The SAGE Handbook of Social Network Analysis

A Brief Introduction to Analyzing Social Network Data

Contributors: Robert A. Hanneman & Mark Riddle
Edited by: John Scott & Peter J. Carrington
Book Title: The SAGE Handbook of Social Network Analysis
Chapter Title: "A Brief Introduction to Analyzing Social Network Data"
Pub. Date: 2014
Access Date: January 3, 2019
Publishing Company: SAGE Publications Ltd
City: London
Print ISBN: 9781847873958
Online ISBN: 9781446294413
DOI: http://dx.doi.org/10.4135/9781446294413.n23
Print pages: 331-339

A Brief Introduction to Analyzing Social Network Data

Robert A.Hanneman and MarkRiddle

Introduction

Social network analysts use two kinds of tools from mathematics to represent information about patterns of ties among social actors: graphs and matrices. In this chapter, we'll provide a very quick sketch about how social networks can be represented with these tools and we will discuss some of the things that these representations let us see more clearly. We'll extend our look at the “toolkit” of social network analysts by looking at how they approach some of the most commonly asked questions about the texture of whole networks and the ways that individuals are embedded in them.

There is a lot more to these topics than we will cover here. The visual representation of social networks as graphs is discussed in depth in the chapter by Lothar Krempel in this volume. Mathematics has whole subfields devoted to “graph theory” and to “matrix algebra.” Social scientists have borrowed just a few things that they find helpful for describing and analyzing patterns of social relations. Representing data as matrices is the basis for manipulating data and calculating the measures we discuss in the next chapter. The chapter by Pip Pattison (this volume) shows some advanced applications.

By the time you've worked through this chapter and the next, we hope that you will have an introductory understanding of the most commonly used formal representations of social networks and some of the most commonly used basic descriptive indexes. This is only an introduction; longer and more complete presentations are available in Wasserman and Faust (1994), John Scott (2000), and our own online text (Hanneman and Riddle, 2005) (this chapter and the following chapter are cut-down and edited versions).

Working with network data and calculating measures of their properties is almost always done with software. We will present a number of examples using the UCINET package (Borgatti et al., 2002), because we are most familiar with it. There are, however, a number of excellent software tools that you will want to review (see Huisman and van Duijn, this volume) when you want to try your hand.

Using Graphs to Represent Social Relations

A good drawing of a graph can immediately suggest some of the most important features of overall network structure. Are all the nodes connected? Are there many or few ties among the actors? Are there subgroups or local “clusters” of actors that are tied to one another but not to other groups? Are there some actors with many ties, and some with few?

A good drawing can also indicate how a particular “ego” (node) is “embedded” in (connected to) its “neighborhood” (the actors that are connected to ego, and their connections to one another). By looking at “ego” and the “ego network,” we can get a sense of the structural constraints and opportunities that an actor faces; we may be better able to understand the role that an actor plays in a social structure.

Graphs representing networks are composed of nodes (the individual actors) and relations. Either arcs (one-directional arrows) or edges (lines without arrow heads) represent which actors are tied to which others, for asymmetric and symmetric relations, respectively. Figure 23.1 shows an example of information sharing among 10 organizations as studied by Knoke and Wood (1981). (The original versions, in color, of the images in this chapter can be seen at http://faculty.ucr.edu/~hanneman/chapter_23_figures.htm.)

Figure 23.1 Directed graph of information ties (Knoke bureaucracies)

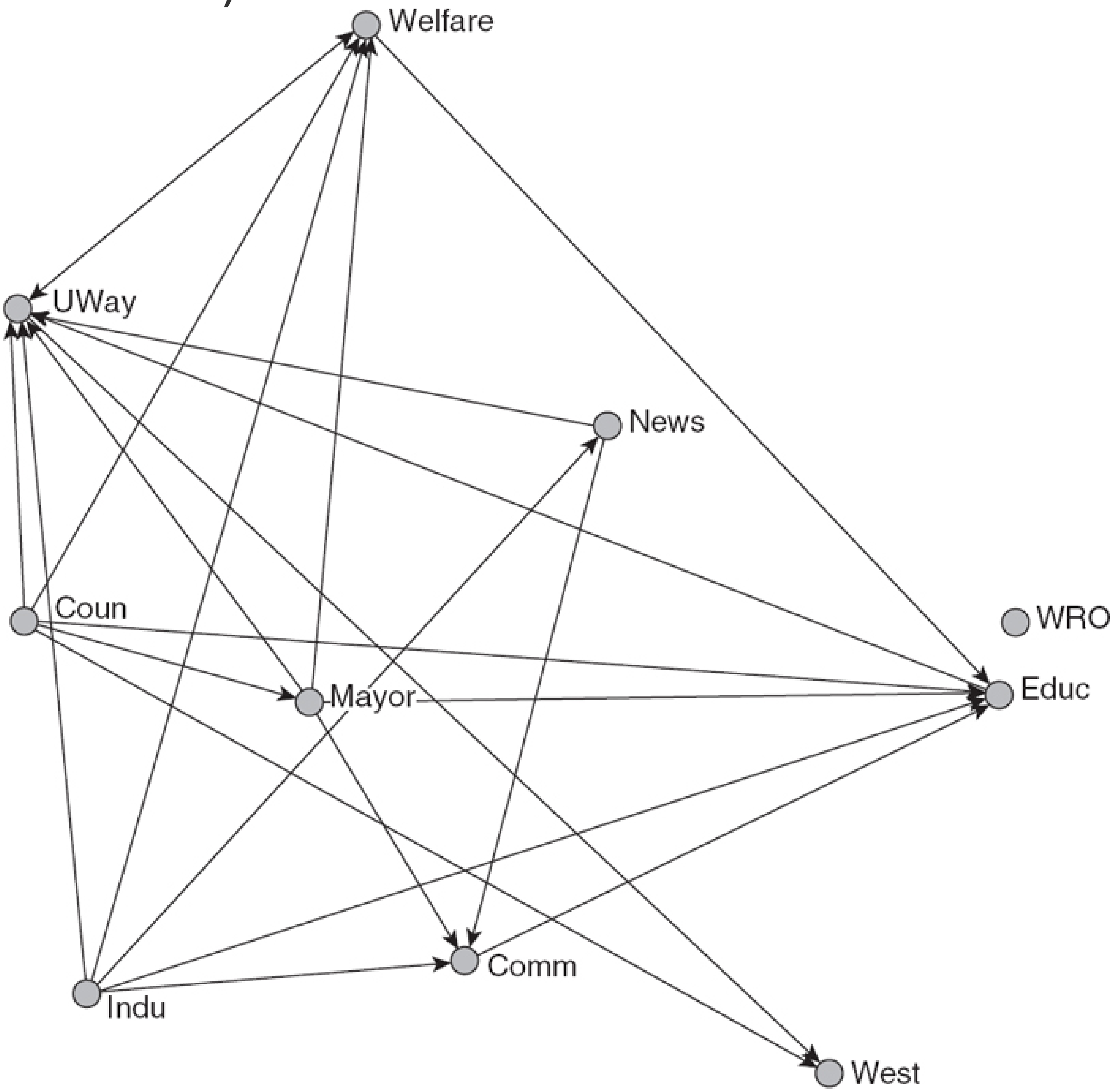


Figure 23.1 is a "directed graph." That is, it provides information that is asymmetric and not necessarily reciprocated. In this example, each node represents an organization, and each relation represents whether or not it provides information to each other organization (if a relation has the value of zero, it is not graphed).

Looking at the "big picture," we note that the "texture" of the network is uneven. All but one organization is connected to others, but overall the density of connections isn't very high. There is a good bit of variability in how connected the organizations are. There seems to be qualitative variation, with different organizations serving as "sources," "receivers," and "transmitters."

Individual organizations are "embedded" in the network in quite different ways. The Educ organization has multiple alternative sources of information from different regions of the network but does not serve as a source of information for others. The WRO (welfare rights organization) is isolated; the UWAY (United Way) seems to be "central."

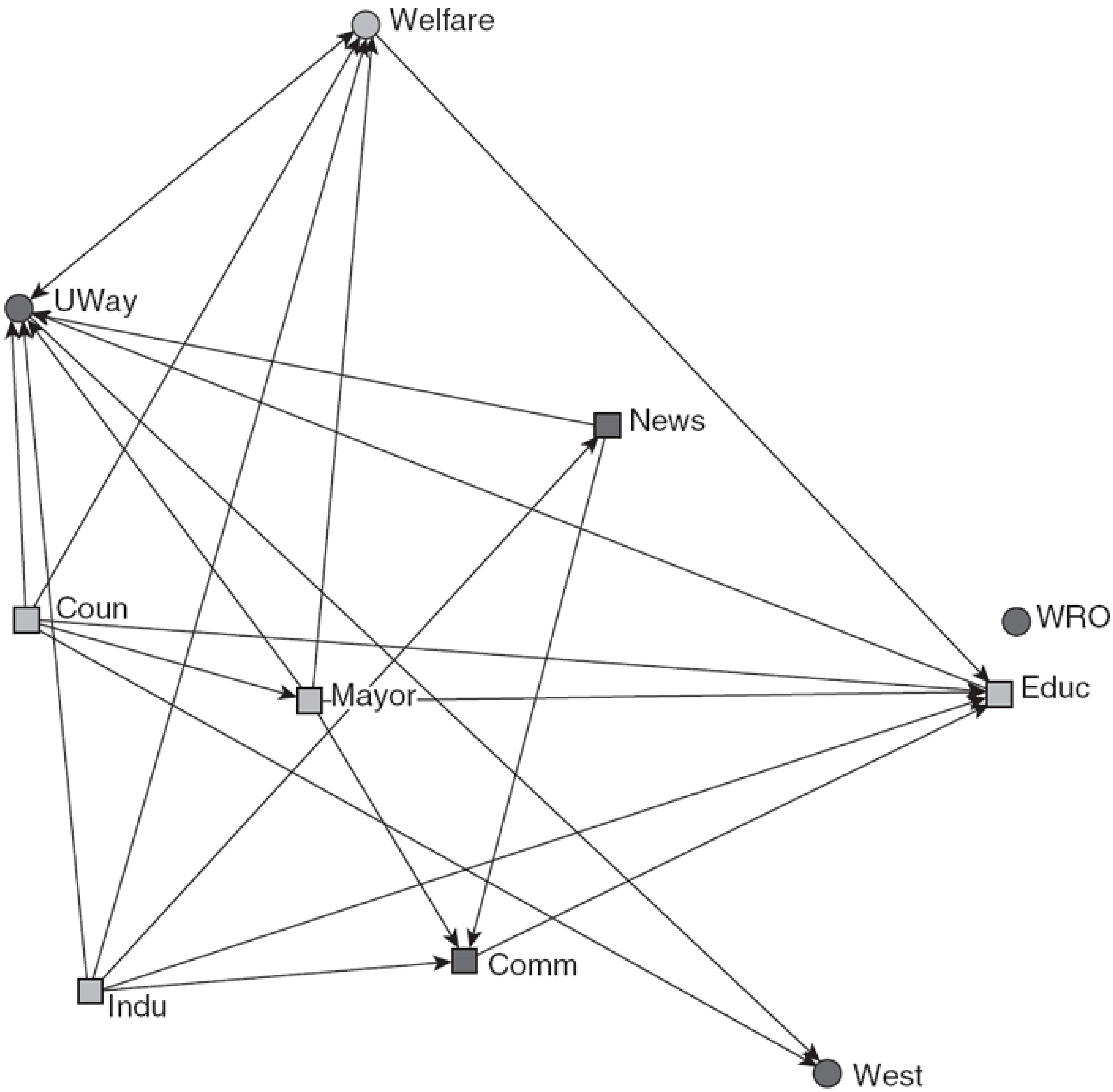
Graphs can be more or less informative and can highlight or obscure features. Krempel (this volume) provides an extended treatment of visualization. But let us identify a few key features of graphs that can be helpful for getting started.

Graphing Node and Relation Attributes

Using colors and shapes are useful ways of conveying information about what "type" of actor each node is. Institutional theory might suggest that information exchange among organizations of the same type would be more common than information exchange between organizations of different types. Some of the organizations here are governmental (Welfare, Coun, Educ, Mayor, Indu); some are nongovernmental (UWay, News, WRO, Comm, West). Ecological theory of organizations suggests that a division between organizations that are "generalists" (e.g., perform a variety of functions and operate in several different fields) and organizations that are "specialists" (e.g., work only in social welfare) might affect information-sharing patterns.

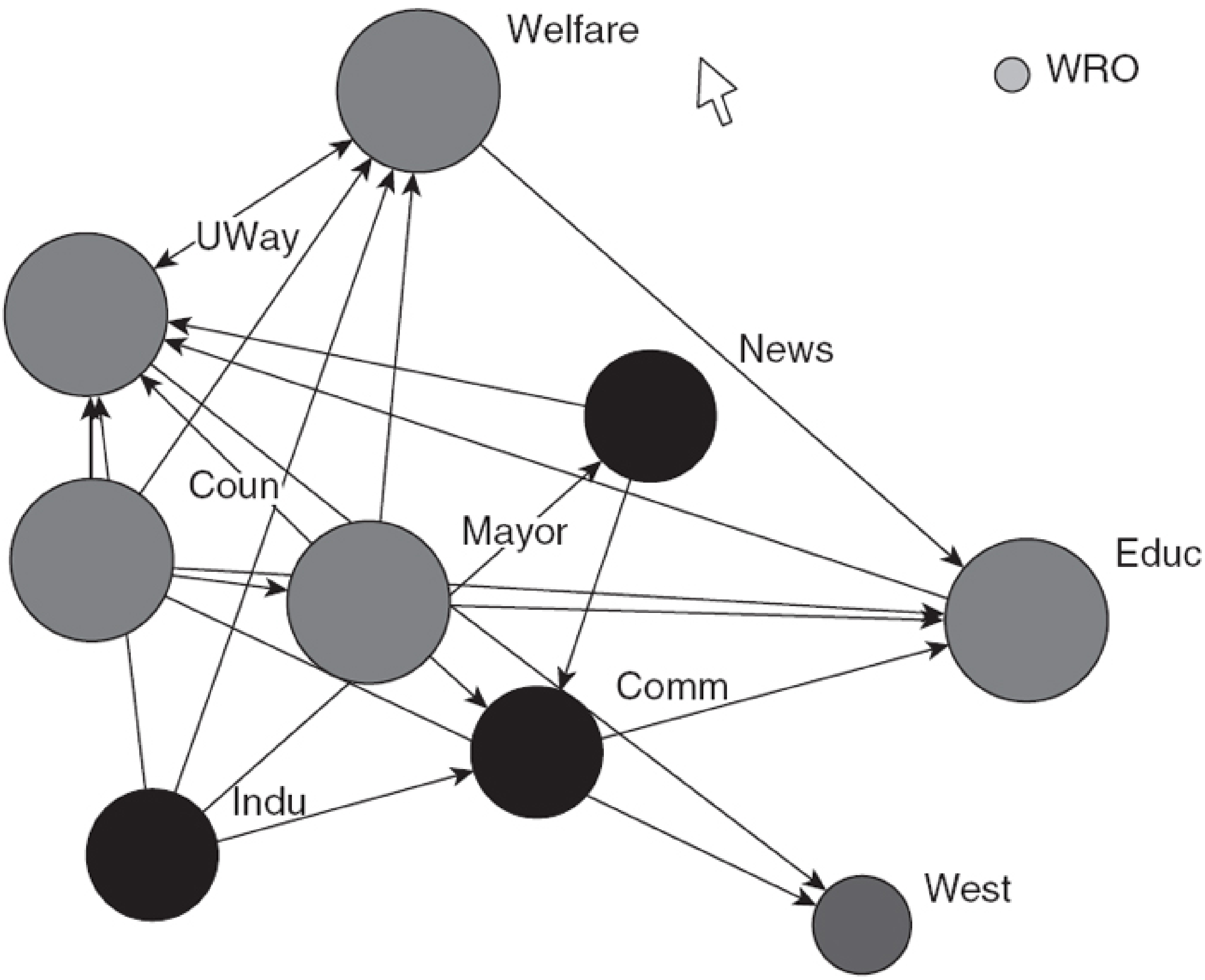
A visual inspection of the Figure 23.2 with the two attributes highlighted by node color and shape is much more informative about the hypotheses of differential rates of connection among black and lightly shaded (red and blue, respectively, in our online version) and among circles and squares. It doesn't look like this diagram is very supportive of either of our hypotheses.

Figure 23.2 Knoke information network with government/nongovernment (solid) and generalist/specialist (shaded) indicated



Nodes may differ quantitatively as well as qualitatively. In a graph of trade flows in the world system of nationally bounded economies, for example, it might be useful to make each node's size proportional to its GDP. The quantities that distinguish nodes may also be based on measures that describe their relational positions in a network; for example, nodes might be shown with sizes proportional to the number of ties each actor has. Figure 23.3 combines the quantitative and qualitative, using color and size to indicate features describing how each node is embedded in the network.

Figure 23.3 Knoke information exchange network with K-cores



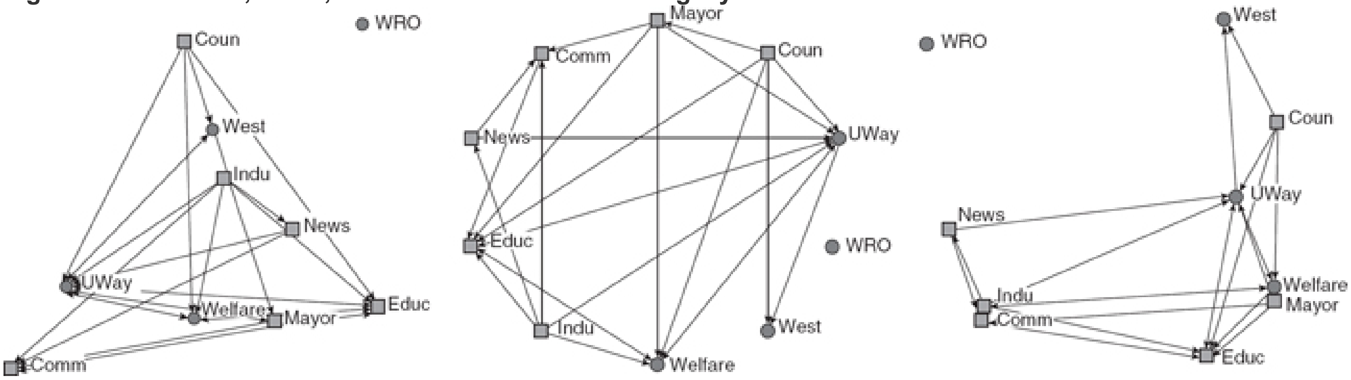
[Figure 23.3](#) shows four subgroups, which are shaded differently to identify which nodes are members of which "K-core" (a K-core is one approach to identifying coherent subgroups in a graph, discussed in the next chapter). In addition, the sizes of the nodes in each K-core are proportional to the sizes of the K-core. The largest group contains government members (Mayor, County Government, Board of Education), as well as the main public (Welfare) and private (United Way) welfare agencies. A second group, colored in solid black, groups together the newspaper, chamber of commerce, and industrial development agency. Substantively, this actually makes some sense!

The relations among the actors can also have "attributes." It can be very helpful to use color and size to indicate difference of kind and amount among the relations. Where the ties among actors have been measured as a value, the magnitude of the tie can be suggested by using thicker lines to represent stronger ties. Dashed lines, or lines of different colors, can be used to indicate different kinds of relations among actors, allowing multiplex data (data with more than one kind of relation) to be displayed in a single graph.

Node Location

Most graphs of networks are drawn in a two-dimensional "X-Y axis" space (Mage and some other packages allow three-dimensional rendering and rotation). Where a node or a relation is drawn in the space is essentially arbitrary — the full information about the network is contained in its list of nodes and relations. Figure 23.4 shows exactly the same network (Knoke's money flow network) that has been rendered in several different ways.

Figure 23.4 Random, circle, and multidimensional scaling layouts



The first drawing locates the nodes randomly, the second drawing uses a “circle,” and the third locates points according to its scores on a two-dimensional nonmetric scaling of the similarity of the node's tie profiles. It may be helpful to create graphs (based on the random graph) that group certain cases close together, or create clusters so that we can see differences in the patterns of ties within and between groups. Circle graphs highlight which nodes are highly connected and which are less so. Drawings that indicate scaling or clustering of network data, like the third one, may be particularly illustrative. In this drawing, the closeness of points indicates similarity (by some definition, and there are many alternative definitions) of nodes. The X and Y axes may also suggest patterns to be explored. Note that the public nodes tend to be grouped to the lower right, and that the private organizations are not very clustered.

Ego Networks (Neighborhoods)

A very useful way of understanding complicated network graphs is to see how they arise from the local connections of individual actors. The network formed by selecting a node, including all actors that are connected to that node, and all the connections among those other actors is called the “ego network” or (one-step) neighborhood of an actor (“neighborhoods” can also be found for two or more degrees of distance from ego).

To visualize the way in which individual nodes are embedded in the whole network, drawings of their ego networks are often very helpful. One of the basic insights of the social network perspective is that actors’ attributes and behaviors are shaped by those with whom they have direct relations, and actors may act to re-shape these constraints. Graphing individuals’ local networks and comparing them (e.g., are the ego neighborhoods of all government actors bigger than nongovernment actors? Which actors have neighborhoods that are composed mostly of actors of the same type?) can give great insight into similarities and differences among actors.

A lot of the work that we do with social networks is primarily descriptive or exploratory, rather than confirmatory hypothesis testing. For small networks, visual inspection of graphs can give a feel of the overall “texture” of the social structure and can suggest how individuals are “embedded” in the larger structure. All of the numerical indices that we will discuss in the next chapter are really just efforts to attach numbers to features that we naturally “see” in graphs. Working with drawings can be a lot of fun, and it is a bit of an outlet for one's creative side. A well-constructed graphic can also be far more effective for sharing your insights than any number of words. Large networks can be difficult to study visually, however. Formal description of the properties of networks and testing hypotheses about them require that we convert our graphs to numbers.

Using Matrices to Represent Social Relations

Graphs are very useful ways of presenting information about social networks. But when there are many actors or many kinds of relations, they can become so visually complicated that it is very difficult to see patterns. It is also possible to represent information about social networks in the form of matrices. We'll briefly review some of the most commonly used matrix representations of social network data. The language of matrices and matrix operations is important if you are going to work with network data. You don't have to do the math (that's why we have computers), but the mathematical concepts provide an efficient way to think about data handling and analysis.

A *matrix* is nothing more than an array (or list) of data (usually named with a bold letter). We might call a list of the 10 organizations names in the Knoke bureaucracy study **A**. Each *element* (organization name) can be *indexed* by its place in the list (name 1, name 2, etc.). There are a several types of matrices that are used, often in combination, in social network analysis.

Vectors

Matrices can have a single dimension (e.g., a list of names). Matrices with one dimension are called *vectors*; *row vectors* are "horizontal" lists of elements, and *column vectors* are "vertical" lists of elements. Row or column vectors are most commonly used in social network analysis to present information about the attributes of nodes. Consider the list of the Knoke organizations and a dummy coding of whether they are governmental (1) or not (0) in Figure 23.5.

Figure 23.5 Attribute vector for Knoke bureaucracies data

```
DISPLAY
------------------------

             1
             G
             -
  1 COUN     1
  2 COMM     0
  3 EDUC     1
  4 INDU     0
  5 MAYR     1
  6  WRO     0
  7 NEWS     0
  8 UWAY     0
  9 WELF     1
 10 WEST     0

------------------------
```

Figure 23.5 actually has two column vectors: organization ID and coding of government or not (the row and column labels 1–10 and 1G were added by UCINET). This makes the data array a ten-by-two (number of rows by number of columns) rectangular matrix. Rectangular matrices are simply "lists of lists" (either horizontally or vertically).

Most network data sets include arrays of variables that describe the attributes of the nodes. These attributes can be purely qualitative (e.g., the name of the organization), or nominal, ordinal, or interval (e.g., number of employees). Each variable can be stored in a single vector, or the vectors can be gathered into a rectangular array like in the example in Figure 23.5. These attribute data sets look very much like conventional social science data: rows representing cases, each coded with columns representing variables.

Arrays like Figure 23.5 are used to provide information about each node. We may code this information from our independent observations, as in the example. Each node might also be described by a variable that is based on its relational properties, such as its number of ties to other nodes, or its "betweenness centrality." A very common use of attribute vectors in social network analysis is to indicate the "group" to which a case belongs. There is a special name for these kinds of codes: partitions. Partitions (like the codes of zero or one in Figure 23.5) can be used to select subsets of cases, rearrange the data, and compute summary measures (e.g., what is the relative density of ties among government organizations relative to their average densities of ties to nongovernmental organizations?).

Square Matrices

In formal mathematics, a network is defined as a collection of nodes and relations. Generally, vectors or rectangular collections of vectors are used to describe the attributes of the nodes. Square matrices of two dimensions, in which both rows and columns contain lists of the same nodes, are used to describe the relations or connections between each pair of actors. Consider the example in Figure 23.6.

Figure 23.6 Adjacency matrices of information and money ties among Knoke's bureaucracies

Matrix #1: KNOKI

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	0	1	0	0	1	0	1	0	1	0
2	1	0	1	1	1	0	1	1	1	0
3	0	1	0	1	1	1	1	0	0	1
4	1	1	0	0	1	0	1	0	0	0
5	1	1	1	1	0	0	1	1	1	1
6	0	0	1	0	0	0	1	0	1	0
7	0	1	0	1	1	0	0	0	0	0
8	1	1	0	1	1	0	1	0	1	0
9	0	1	0	0	1	0	1	0	0	0
10	1	1	1	0	1	0	1	0	0	0

Matrix #2: KNOKM

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	0	0	1	0	1	0	0	1	1	1
2	0	0	1	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	1	0	0
4	0	1	1	0	0	0	1	1	1	0
5	0	1	1	0	0	0	0	1	1	0
6	0	0	0	0	0	0	0	0	0	0

The simplest and most common matrix to show relations is binary. That is, if a tie is present, a one is entered in a cell; if there is no tie, a zero is entered. This kind of a matrix is the starting point for almost all network analysis and is called an “adjacency matrix” because it represents who is “next to” (adjacent) to whom in the social space mapped by the relations that we have measured. Figure 23.6 shows two such adjacency matrices.

A matrix may be symmetric or asymmetric. By convention, in a directed (i.e., asymmetric) matrix, the sender of a tie is the row and the target of the tie is the column. Let's look at a simple example. In Figure 23.6, we see that organization 1 sends information to organization 2 (i.e., the entry in cell 1, 2 = 1), and this relation is reciprocated (i.e., the entry in cell 2, 1 =1). Organization 1 also sends information to organization 7, but this tie is not reciprocated. Asymmetric ties are not necessarily reciprocated, though they may be. A symmetric matrix represents “bonded ties” or “co-membership” or any kind of social relation in which, if A is tied to B, B must logically be tied to A (which is the case in most institutionalized or role relations).

Each node-by-node square matrix represents a map of a particular relation between all pairs of actors in the network. If there are multiple relations, as in Figure 23.6, these are represented as a third dimension (or slice or stack). The quantity in each cell of the matrix represents the strength of the tie between two actors (in symmetric data) or the quantity directed from the node on the row to the node on the column. The strength of relations can be measured as a nominal dichotomy or adjacency (as in our example), or at higher levels. A more in-depth discussion of data types is provided by Peter Marsden (this volume).

In representing social network data as matrices, the question always arises: what do I do with the elements of the matrix where $i = j$? That is, for example, does organization 1 “send information” to itself? This part of the matrix is called the *main diagonal.* Sometimes the value of the main diagonal is meaningless, and it is ignored (and left blank or filled with zeros or ones). Sometimes, however, the main diagonal can be very important and can take on meaningful values.

Multiplex Matrices

Many social network analysis measures focus on structures defined by patterns in a single kind of relationship among actors: friendship, kinship, economic exchange, warfare, and so on. Social relations among actors, however, are usually more complex, in that actors are connected in multiple ways simultaneously. In face-to-face groups of persons, the actors may have emotional connections, exchange relations, kinship ties, and other connections all at the same time. Organizations exchange personnel, money, and information and can form groups and alliances. Relations among nation-states are characterized by numerous forms of cultural, economic, and political exchange.

Multiplex data consist of a series of matrices (or “slices”), each of which is a square matrix describing a single type of tie among all pairs of actors. The various relations may be symmetric or asymmetric and may be scored at different levels of measurement. The two matrices shown in Figure 23.6 are two slices of the Knoke bureaucracies data, showing information sending/receiving and money sending/receiving relations among the 10 organizations.

One may apply all the tools of network analysis to each matrix in multiplex data, separately. For example, is there greater network centralization in the flow of money than there is in the flow of information among organizations? But we may also wish to combine the information on multiple relations among the same actors. There are two general approaches: reduction and combination. The reduction approach seeks to combine information about multiple relations among the same set of actors into a single relation that indexes the *quantity* of ties. The combination approach also seeks to create a single index of the multiplex relations, but attempts to represent the *quality* of ties, resulting in a qualitative typology. Role algebras (see Pattison, this volume) are a particularly important approach to qualitative reduction of multiplex data.

A special type of multiplex data arises when we obtain multiple reports or views about the same social structure. This type of cognitive social structure data maps the ties among pairs of actors in each slice and has a slice for each perceiver. One may wish to use matrix operations to combine the multiple cognitive maps (e.g., averaging, minimum value, maximum value, etc.), or one may wish to find groups of perceivers who have more or less similar views of the social structure.

Affiliation (Two-Mode) Matrices

A central focus of sociological analysis is the embedding of individuals in larger structures (e.g., families, organizations, communities, networks, identity categories). These larger structures are often seen as arising from the agency of individuals. Social network analysis may be used to map and study the relations within and between multiple levels of analysis: individuals affiliate with groups and organizations; organizations are linked in community ecology by their overlapping memberships.

Any array of relational data that maps the connections between two different sets of actors is said to be "two-mode." In sociological analysis, it is common for one mode to be individual actors and for the other mode to be sets of events, organizations, or identity categories. These types of data are called affiliation networks because they map the membership of connections, or affiliations, of actors with structures. Borgatti and Halgin (this volume) provide an in-depth treatment of the analysis of two-mode data.

It is not uncommon to translate two-mode data into a series of one-mode matrices. For example, one might transform a matrix that displays which persons were members of which voluntary organizations in a community into a matrix showing how many times each pair of persons happened to be co-members. One can transform the same data into a one-mode matrix that shows how many times each pair of organizations was connected by overlapping memberships. Asymmetric one-mode data can also be viewed as two-mode data in which the same actors are the lists for the two modes.

Image Matrices and Hyper-Graphs

Relational matrices that show which actors are connected to which others are of great descriptive and practical use. Often, though, our interest is in more abstract categories and relations among them. For example, it is interesting to note that the United States imports many relatively low-tech manufactured goods from China, and it sells many high-tech and brand-name goods in return. To a world-systems analyst, however, this relation is simply one example of a larger class of equivalent relationships involving core and semiperipheral nations. Ferligoj et al. (this volume) provide an introduction to methods of identifying and working with "equivalence classes."

Once classes and their member nodes have been identified, we can represent ties among them with graphs. These hyper-graphs have classes as nodes, and edges or arcs as defined by the equivalence relation. Such graphs can greatly reduce the difficulty in visualizing networks with large numbers of nodes and can provide great analytic insight.

The information in a hyper-graph can also be represented in matrix form. The original rows and columns of the actor-by-actor relation matrix are re-arranged (permuted) to group the actors in the same class together. This re-arranged matrix is made up of "blocks" that map relations of members of a class to one another (in the diagonal blocks) and all the relations of the actors in one class to all those in another class in off-diagonal blocks.

The blocked and permuted matrix is then often reduced to a new class-by-class matrix by summarizing the information within each block. Sometimes, the average density of ties, or the average value of tie strength, is used to summarize the blocked matrix. Frequently, some cut-off value is chosen (often the average value of ties for the entire network) and blocks are assigned a value of 1 if ties exceed the cut-off or 0 if they fall below the cutoff. This type of zero-one matrix, with groups or equivalence classes as nodes, is called the “image” matrix. Thoughtfully constructed image matrices can greatly simplify complex patterns among large numbers of actors (and, of course, badly constructed matrices can obscure them).

Dynamic Networks

Increasing attention is being given to the dynamics of networks, with particular attention to understanding how the embedding of actors in a particular place in a network at one time may effect change in their attributes or behavior and how the attributes and behaviors of actors at one point in time may shape the pattern of ties that are built and dissolved over time. Snijders (this volume) discusses approaches to studying network dynamics; many of the statistical models for the analysis of network data discussed in the chapters by van Duijn and Huisman and by Robins (this volume) are specifically aimed at the analysis of change.

A dynamic network can be represented as a series of matrix cross-sections, and it is often the case that dynamic data are observed this way. This approach results in a multiplex matrix in which one dimension (usually the slice) is defined by “time.” Alternatively, we may have information on the exact times at which events began and ceased (e.g., when actors joined the network and departed, when ties were formed or dissolved). Data of these types are becoming increasingly common with the use of digital instruments for network data collection (e.g., computer server logs, video recordings). Generally, such dynamic data are stored as lists of events (the network analysis package Pajek has a number of data formats and algorithms specifically designed for dynamic data), and programming is used to build matrices for analysis from them.

Conclusion

Once a pattern of social relations or ties among a set of actors has been represented in a formal way (via a graph or matrix), we can define some important ideas about social structure in quite precise ways using mathematics for the definitions. In the next chapter we will examine some of the most commonly used approaches to describing the “texture” of whole networks and the positions of individual nodes in them.

References

Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002) UCINET for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.
Hanneman, R. and Riddle, M. (2005) An Introduction to Social Network Methods. http://faculty.ucr.edu/~hanneman/nettext.
Knoke, D. and Kuklinski, J.H. (1982) Network Analysis. Beverly Hills, CA: Sage.
Knoke, D. and Wood, J. (1981) Organized for Action: Commitment in Voluntary Associations. New Brunswick, NJ: Rutgers University Press.
Scott, J. (2000) Social Network Analysis: A Handbook.
2nd ed.

London: Sage.
Wasserman, S. and Faust, K. (1994) Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511815478

- nodes
- matrices
- actors
- network analysis
- networking
- social networks
- social network analysis

http://dx.doi.org/10.4135/9781446294413.n23