, e.g., see Figure 8.
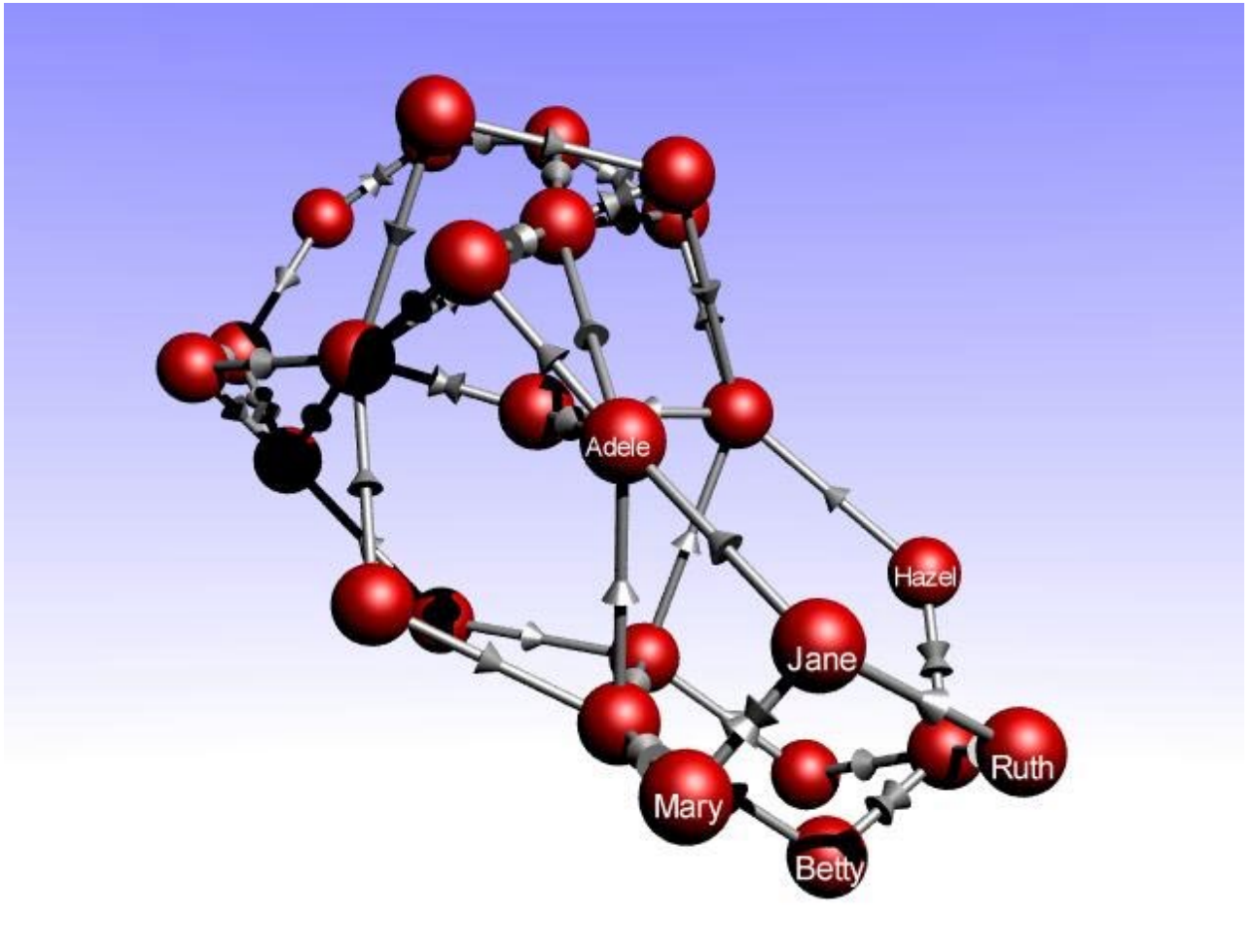
Figure 13 - A 3D rendering of the dining-table partners network.

*Export>EPS/PS',
SVG,
Options*

Vector graphics produce more pleasing results than bitmaps. In a vector graphic, the shape and position of each circle representing a vertex, each line, arc, loop, or label is specified. Because each element in the drawing is defined separately, a picture can be enlarged or reduced without loss of quality and its layout can be manipulated easily. Pajek can export vector graphics in two formats, which are closely related: PostScript (command *EPS/PS*) and Scalable Vector Graphics (command *SVG*). PostScript (PS) and Encapsulated PostScript (EPS) are meant for printing, whereas the Scalar Vector Graphics format is developed for the web. The user can modify the layout of both PostScript and Scalar Vector Graphics drawings in the *Export>Options* submenu, which is covered in detail in Appendix 2.

In this book, we use vector graphics because of their high quality. For example, the sociogram shown in Figure 1 was exported from Pajek as

Encapsulated PostScript. However, PostScript and Scalar Vector Graphics are formats that are not universally supported by Windows software. For example, MS Word 97 cannot display Encapsulated PostScript. Appropriate software, e.g., Adobe Illustrator or CorelDraw, is necessary to edit vertices, labels, and lines, or to convert a vector graphic to a format which your word processor can handle (e.g., Windows MetaFile). To display Scalar Vector Graphics in a web-page, a plug-in is needed which can be downloaded from the website of Adobe Systems Incorporated (see Appendix 2 for details).

1.4   *Assembling a social network*

In order to perform network analysis, social relations must be measured and coded. In this section, we will briefly discuss data collection techniques for social network analysis and explain how to convert data to a network file for Pajek.

There are several ways to collect data on social relations. Traditionally, sociometrists focus on the structure of social choice within a group. They gather data by asking each member of a group to indicate his or her favorites (or opponents) with respect to an activity that is important to the group. For example, they ask pupils in a class to name the children whom they prefer to sit next to. In a questionnaire, respondents may write down the names of the children they choose, or check their names in a list. These methods are called *free recall* and *roster* respectively. The latter method reduces the risk that respondents may overlook people.

Sometimes, the respondent is asked to nominate a fixed number of favorites. In sociometry, it was very popular to restrict the number of choices to three. For example, in the dining-table partners network each girl was asked to make three choices. This restriction is motivated by the empirical discovery that the more choices that are allowed the more they concentrate on people who are already highly chosen. When asked for their best friends, most people mention four people or less. If they have to mention more people, they usually nominate people whom they think they should like because they are liked by many others. However, restricting the number of choices reduces the reliability of the data: choices are less stable over time and correlate less well than other measurement techniques such as unrestricted choices, ranking (rank all other group members with respect to their attractiveness), or paired comparison (list all possible pairs of group members and choose a preferred person in each pair). A researcher who fixes the number of choices available to respondents eliminates the difference between a respondent who entertains many friendships and a loner.

Fixed and free choice, ranking, and paired comparison are techniques that elicit data on social relations through questioning. However, there are several data collection techniques that register social relations rather than elicit them. For example, the amount of interaction between pupils in a class may be observed by a researcher. Respondents may be asked to register their contacts in a diary, membership lists and files that log contacts in electronic networks may be coded, family relations or transactions can be retrieved from archives and databases, et

cetera. The rapid growth of electronic data storage offers new opportunities to gather data on large social networks.

Indirect data are usually better than reported data, which rely on the often inaccurate recollections of respondents. However, it is not always easy to identify people and organizations unambiguously in data collected indirectly: is a Mr. Jones on the board of one organization the same person as the Mr. Jones who is CEO of another firm? It goes without saying that network analysis demands correct identification of vertices in the network.

Application

Now that you have collected your data, it is time to create a network that can be analyzed with Pajek. In Section 1.3.1, we discussed the structure of a Pajek network file. This is a simple text file, which can be typed out in any word processor that exports plain text. Do not forget to attribute serial numbers to the vertices ranging from 1 to the number of vertices. Save the network from the word processor as plain, unformatted text (DOS text, ASCII) and use the extension *.net* in the file name.
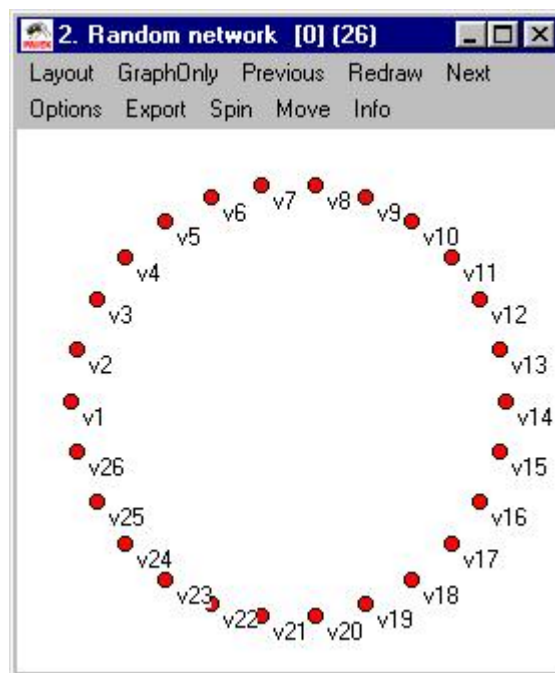
Figure 14 - Random network without lines.

It is also possible to produce the network file in Pajek. First, make a new random network with the command *Net>Random Network>Total No. of Arcs*. In the first dialog box, type in the number of vertices you want. In the second dialog box, request zero arcs to obtain a network without lines. The command creates a new network and adds it to the *Network* drop list in the Main screen. When you draw it (*Draw>Draw*), you will see that the vertices are nicely arranged in a circle (Figure 14).
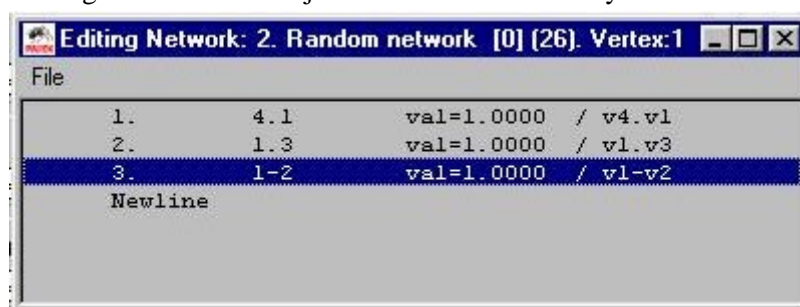
As a second step, add lines to the network, which may be done in the Main screen or in the Draw screen. In the Main screen, both the *Edit* button at the side

of the *Network* drop list (a picture of a writing hand) and the command *File>Network>Edit* open a dialog box which allows you to select a vertex by serial number or by label. Next, the Editing Network screen is shown for the selected vertex (Figure 15). In the Draw screen, you can open the Editing Network screen by right-clicking a vertex.

Double-clicking the word *Newline* in the Editing Network screen opens a dialog box which allows the user to add a line to or from the selected vertex. To add an edge type the number of another vertex. Type the number of the vertex preceded by a + sign to add an arc *to* the selected vertex and use a - sign to add an arc *from* the selected vertex. Each new edge and arc is displayed as a line in the Editing Network screen. For example, Figure 15 displays an arc from vertex 4 to vertex 1, an arc from vertex 1 to 3, and an edge (indicated by a dash between vertex numbers) between vertices 1 and 2. Also, you can delete a line in the Editing Network screen: just double-click the line you want to delete.

```
Editing Network: 2. Random network [0] [26]. Vertex:1    _ □ ✕
File

     1.          4.1          val=1.0000   / v4.v1
     2.          1.3          val=1.0000   / v1.v3
     3.          1-2          val=1.0000   / v1-v2
  Newline
```

Figure 15 - Edit Network screen.

By default, all lines have unit line value as indicated by the expression val=1.000 in the Editing Network screen. Line values can be changed in this screen by selecting the line (left-click) and right-clicking it. A dialog box appears that accepts any number (positive, zero and negative) as input.

As a final step, save the network. Networks in Pajek are not automatically saved. Because network analysis usually yields many new networks, most of which are just intermediate steps, Pajek does not prompt the user to save networks. Save the new network as soon as you finish editing it. Use *Save* from the submenu *File>Network* or use the save button (the picture of a diskette) at the side of the Network drop list. We recommend that you save a network in Pajek Arcs/Edges format for easy manual editing and for a maximum choice of layout options (see Appendix 2). Give the file a meaningful name and the extension *.net*.

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 especially, processing the data as a spreadsheet or relational database helps data cleaning and coding. We will, however, not go into details here.

1.5   Summary

This chapter introduces social network analysis and emphasizes its theoretical interest in social relations between people or organizations, and its roots in mathematical graph theory. A network is defined as a set of vertices and a set of

lines between vertices with additional information on the vertices or lines. This flexible definition permits a wide variety of empirical phenomena, ranging from the structure of molecules to the structure of the universe, to be modeled as networks. In social network analysis, we concentrate on relations between people or social entities that represent people, e.g. affective relations between people, trade relations between organizations, or power relations between nations.

This simple definition of a network covers all types of networks which we will encounter in this book: directed and undirected networks, networks with and without loops or multiple lines. Most social networks are simple undirected or directed networks, which do not contain multiple lines, and loops usually do not occur. As a result of transformations, however, networks may acquire multiple lines and loops. The results of analytic procedures may depend on the kind of network we are analyzing, so it is important to know what kind of network it is.

The mathematical roots of network analysis permits powerful and well-defined manipulations, calculations, and visualizations of social networks. Exploratory social network analysis as we present it here makes ample use of these three techniques. In exploring a social network, we will first visualize it to get an impression of its structure. The sociogram, which originates from sociometry, is our main visualization technique.

We are convinced that doing social network analysis is a good way of learning about it. Therefore, the program Pajek for network analysis is an integral part of the course. We urge you to practice the commands demonstrated rather than just read about them. The data related to each example in the book are supplied on the CD ROM. This chapter introduces the menu structure of Pajek and some basic commands for visualizing networks. You are now ready to start analyzing social networks.

1.6   Exercises

1   A sociogram is:
    a   an index of sociability
    b   a graphic representation of group structure
    c   the structure of sociometric choices
    d   a set of vertices connected by edges or arcs
2   Which of the following definitions is correct?
    a   a graph is a set of vertices and a set of lines
    b   a graph is a set of vertices and a set of edges
    c   a graph is a network
    d   a graph is a set of vertices and a set of pairs of vertices

3   Have a look at the networks depicted below and choose the correct statement.
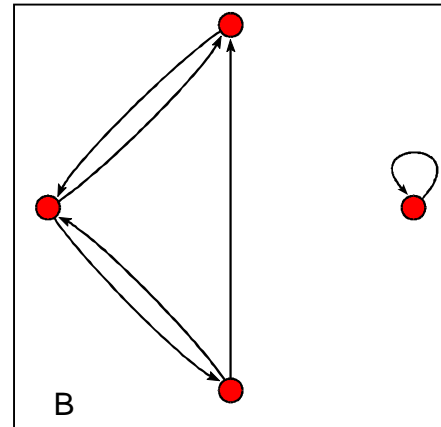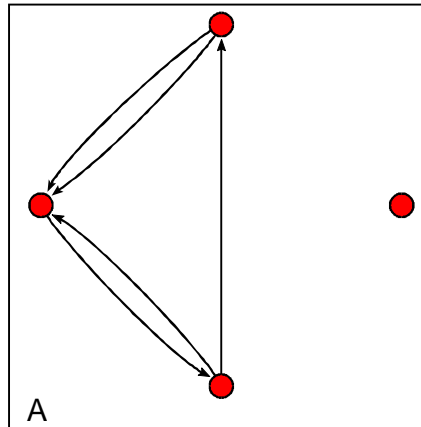
a   neither A nor B is a simple directed graph

b   A is not a simple directed graph but B is

c   A and B are simple directed graphs

d   A and B are networks but not graphs

4   Social network analysis is exploratory if:

a   the researcher has no specific ideas on the structure of the network beforehand

b   it deals with social settings that are unexplored by social scientists

c   the network is studied outside a laboratory

d   it does not try to predict network structure from a sample

5   Which of the following statements is correct?

a   a line can be incident with a line

b   a line can be incident with a vertex

c   an edge can be incident with a line

d   a vertex can be incident with a vertex

6   Open the dining-table partners network and improve it according to the esthetic criterion for minimizing crossing lines. Which vertex can you move to reduce the number of crossing lines?

7   Open the dining-table partners network and remove all second choices with the command *Net>Transform>Remove>lines with value>higher than*.

a   Why is this command part of the *Net* menu?

b   What in your opinion is the most striking result, if you draw the new network?

c   Which energy command do you recommend to optimize this drawing? Explain your choice.

8   Use the *Fruchterman Reingold* command with random starting positions and its *Factor* option to get a sociogram from the dining-table partners network with a clear distinction between vertices in the center and vertices in the margin or periphery. Do you think your drawing is better than the original? Justify your answer.

1.7   Assignment

Your first social network analysis: investigate the friendship network within your class as they are or as you perceive them. Choose the right kind of network to conceptualize friendship relations (directed, undirected, valued, with multiple lines and loops?), collect the data (design a questionnaire or state the friendships which you observe), make a Pajek network file, draw and interpret it. If you use perceived friendship relations, it is worthwhile comparing your network to networks made by other students. Where do they differ?

1.8   Further Reading

- The example of the girls' school dormitory is taken from J.L. Moreno, *The Sociometry Reader*. Glencoe (Ill.), The Free Press, 1960, p. 35.
- A brief history of social network analysis can be found in J. Scott, *Social Network Analysis. A Handbook*. London, Sage Publications, 1991 (2nd ed. 2000), Chapter 2.
- For the problem of boundary specification and sampling see S. Wasserman & K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge, Cambridge University Press, section 2.2, p. 30-35. Information on data collection can be found in section 2.4 (p. 43-59). This impressive monograph is a good starting point for further reading on any topic in social network analysis.

1.9   Answers to questions

1   Answer b is correct because a sociogram is a specific way of drawing the structure of relations within a human group.
2   A graph is defined as a set of vertices and a set of lines between pairs of vertices (see Section 1.3.1), so answer d is correct: lines can be defined as pairs of vertices. Answer a does not specify that the lines must have the graph's vertices as their endpoints, so this answer is not correct. Answer b has the same fault and it restricts the definition to undirected graphs, which is not correct. Answer c ignores the difference between a graph and a network.
3   Answer b is correct. A simple directed graph contains at least one arc (so it is directed) and no multiple lines; it may contain loops. Therefore, network B is a simple directed graph but network A is not due to its multiple lines. Answer d is wrong because there is no additional information on the vertices and arcs, such as vertex labels or line values.
4   Answers a and d are correct.
5   Answer b is correct.
6   Select option *No. of crossings* in the *Info* menu of the Draw screen. The vertices in the center are colored pink now: their lines cross. The Report screen shows the number of crossing lines to be 13. The easiest improvement is to drag Alice to a position above the arc from Laura to Eva. Now, Alice's

arcs do not cross any other and the number of crossing lines is reduced to 12. However, it is even better to drag Alice and Martha to the left of the arc from Francis to Eva, because Martha and Marion are connected by mutual choice, which counts twice at every crossing. If you do this, it is better to drag Maxine below the arc from Martha to Anna, because Eva and Maxine are doubly connected. Another improvement is made if Adele is placed to the left of the arc from Anna to Lena. We reach a minimum of 7 crossing arcs. Can you do better?

7   Note: since first choices have line values 1 and second choices have line values 2, you must type 1 in the dialog box which appears on activation of the command *Net>Transform>Remove>lines with value>higher than*.

   a   This command operates on one object, namely a single network.

   b   The network is no longer connected. Some girls and pairs of girls are disconnected, e.g., Hazel, Cora & Ada, and Jean & Helen.

   c   We prefer Fruchterman Reingold because Kamada-Kawai sometimes draws disconnected parts on top of one another.

8   Setting *Factor* to 2 or higher will yield a drawing with a lot of vertices placed on a circle and few vertices in the middle, notably Anna, Eva, or Edna. This drawing is not an improvement, the selection of central vertices is arbitrary. Repeat the energy command several times (with random starting positions) and you will find different persons in the center. Also, several esthetic criteria are violated more than in the original drawing.