Looking Back, Looking Ahead

In: Social Network Analysis and Education: Theory, Methods & Applications

By: Brian V. Carolan
Pub. Date: 2013
Access Date: July 15, 2021
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412999472
Online ISBN: 9781452270104
DOI: https://dx.doi.org/10.4135/9781452270104
Print pages: 261-286

Looking Back, Looking Ahead

Objectives

This book has introduced the theories, methods, and applications that constitute the interdisciplinary field of social network analysis and its potential uses in educational research. The purpose of this chapter is to succinctly summarize these ideas, reiterate its limitations, and ultimately prod you to start thinking about and doing social network analysis in educational research. After reviewing the main themes across this book's chapters, this final chapter provides you with a concise series of steps that provide a blueprint to conduct your own social network analysis study. The chapter also provides a summary of the different software tools that are available to assist with these studies. These different applications are presented as either general packages that perform an array of routine and advanced analyses or specialized packages that are intended to conduct specific, more advanced analyses. Next, the chapter revisits some of the ethical challenges of conducting social network analysis in and around educational settings and offers practical guidance when proposing a social network analysis to your institutional review board. Finally, in this chapter, you will learn about some of the areas in which social network analysis is likely to lead to new insights that are of interest to educational researchers.

Chapter Summaries

This book has introduced an assortment of theories, procedures, techniques, methods, measures, and applications of social network analysis to topics related to education. In Part I, Chapter 1, you learned about the social network perspective and the importance it places on how relationships influence a person's behavior above and beyond the influence of one's individual characteristics. To help develop this perspective, this chapter also introduced you to several foundational concepts, including actor, ties, groups, relation, and social network. These concepts provided the vocabulary that was used throughout the text. Once these concepts were defined, this chapter argued that social network analysis is a powerful tool for empirically studying the dynamic and fluid view of schooling that is at the core of contemporary educational theory. After introducing a few exciting areas in which social network analysis has generated insights, this chapter closed by delineating between the different analytical levels at which social network analysis operates.

Chapter 2 provided a concise summary of the interdisciplinary origins of social network analysis. From its early ideas and practices to its contemporary applications that emphasize the mathematical modeling and visualization of social networks, social network analysis has sought to integrate theory with method. This is the core idea developed in Chapter 2: Social network analysis goes beyond methodology to inform a new theoretical paradigm that emphasizes that the elemental unit of social life is the social relation. Oddly enough, social relations were often de-emphasized throughout much of the history of educational research. After speculating on reasons why this has been the case, Chapter 2 offers three specific ways in which social network analysis can further advance educational research: (1) bridge theory and data; (2) develop and employ advanced statistical techniques; and (3) refine and reconceptualize varied phenomena. This chapter also briefly discussed some of the ethical challenges that educational researchers encounter when conducting social network analysis.

The final chapter in Part I, Chapter 3, introduced basic network concepts and surveyed the means through which network data and their characteristics at either the ego or complete network level are represented. Furthermore, this chapter provided an overview of additional foundational ideas and concepts. For example, in this chapter, you learned about how network data can be represented as a graph or matrix and the ways in which these data differ from traditional actor-by-attribute data that dominates most social science. Additionally, you learned the difference between one-mode and two-mode matrices and multiplex matrices. This chapter ended with a discussion of the three types of variables that are often used in network studies: relational, structural, and attribute. As you become more proficient, it is likely that your own network-based research will ultimately incorporate all three kinds.

Part II of this book consisted of Chapters 4 through 7 and focused on the methods and measures associated with ego-level or complete network data. Chapter 4 focused on the collection and management of network data. Key ideas were introduced in this chapter, including boundary specification, sampling, measurement, collection, storage, and measurement. While these ideas are germane to any research endeavor, Chapter 4 addressed specific issues that you need to consider when collecting and working with network data. This chapter began with a discussion of the importance of specifying the network's boundary and the three strategies through which this can be done: positional, relational, or event-based. The chapter then went on to discuss the ways in which you can collect or generate relational data for complete and egocentric network analyses, including archives, sociometric instruments, and name generators/interpreters. This chapter also discussed important issues related to the quality of network data, including validity, reliability, measurement error, and patterns of missingness.

Chapters 5 through 7 introduced an array of different ways in which the static properties of egocentric or complete networks can be mathematically modeled. Using Newcomb's Fraternity Data, Chapter 5 specifically focused on structural measures that are calculated on the complete network. These measures include size, density, and reciprocity, among others. This chapter then went on to identify the ways in which groups (substructures) of actors can be identified in complete networks using relational data. These approaches were presented as either “bottom-up” (e.g., *K*-cores and *N*-cliques) or “top-down” (e.g., components and bi-components).

Chapter 6 continued addressing the analysis of complete network data but turned toward the issues of identifying positions and the relations among positions. Here, the key distinction between groups and positions is that a network position is a set of actors that occupy the same place or have similar patterns of relations with others. That is, positions consist of actors who share the same place in the network, but they need not be connected to each other (like actors in groups), though they very well might be. This is what distinguishes a position from a group. A number of different techniques to conduct a positional analysis were introduced, each based on a different definition of equivalence.

The final chapter in Part II, Chapter 7, shifted your attention to egocentric network measures. The point was made that ego-level network analysis can be done using ego data that are collected as part of a complete network study or (ideally) randomly selected from some target population. Regardless of the method through which the ego data are collected, most egocentric indices require information on an ego's alters and preferably information on each alter-alter pair (this is much easier when the ego data are collected as part of a complete network study). Once the egocentric data have been collected and organized, a number of different measures can be calculated to reveal different properties. These properties are related to connectivity (size and density), centrality (degree and betweenness centrality), structural holes (constraint), and brokerage (coordinator and gatekeeper). These different measures were used to describe the content and contours of an ego's local neighborhood.

Part III represented an effort to bring together much of what was covered in the book's previous chapters. The overarching goal of this section was to show you how the different conceptual and methodological tools you have learned can be applied to educational research. The first chapter in this section, Chapter 8, differentiated between the mathematical and statistical approaches to social network analysis. Whereas much of the material in Part II focused on calculating indices that reflected what a complete or ego network “looks like,” this brief chapter introduced you to the logic of statistical inference with network data. While most applications of social network analysis in education continue to be descriptive, educational research is poised and pressured to adopt some of the exciting inferential techniques that allow you to assess the likelihood of a certain network configuration, or whether a certain configuration is expected and “normal.” The use of these relatively new classes of inferential techniques, which are based on simulations, will allow you to make and test predictions using network data and move beyond techniques that describe static network properties. Using ERGMs as an example, the goal of this chapter was to not get mired in the technical aspects of this and other related techniques, but rather to sell you on the idea that the use of the inferential modeling techniques facilitates an important shift from description to explanation.

Chapter 9 extended this discussion by demonstrating a number of different techniques that make use of simulations when statistically modeling network data. These techniques were introduced according to the types of questions that are asked and whether the focus is on (1) ties between actors in complete networks; (2) individual attributes; or (3) relations within and between groups. Results of these different analytical techniques, including QAP, p_1, and p^* on data from Daly's School Leaders data set, were presented.

Chapter 10 reflected an effort to connect the methods, measures, and techniques presented in the previous five chapters to a rich theoretical area—social capital—to which educational researchers have made important contributions. This chapter reviewed the different ways in which social capital is defined and measured and the implications these differences have on the study of educational outcomes, particularly students' achievement. Using a number of empirical examples drawn from the educational research literature, the goal of this chapter was to help you understand what social capital is (and isn't) and how the conceptual and methodological tools associated with social network analysis are essential to testing and refining this rich theory. This chapter also revisited some ego-level indices (e.g., constraint and density), first introduced in Chapter 7, that are often employed in studies that operate from a social capital framework.

Similar to the previous chapter, Chapter 11 revisited the methods, measures, and techniques presented in earlier chapters but connects these to a second rich theoretical area—diffusion—that is very relevant to a range of educational processes. Chapter 11 outlined the core elements of diffusion theory and the major models used to understand how diffusion through networks occurs. After introducing four network-based diffusion models (integrated/opinion leaders, structural, critical values, and dynamic), two empirical examples were used to highlight the ways in which network data were collected, measured, and modeled to explain diffusion processes. Caution regarding the data requirements for such diffusion studies was urged and the potential of ABMs was introduced as a way to alleviate some of these concerns in order to study diffusion processes.

As this summary suggests, social network analysis is a broad field that is rapidly expanding, as theoretical insights and methodological techniques and tools from diverse academic disciplines seem to percolate almost daily. Coupled with this is the fact that the last several decades have witnessed an explosion of awareness about networks, not only within these various academic disciplines but within the larger cultural consciousness as well. It is now common to regularly speak of one's social network because of the ubiquitous use of networking platforms (e.g., Facebook) as well as the range of social media affixed to mobile phones (McFarland, Diehl, & Rawlings, 2011). This awareness and popularity are understandable, but they do not translate into making things any easier for you as a researcher. Like any research endeavor, selecting the appropriate methodology for your research questions is a big, important challenge. Social network analysis, in fact, may compound this problem by providing a theoretical perspective and methodology that is, at least on the surface (1) intuitively appealing; (2) appropriate for many empirical settings and questions; and (3) seemingly easy to do. After all, if interested in peer relations in a diverse elementary school classroom, you simply need to ask students in the class to name their closest friends. With this information, you then have access to a bunch of standard measures that help unpack that classroom's hierarchy, groups, positions, and so forth. If you also collect information on students' attributes like gender, for example, you can examine whether there is a tendency for ties to be reciprocated among those who share the same gender. While on the one hand, this seems straightforward, on the other hand, such a study can lead to a dizzying amount of complexity.

Looking Ahead

So where do you begin? Integrating much of the material covered throughout this text, I have adapted Prell's (2012) suggested process, which walks you through the nine steps—from theory to design—that are necessary to conduct your own social network study. These steps confront basic issues such as: How do I design a social network analysis? How do I gather data? In what ways can I address some of the ethical, validity, and reliability issues of social network analysis? Once I gather my data, then what?

These are critical questions to ask and relevant for just about any research that you may conduct. However, within the context of social network analysis, concerns surrounding these issues are handled slightly differently. This section revisits some of the tools covered throughout this text that will help you address these issues. While the steps outlined below can be followed in any order, especially depending on “where” you enter the research process, I recommend that they be followed in as a linear manner as possible, regardless of your level of experience or familiarity. Table 12.1 lists these steps in this recommended order.

Table 12.1 Nine-Step Process for Conducting a Social Network Analysis. Adapted from Prell (2012).

1. Consult the literature
Your search for relevant empirical literature on your topic of interest should also include journals and databases that are not solely focused on education.

2. Develop a theoretical frame
What is the logical explanation that proposes a causal process or mechanism that produces an outcome of interest? What this means is that you as a researcher should have a fairly clear idea of the causal explanation you would like to test through the collection and analysis of social network data. This frame will be informed by your review of the literature.

3. Draft research questions or hypotheses
When constructing research questions or hypotheses pertaining to the study of social networks, the key in this process is to think about how a network variable (either relational or structural; Chapter 3), relates to, affects, or is affected by another set of variables.

4. Select a sample
Whom or what do you plan to study? In making this decision, you will have to consider issues of boundary specification, sampling, and populations (Chapter 4). There is another important issue to consider, one that will affect the five remaining steps: Will you examine ego networks or complete networks?

5. Collect data
With a clear question of hypothesis, you can develop a better image of the type of data you are interested in gathering. These data should reflect your variables of interest and include relational data on how actors are connected to one another.

6. Data considerations
Will you collect directed or undirected relational data? Will these network data be valued or binary? How valid and reliable are the network data? In what ways will you address the issue of missing data?

7. Input and manage data.

There are several options for inputting and structuring your data (reviewed in the second half of Chapter 3). The most common method is your basic network data matrix, also known as an adjacency matrix. Remember that each relation has its own matrix, and attribute data are stored in a separate rectangular matrix with the rows in the same order as the adjacency matrix.

8. Visualize network

So long as the network isn't too large and dense, visualizing a network will give you an intuitive feel for the network's topography.

9. Further descriptive or inferential analyses

Your analysis depends on your questions. But chances are these questions will require you to perform tests of statistical significance to examine the likelihood of (1) ties between actors in complete networks; (2) certain individual attributes predicted from relational data; or (3) relations within and between groups.

Source: Prell, C. (2012). *Social network analysis: History, theory, and methodology*. Thousand Oaks, CA: Sage Publications Inc.

Step 1: Consult the Literature

This first step sounds silly, but I am often taken aback by how often novice and senior researchers ignore—or give little attention to—this first step. You will first need to become familiar with the social network studies conducted on a topical interest similar to your own. This should happen even before you start formulating your own research questions. Of course, all the standard tricks for searching the empirical literature apply, such as using “social networks” or “social network analysis” as your key words. While you should expand your search to include journals that specialize in social network analysis (e.g., *Social Networks* and *Social Network Analysis and Mining*), you should also consider searching journals that are more closely aligned with a specific discipline (e.g., *American Journal of Sociology* and *American Sociological Review*) or are more broad in scope in that they may address issues related to children and schools (*Journal of Research on Adolescence* and *Youth and Society*, e.g.). The point I am making is that your search for relevant empirical literature should also include journals and databases that are not solely focused on education. You will be somewhat surprised how expanding your search in these ways will yield a bigger, better, and deeper pool of literature that can be used to inform your research questions and design. Of course, this first step will help you refine your ideas before you move forward, but you will also find useful concrete examples of how previous researchers have formulated their research questions, collected data, and analyzed those data.

Step 2: Develop a Theoretical Frame

Many of the early examples of social network analysis presented in Chapter 2 were inductive; that is, there was an interest in inducing theoretical concepts and frameworks based on the network data they had collected and analyzed through early mathematical and algebraic models. More commonly, however, social network analysis studies adopt a deductive, theory-driven approach. Remler and Van Ryzin (2010) define theory as a logical explanation that proposes a causal process or mechanism that produces an outcome of interest. What this means is that you as a researcher should have a fairly clear idea of the causal explanation you would like to test through the collection and analysis of social network data. Part III introduced two of these theoretical areas (social capital and diffusion of innovation), but there are many others. These others include network exchange theory (Blau, 1964; Cook, Emerson, Gilmore, & Yamagishi, 1983), which focuses on how a network's structure influences whom in the network emerges as powerful, and social influence theory (e.g., Friedkin, 1998), which considers how actors influence one another's thoughts. Before considering how you might test propositions derived from these theories, it is critical that you immerse yourself in the relevant literature that has developed and tested various components related to these and other network-based theories.

Step 3: Draft Research Questions or Hypotheses

Research questions or hypotheses flow directly from your preferred theoretical framework—assuming, of course, you are operating in a deductive manner. First, start with an initial topic of interest and then turn this topic into a question. The third and final step is to identify the problem your selected question helps resolve—its significance (Booth, Colomb, & Williams, 2008). When constructing research questions pertaining to the study of social networks, the key in this process is to think about how a network variable (either relational or structural, Chapter 3), relates to, affects, or is affected by another set of variables. It is this step in the process that will align your work with the social network perspective and important subsequent data considerations. Putting these pieces together, you have the following framework for focusing your research question and clarifying its significance:

> General topic: Adolescent peer influences
>
> Question: How do the characteristics of an adolescent's closest friends shape that student's attachment to school?
>
> Problem (significance): Schools can more effectively and explicitly design learning environments that mitigate the negative effects of peers while also leveraging their positive effects.

When asked in this manner, the implied causal mechanism is the adolescent's friendship network, requiring your data to include a variable that measures the size, structure, and/or content of that local network. This question implicates social networks as affecting variables (e.g. attachment to school), yet this order could easily also be reversed. That is, you could ask how an adolescent's level of school attachment influences the likelihood of forming friendship ties.

Regardless of whether your questions treat the relational or structural variables derived from network data as predictor, outcomes, or both, they require information about an ego and his or her alters. Questions can also be framed in terms of hypotheses, a predictive statement about how you envision the relationship between variables (positive/negative or weak/strong, for example). Using the question posed above, we could test the following hypothesis: If an adolescent's closest friends' parents are involved in school, then that adolescent will demonstrate higher levels of school attachment. This hypothesis has further specified the characteristics on which you will be focusing and makes a prediction about the relationship. Good hypotheses also have several other characteristics—they are declarative, firmly grounded in theory, brief, and testable (falsifiable). Again, consulting the relevant empirical literature (Step 1) will provide some useful clues as to how to best craft these questions and hypotheses pertaining to the study of social networks.

Step 4: Select a Sample

So far in the design process, you have done some background reading and developed questions/hypotheses that examine an issue derived from a larger theoretical perspective. In this next step, you will determine whom you plan to study. In making this decision, you will have to consider issues of boundary specification, sampling, and populations (Chapter 4). There is another important issue to consider, one that will affect the five remaining steps—will you examine ego networks or complete networks? With ego networks, issues related to populations and samples are fairly straightforward. As with standard survey methods, you randomly sample a certain number of egos from a given population. Once you have a sample of egos, you gather your network data for each of your responding egos. Because these network data are on an ego's immediate, local neighborhood, they are likely to include ego-alter information and perhaps even information on alter-alter pairs. Thus, you will be able to calculate a few of the indices presented in Chapter 7, including size, density, and distance. If, however, you perform an ego-level analysis using data derived from complete network study, you have more options to calculate ego-level indices.

Issues related to sampling and populations for complete network data, on the other hand, are negotiated differently. Recall that a network's boundary refers to the set of actors that you consider to be a complete set of actors for the network study. Where do you set the limits when collecting complete network data when, in theory, there are no limits (Barnes, 1979; Knoke & Yang, 2008)? There are three generic approaches to addressing this issue: positional, relational, and event-based. The first of these approaches, positional, demarcates a network's boundary by including those actors bound together by some common attribute (e.g., teachers within school district). The relational approach is based on your knowledge about relations among a set of actors or relies on the actors themselves to nominate additional actors for inclusion. For example, you may specify the boundary of a network of parents within a school district by relying on a small number of influential parents—key informants—to identify those for inclusion in a study on the influence of parent networks on local educational policy decisions. This relational strategy relies on what is referred to as a reputational method, but relational strategies also include fixed lists, expanding selection approaches, and snowball samples. The third approach you may employ to address the boundary specification problem is event based. In this approach, your network would include only those actors who participated in an "event" at a specific time and place. For example, in a study on the relationship between parent networks and local educational policy decisions, your network could choose to include only those parents who physically attended a local board of education meeting within the past 12 months.

Step 5: Collect Data

Now that your network's boundary has been specified, you are ready to collect your data. These data should reflect your variables of interest and include relational data on how actors are connected to one another. Here, your research questions/hypotheses are of critical importance. Are you interested in how collaboration ties affect some behavior; for example, the likelihood of whether a teacher adopts a more student-centered instructional approach? Or are you interested in seeing how actors acquire various types of resources through different kinds of ties to others? For example, are students with more diverse ties likely to get more assistance outside of class from their peers? With a clear question, you can develop a better image of the type of data you are interested in gathering.

As this imagery comes into sharper focus, you will increase your likelihood of creating measurement instruments and a methodological approach that will result in the collection of high-quality network data that address your specific research questions or hypotheses. Chapter 4 describes these different data sources for either egocentric or complete network studies. These sources include census, archives, sociometric instruments, name generators, position generators, and resource generators, with these latter three being most relevant for egocentric studies. Network data can also be gathered through observations, interviews, or extracted from the back end of electronic databases. Qualitative work could also serve as a useful complement and helpful precedent to designing a quantitative social network study (see Hollstein, 2011, for a review of these strategies). Special attention should be given to issues of validity, reliability, and accuracy, which are also discussed in Chapter 4.

Step 6: Data Considerations

This next step is integrated with issues surrounding the collection of network data. In addition to issues of validity, reliability, accuracy, and patterns of missingness, there are two other important issues to consider. The first of these is whether you will collect directed or undirected relational data. If you are observing faculty members in a teachers’ lounge, are you interested in who speaks with whom, or are you also interested in who initiated the conversation? Who initiated the conversation and who responded might be more interesting than just recording the pairs of teachers who engaged in conversations. This scenario reflects the difference between directed (arcs) and undirected relational data (edges) data—a point that was emphasized in Chapter 3. Therefore, directed network data includes who sought professional advice from a colleague or who trusts whom. Directed data such as these are typically preferable because more information is held within directed data, as you not only have information on the tie being present/absent but also on the direction of that tie. Directed data also have the benefit of being able to be transformed into to undirected data. For example, if two actors nominate each other as friends (a reciprocal relationship), this can be transformed into an undirected tie in which the actors are coded as friends (the top and bottom halves of the sociomatrix would therefore look the same). However, the reverse is not true; undirected data cannot be used to retroactively attempt to discern the senders and receivers of ties.

A second related issue to consider is whether your network data will be valued or binary. The distinction, also elaborated in Chapter 3, has important implications for how you collect and record your network data. With binary data, your concern is recording the presence or absence of ties (represented as 1/0). For example, you may ask and record whether a teacher discusses nonschool issues with another teacher (of course, this can also be directed or undirected). For many analyses, binary data are sufficient for unmasking many network properties; in fact, many network measures require that the data be in binary form. However, it is often preferable and desirable to get a deeper look into the relation by measuring the strength of the tie. This requires that you collected valued network data. Valued data reflect the relative strength, frequency, or duration of a relationship between a pair of actors. Different options for gathering valued network data include the use of Likert-type scales that assess the frequency with which one engages in a behavior with someone else: 1, 2, 3, or 4; 1 = never, 4 = frequently. This is similar to the ways which relations were measured in Daly's School Leaders data set. Another option for gathering valued network data is to rank actors. This is the method employed in the Newcomb Fraternity data set in which fraternity members were asked to rank all other members in terms of friendship preferences. Another slight twist is that in both of these data sets, data on these relations were measured at more than one point in time, which allows you to best address questions that deal with network dynamics. One last point about valued data is that they can always be transformed into binary data; however, the reverse is not true.

Step 7: Input and Manage Data

Once your egocentric or complete network data are gathered, they are ready to be organized in a manner that is suitable for analysis. There are several options for inputting and structuring your data (reviewed in the second half of Chapter 3). I'll briefly focus on one method that—with a little processing—is readable by most general social network software packages (reviewed later). This method is your basic network data matrix. In this section, I will gently walk you through the creation of a network data matrix.

Table 12.2 Sample Sociometric Questionnaire. This is an example of a questionnaire that can be used to elicit sociometric data from a complete network of five teachers. Each row represents the network's actors (Teachers A, B, C, D, and E). The columns represent the three types of ties on which relational data are being collected. The values in the cells are selected by respondents, which generate valued and directed relational data on each pair of actors.

Teacher	*I consider this person my friend.*				*This person is someone with whom I enjoy collaborating.*				*I could turn to this person to get professional advice.*			
A	1	2	3	4	1	2	3	4	1	2	3	4
B	1	2	3	4	1	2	3	4	1	2	3	4
C	1	2	3	4	1	2	3	4	1	2	3	4
D	1	2	3	4	1	2	3	4	1	2	3	4
E	1	2	3	4	1	2	3	4	1	2	3	4

Please rate on a 4-point scale the extent to which you agree of disagree with each statement for each and every teacher listed (excluding yourself). Please circle a number for each teacher where 1 = strongly disagree; 4 = strongly agree.

Source: Prell, C. (2012). *Social network analysis: History, theory, and methodology*. Thousand Oaks, CA: Sage Publications Inc.

Table 12.2 presents a hypothetical example of a questionnaire that gathers relational and data for each teacher in a complete network. This questionnaire can also include items that elicit standard attribute data, including, for example, years of experience and whether the teacher has a master's degree (these are not shown on Table 12.2). The sociometric component of the questionnaire gathers data on three different relations: friendship, collaboration, and professional advice. It asks teachers to rate on a four-point scale the extent to which they agree or disagree with each statement for each and every teacher that is listed (excluding themselves) where 1 = strongly disagree; 5 = strongly agree. In addition to these data being directed, they are also valued, providing an indicator of strength of this tie.

The next step is to represent these data in the form of an adjacency matrix, the most common network data matrix in which two actors are considered adjacent if they are structurally “near” each other. When constructing an adjacency matrix, which can be done in any spreadsheet program, you first enter actors’ names in both the rows and columns (referred to as a single-mode matrix, Chapter 3). Table 12.3 shows what an adjacency matrix looks like using those same teachers listed on the sociometric questionnaire in Table 12.2. Rows represent senders while columns represent receivers of ties. An important point to reinforce is that each relationship has its own matrix. So, given that the questionnaire measured three types of ties, each would require its own matrix. In this example, there would be three separate matrices on the same set of five teachers: A, B, C, D, and E.

Table 12.3 Example Adjacency Matrix. This one-mode adjacency matrix consists of five actors (teachers). The rows of the matrix “send” ties, and the columns “receive” ties. These data can also be recorded in an edge-list format, which is appropriate for valued and directed network data such as these (described in Chapter 4).

Teacher	*A*	*B*	*C*	*D*	*E*
A					
B					
C					
D					
E					

Source: Prell, C. (2012). *Social network analysis: History, theory, and methodology*. Thousand Oaks, CA: Sage Publications Inc.

Working with only one of these ties (friendship), you next enter the nominations of each respondent, starting with the row representing that respondent and then inserting a value under the column representing that respondent's choice. For example, Table 12.4 shows that Teacher A nominated Teachers D and E as friends and “sent” each of them a 4. Teacher A did not consider Teachers B and C as friends; thus, a 1 is recorded in those cells. Of course, self-nominations in this instance make little sense, so a 0 is recorded in the appropriate cell. In most instances, in fact, the diagonal of the matrix is of little interest and is ignored or filled with zeros. The next step is to do this for each respondent in the network until you have a complete adjacency matrix in which all cells (those not on the diagonal) are filled with a 1, 2, 3, or 4, indicating the strength of the tie that has been sent by each row and received by each column.

Table 12.4 First Row of the Example Adjacency Matrix With Valued and Directed Data on Friendship Ties. This row shows the friendship ties “sent” by Teacher A to the four other teachers in the complete network. For example, Teacher A strongly disagrees with the statement that “I consider this person my friend” for Teacher B (strongly disagree = 1). Conversely, Teacher A strongly agrees with the statement that “I consider this person my friend” for Teacher D (strongly agree = 4). The 0 in this row indicates that self-nominations, in this instance, are illogical; that is, Teacher A cannot evaluate friendship with him/herself.

Teacher	*A*	*B*	*C*	*D*	*E*
A	0	1	1	4	4

Source: Prell, C. (2012). *Social network analysis: History, theory, and methodology*. Thousand Oaks, CA: Sage Publications Inc.

In addition, you can record your attribute data in a matrix that consists of three vectors (columns). The first vector is the respondent's ID, the second is a vector for the variable "years of experience," and the third is a vector for the binary variable indicating whether they have a master's degree. Table 12.5 shows how these attribute data are structured in matrix format. The complete data files for this hypothetical study, therefore, consist of four matrices: three adjacency matrices containing valued and directed relational data and one rectangular matrix with three vectors containing information on two attributes for each of the five teachers in the network.

A quick note about inputting and managing egocentric network data. The above example focused on the collection of complete network data. If collecting ego-level network data, each ego would have its own data matrix and attribute file. Then, local network properties can be calculated for each ego and exported into a standard statistical software package (with some savvy programming skills, you can even calculate some of these properties directly in standard statistical software) and merged with the attribute data file for each ego. In addition, as mentioned in Chapter 4, after the data file has been converted to a long format, the statistical analysis of egocentric data likely has to be done in a multilevel framework in order to account for the nonindependence of observations.

Table 12.5 Data Matrix with Three Vectors: Actor's ID, Years of Experience, and Master's Degree (1 = yes, 0 = no). These attribute data would need to be collected as part of the sociometric questionnaire, or gathered from existing data sources (school personnel records, for example). These data are stored in a matrix that is separate from the relational data. The important point for subsequent analyses is that the rows of this data matrix are in the same order as those of the adjacency matrices that contain the relational data. For example, Teacher A has 5 years of experience and a master's degree.

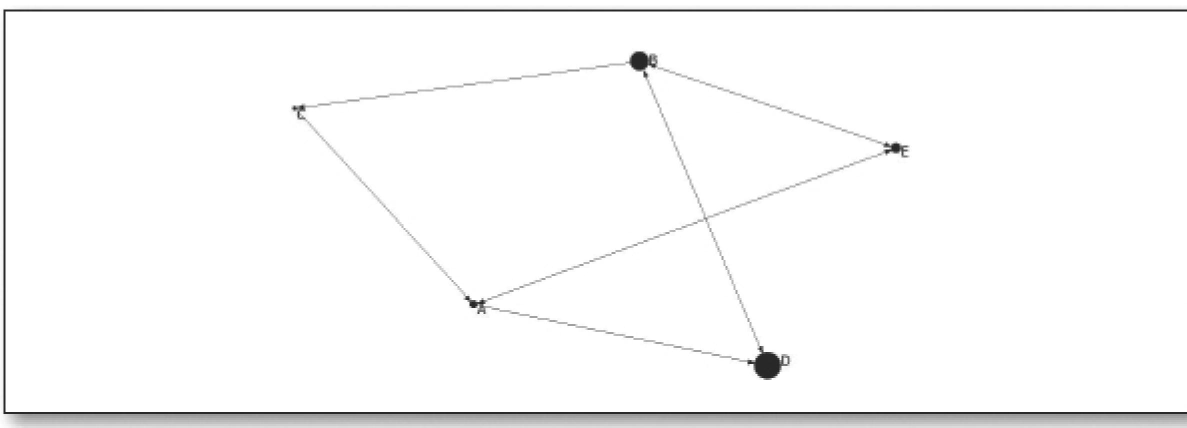
Teacher	*Years of Experience*	*Master's Degree*
A	5	1
B	14	1
C	2	0
D	21	1
E	7	0

Step 8: Visualize the Network

As you get started, however, I will keep things simple and focus on complete network data, specifically our complete network of five teachers. Before calculating any of the measures mentioned in Chapters 5 and 6, it is very helpful to visualize the network as a graph, which gives you an initial peek at what the network looks like. So long as the network isn't too large and dense, visualizing a network will give you an intuitive feel for the network's topography. Some guidelines for creating useful visualizations from network data are discussed later in this chapter.

For example, using our hypothetical network of five teachers and the valued and directed matrix that measures friendship ties, we can visually discern a few interesting features. Figure 12.1 represents this network as a graph. This graph only shows strong friendship ties between teachers (only those originally coded as a 4 in the adjacency matrix). First, the network consists of one weak component in which a number of the friendship ties are not reciprocated. For example, Teacher A nominated Teacher D, but not vice versa. Second, Teachers B and D have a reciprocal relationship, as they have both nominated each other. They are also the most experienced teachers in the network, suggesting that this mutual friendship may have something to do with their similar lengths of experience. Finally, this network is not very dense (density = 45%), suggesting that this is not a very cohesive network of teachers. Before calculating any descriptive network properties or performing any inferential analyses, visualizing your network in the form of a graph provides some hints as to whether your motivating research questions or hypotheses are worth pursuing or in need of revision.

Figure 12.1 Graph of Teachers' Friendship Network. Each directed line represents a friendship tie that has been sent from one teacher to another. A friendship tie is considered present only if the respondent selected a 4 on the sociometric questionnaire (Table 12.2). The size of each node reflects the attribute "number of years teaching," with bigger nodes equaling higher values on this attribute.



Step 9: Further Descriptive or Inferential Analyses

This last phase involves analyzing the networks in light of your research questions/hypotheses. The first phase of your analysis will focus on the descriptive properties of the networks. These properties reflect the network's structure, groups, and positions (Chapters 5 and 6). Depending on your questions, this type of descriptive analysis may be more than adequate. However, your research questions may be nudging in an inferential direction in which you'll want to perform tests of statistical significance to examine the likelihood of (1) ties between actors in complete networks; (2) certain individual attributes predicted from relational data; or (3) relations within and between groups. To do these types of analyses, you should have a firm handle on the material in Chapters 8 and 9.

A final word on getting started. While the above section outlines a series of steps to get you started in your own network study, an alternative may be to directly work with the data files that are available to you at the book's companion website, available at http://www.sagepub.com/carolan. These are some of the same files that have been used to perform the analyses cited throughout this text. These files can be downloaded and copied into UCINET or imported into your preferred software package (described below). These files, which include both relational and attribute data in matrix format, can be used to understand and explore network data. In addition, the companion website also provides instructions for how to get stated with UCINET, conduct a *p** analysis with PNet, and detect and visualize groups in NetMiner.

Software Sampler

Fortunately all the mathematics, statistics, and visualization associated with social network analysis—especially relevant in regard to Steps 7 through 9 above—are built into a number of general software applications that have been explicitly designed and marketed to social network analysts. In addition to performing these functions without needing to know any programming whatsoever, most applications come with a suite of powerful visualization tools that effectively communicate the static and even dynamic properties of social networks. This section reviews three of these general packages, which can be regarded as the major general packages with respect to functionality and provide both novice and advanced users with more than sufficient capability to perform both routine and advanced analyses. Then this section covers a small number of specialized programs that perform specific types of analyses. Finally, this section introduces two applications designed for network data collection. These programs are not presented in a manner that compares or ranks them, as this depends very much on your purposes and preferences. A more detailed treatment on these applications can be found in Huisman and van Duijn (2011).

General Packages

In this section, I review three of the most widely used general social network analysis software packages. By general, I mean that they posses ample capabilities for general exploration and analysis of network data. These three have been selected from approximately 15 others that are also considered general packages (e.g., InFlow, Network Workbench, Blue Spider). I focus on these three because they (1) have been available for many years; (2) have been extensively reviewed by others; and (3) are updated on a regular basis.

About UCINET

UCINET 6 (Borgatti, Everett, & Freeman, 2006) is the most popular, comprehensive package for the analysis of one-mode or two-mode social network data. In fact, most of the analyses presented throughout this book were performed with this application. This menu-driven Windows-based program has a number of strengths. First, it can handle fairly large networks. Second, through its easy-to-use menu, you can access a number of social network analysis methods, including routines for identifying subgroups, equivalence, and ego-level properties. Additionally, the program has excellent matrix routines that allow you to easily manipulate and transform data. Although its statistical techniques are limited, it does have strong permutation-based testing procedures, including QAP and MR-QAP. Also built into UCINET is NetDraw, a visualization tool that has advanced graphic properties. For these and other reasons, UCINET is an excellent choice, especially for someone new to network analysis.

NetMiner

A second excellent option for someone new to social network analysis is NetMiner (Cyram, 2009). In part, this is an excellent option due to its user friendliness, which is complemented by good user support and documentation. This application allows you to interactively explore network data in a way that integrates analysis with visualization methods. Data can be entered directly through a spreadsheet editor or by opening data sets in different formats, including NetMiner NTF files, Excel spreadsheets, or UCINET data sets. Able to handle large data sets, it contains a number of procedures for investigating a network's connectivity and local neighborhood structure, subgraph configurations, cohesion, and centrality. Like UCINET, it has advanced graphical features (and goes further by supporting 3D visualizations), but it also supports a larger number of statistical procedures. These procedures include correlation and regression analysis and even more complex p_1 and p^* models. These and other features make this package appropriate for both novice and advanced users. It is also available in a free trial version.

MultiNet

A third general and popular package for social network analysis is MultiNet (Richards & Seary, 2009). Designed for the analysis of large networks (like Pajek, another general package), MultiNet is a good choice when the interest is in exploratory network data analysis. Like UCINET and NetMiner, it supports the analysis of ego-level and complete network data. In addition to handling large networks, it can also handle a large number of variables that can be easily recoded, transformed, or manipulated in other ways. Foremost among its strengths is that it combines attribute and relational data into one model to perform a context analysis: It integrates data that describe people with data that describe relationships between people into a single analytic model. In addition, results are presented textually and visually, which will assist you in interpreting the results. It also recently integrated other programs such as PSPAR, which fits p^* models (Chapter 9), and NEGOPY, a program that finds cohesive subgroups (Chapter 6). One drawback for more advanced users is that its statistical procedures on network ties are limited. However, the package supports the user to an unusual extent, providing excellent online help and extensive documentation.

As you consider the package that best suits your needs, consult Huisman and van Duijn's (2011) excellent comparison of these three (and numerous others) different programs on nine criteria: (1) data manipulation; (2) network visualization; (3) network descriptives; (4) procedure-based methods; (5) statistical methods; (6) network dynamics; (7) availability of documentation; (8) online help; and (9) user friendliness. They use a + to denote that it is good, and ++ indicates that it has been scored as very good. In addition, a – indicates that the program has shortcomings in that area, a 0 indicating that the aspect under consideration is absent, and a +− indicating that the issue is undecided (having a mix of both positive and negative). Table 12.6 recaps Huisman and van Duijn's (2011) comparison of the three packages reviewed above. This can be used to help you select the one that is most appropriate for your analytical needs.

A quick note on network visualization. All three general packages also incorporate tools for visualizing networks. While the graphical representations of network data are easier to produce than ever before with the assistance of these embedded tools, the quick dissemination of these technologies has resulted in network images that do not adhere to any consensual set of basic principles (Krempel, 2011). However, as you explore your data through these visualization tools, it is useful to consider the following principles, which will result in better, more useful visualizations. First, the location of nodes and the distance between them should represent similarities and distances based on statistical procedures. Most visualization packages address this placement problem by employing various kinds of spring embedders (e.g., Kamada & Kawai, 1989), which ultimately provide information about local connections, that is, who is connected with whom and the strength of this connection. Closer nodes are typically placed close to one another. Second, shapes and colors should be used to communicate different classes or groups of nodes. Lines can also have different sizes and colors, which help delineate different types and strengths of relations. Altering the visual layers of these network attributes (attributes related to nodes and/or their relations) permits you to convey several pieces of numerical information simultaneously, providing a more complete multivariate view. Finally, extremely large networks with thousands of nodes should be filtered, which reduces the networks to the most connected nodes or most dominant lines. Keeping these principles in mind, the potential of network visualizations is that they are able to identify local combinations of external attributes that are connected to emerging social processes. In addition, network visualizations hint at where additional information is needed and direct our attention to areas of the network that need further theorizing and empirical investigation. Mapping networks is an important early step in the research process for all sorts of inquiries.

Table 12.6 Review of Three General Social Network Software Packages. Adapted from Huisman and van Duijn (2011). A + is used to denote that it is good, and ++ indicates that it has been scored as very good. In addition, a – indicates that the program has shortcomings in that area, a 0 that the aspect under consideration is absent, and a +− that is undecided (having a mix of both positive and negative).

	Functionality						*Support*		
	Data	*Visualization*	*Descriptives*	*Procedure-based Methods*	*Statistical Methods*	*Network Dynamics*	*Documentation*	*Online Help*	*User Friendliness*
UCINET	+ +	+ +	+ +	+ +	+	0	+ +	+	+
NetMiner	+ +	+ +	+ +	+ +	+ −	+ −	+	+	+ +
MultiNet	+ −	+	+ −	+	−	0	+	+ +	+

Source: Huisman, M., & van Duijn, M. A. J. (2011). A reader's guide to SNA software. In J. Scott & P. J. Carrington (Eds.), *The Sage handbook of social network analysis* (pp. 578-600). Thousand Oaks, CA: Sage Publications Inc.

Specialized Programs

Chapter 9 noted that the field of statistical methods is one in which major advances in social network analysis have occurred within the last 10 years. These advances have given way to a big increase in statistical routines and specialized packages needed carry them out. These specialized routines, for the most part, were developed to perform a small number of relatively advanced descriptive and statistical procedures, including *p** models (ERGMs), the analysis of longitudinal network data (actor-oriented models), and various measures of ego networks. Therefore, by *specialized*, I mean that they contain a few distinctive procedures for network analysis or a limited range of analytical procedures to perform a specific type of analysis.

PNet

PNet (Wang, Robins, & Pattison, 2008) is a specialized program that is used exclusively for the simulation and estimation of ERGMs, or *p**. It has three major functionalities: (1) simulating network distributions with specified model parameter values; (2) estimating specified ERGM parameters for a given network; and (3) testing the goodness of fit of a specified model to a given network with a particular set of parameters. This free, Windows-based software is a good first option to get started with your own ERGMs, as its supporting documentation is very helpful. It is available at http://www.sna.unimelb.edu.au/pnet/pnet.html.

RSiena

RSiena (Ripley & Snijders, 2010) is a specialized program that evolved from Siena, which was one of the first applications to analyze longitudinal network data and was originally implemented within the StOCNET package. Specifically, this program analyzes the co-evolution of networks and behaviors, that is, how behavior and network ties mutually influence each other. Siena methods are now available in RSiena, which is a package of the statistical system R. This replaced the older Windows-based Siena version 3, which still is available but no longer maintained. RSiena can be executed on all platforms for which R is available: Windows, Mac, and Unix/Linux. R, a free software environment for statistical computing and graphics, can be downloaded from http://cran.r-project.org, and RSiena can be loaded as one of the packages in R. In fact, the number of packages for social network analysis in R has grown tremendously. But using these R packages requires that you be very familiar with the R environment.

E-Net and EgoNet

Whereas the two aforementioned specialized programs focus on specific statistical methods, these final two specialized programs provide a suite of routines tailored specifically to measure various aspects of ego networks. Made by the developers of UCINET, E-Net (Borgatti, 2006) uses attribute data of ego and alters, as well as ties among alters. Measures of the network's composition, heterogeneity, homophily, and structural holes are calculated on all selected egos, and ego networks can also be visualized. Because ego networks more readily meet the independence assumption required for OLS models, analyses using these ego-level measures can be done using statistical packages like SPSS, Stata, SAS, and so forth. One useful feature of

E-Net is that it facilitates the process of performing these analyses by getting the data into a format that is readable by these statistical packages.

EgoNet (McCarty, 2003) also provides these same analytical functions, but it also contains helpful routines that assist with questionnaire development and data collection. These questionnaires can incorporate name generators and interpreters and provide a very helpful framework to configure and perform a survey interview. Once the data are collected, it calculates general ego network measures as a first step in data exploration. Another nice feature is that the data can be easily exported in formats that are readable by other general (e.g., Excel) or network software packages (e.g., UCINET).

Programs for Network Data Collection

These general and specialized programs for analyzing network data have made significant advances in recent years. In addition to these improvements in analytical capabilities, there have been developments in the tools needed to collect relational data. In addition to the data-collection capabilities of EgoNet, two other programs developed specifically for network data collection through surveys are noteworthy (Huisman & van Duijn, 2011). First is Network Genie, a web-based application that facilitates the design and management of social network survey projects (Hansen & Reese, 2008). With this package, you can create online survey questionnaires that collect network data on egocentric or complete networks. Either type of survey can include social network items (questions about people in the network) or person-centered items (questions about the person who is completing the survey). Once these different types of data are collected, the program also allows formatted data to be exported to the social network analysis programs of your choice.

Second is the Organization Network Analysis (ONA) survey tool (Optimice, 2012), which was first designed to help organizations gather information about people within and across formal organizational boundaries. Similar to Network Genie, this application can create web-based network surveys and process these data prior to analysis. The surveys designed with ONA can either be (1) person centric, consisting of a number of questions about each respondent, or (2) question centric to evaluate the relationship for each question. In addition, the sampling approach can be bounded, measuring relations within a predefined group, or snowball, where an initial group of respondents is asked to nominate alters, who are then asked to nominate even more alters. Conveniently, this application also has export functions that prepare data for analysis in other applications. The interfaces of these other applications have significantly altered the amount and quality of network data that can be collected. It is strongly advised that network data be collected through means such as these because they reduce respondents' burden and errors in data entry.

Future Directions in Educational Applications

While advances in software—both for network data collection and analysis—will certainly influence the number of educational applications that employ social network analysis, the growth of social network analysis will be influenced by even more substantive reasons. Specifically, as emphasized in Chapter 2, social network analysis has the potential to become an increasingly central part of educational research, as it facilitates a paradigm shift from methodological individualism to methodological transactionalism, in which dynamic networks and communication processes are the primary focus of data collection (McFarland, Diehl, & Rawlings, 2011). This shift can be seen in diverse empirical examples, ranging from teacher professional communities (Penuel et al., 2009), school redesign networks (Daly & Finnigan, 2010), and cyber-bullying (Kowalski & Limber, 2007) to the integration of technology into schools (Frank et al., 2004). The growth of network thinking in educational research will continue, as social network analysis offers educational researchers a means for better capturing complex interdependencies and fluid dynamics than many other current and more utilized methods. As this potential becomes realized, what does the future hold for social network analysis and educational research? McFarland, Diehl, and Rawlings (2011) offer the following two hints.

The Collection and Analysis of Large-Scale Dynamic Behavioral Data

One of the most important developments in recent years is simply that there is more and significantly richer behavioral and communication data for researchers across all fields to work with. Most obviously, this refers to the familiar assortment of streaming and interconnected information that is readily available on the Internet in the form of information including text, images, videos, communication, and organizational records that can be rendered into network relations. Even beyond this, already-available data are the technological advancements that are making the collection of streaming behavioral more feasible. One well-known example of this kind of work comes from the Reality Mining project at MIT (Eagle & Pentland, 2006). As part of this project, research participants were given cell phones that continuously recorded their location, the presence of other participants, and all phone calls and text messages. Using these data, researchers could directly model the network of interaction between participants and study its contents in terms of communicative features like expressions of sentiment in text usage and voicing qualities. You could easily imagine extensions that would also allow, for example, the collection of biophysical data from students in classrooms, including physiological change and shifts in body position during interactions.

Longitudinal Analyses

A second future direction identified by McFarland, Diehl, and Rawlings (2011) for social network analysis in education is in the use of longitudinal network models, which help distinguish selection versus influence processes in tie formation. The misattribution of selection effects to social influence, or vice versa, has often led educational researchers to the wrong conclusions about potential causal mechanisms. The longitudinal analysis of social networks, the Holy Grail for network researchers (Wasserman & Robins, 2005, p. 6), has been made possible in the last few years since the development and availability of accessible methods for longitudinal network analysis. The most popular of these methods are the actor-oriented models developed by Tom Snijders and his colleagues available in the statistical package RSiena (reviewed above). These models, first introduced at the end of Chapter 10, are essentially longitudinal ERGMs that combine regular panel data (e.g., individual attitudes) with network panel data (i.e., relational measures collected at separate time points). Importantly, even though network data in this work are generally measured at discrete intervals, the methodological assumption is that relationships are (potentially) evolving states that may change between observations.

The empirical work utilizing longitudinal network methods is just in its earliest stages, but early work on adolescent friendship networks is already beginning to tease apart selection and influence processes related to issues relevant to adolescents, such as drug use (Pearson, Steglich, & Snijders, 2006) and smoking (Mercken et al., 2009). In both cases, the authors find that over time, there is a process of both selection and influence as peers both seek out other “deviants” as well as influence each other's behavior. It is easy to imagine how such models could be employed to examine networks’ influences on students’ achievement, attitudes toward school, and degree attainment, among others. Existing work outside of the network tradition has already argued for this reciprocal relationship between selection and influence, but utilizing dynamic network analysis allows educational researchers to better specify the mechanisms at work and understand how they shape each other through time.

Limitations

As social network analysis becomes more prevalent in educational research and advances in these directions, it will still be inhibited by a number of limitations that will shape both how it is practiced and the degree with which it is accepted. Valente (2010) encourages you to consider three of these as you design your network studies. First, most complete network studies are performed on populations in settings with clearly demarcated boundaries (students in a classroom, for example). This lacks the inferential abilities of random sample designs, as each of these settings is unique, thereby questioning whether even inferential results are generalizable to other populations. Second, many educational researchers are understandably uncomfortable with the tendency of social network analysis to quantify—in essence, reducing to a number—the complexity of interpersonal relations. Finally, despite the impressive advances in statistical tests involving nonindependent network data, there are lingering concerns regarding their appropriateness and value.

A Second Note on Research Ethics

In addition to these limitations, there are still a number of ethical concerns surrounding the collection of primary data and social network analysis. Chapter 2 noted that conducting social network analysis poses some ethical challenges for both researchers and institutional review boards. It is worth reiterating a few key points about the ethical issues related to social network analysis, especially when working with children in classrooms and schools. As with all social research, your primary concern must be on the potential harm to your study's participants. One potential harm is that it is more likely than not that a participant will know who else is participating in the study. Here, the process of informed consent is critical (Prell, 2012). Be sure to clearly explain to your participants how you plan on using the data and that the data will be treated as anonymous once they have been collected and processed. Also be sure to guarantee confidentiality—you will not share information gained from one respondent with another. In addition, as it becomes easier to mine electronic data for use on social network studies (for example, digital exchanges between a school's teachers and parents), the issue of privacy becomes critically important. These data cannot be collected without the consent of the entity that “owns” them, which in this instance can be the district, individual teacher, or even media platform through which these exchanges occur. It is advisable to work with your institutional review board as you design your study.

Final Thoughts

There is an increasing recognition and appreciation among educational researchers that relationships are key antecedents and determinants of individual behaviors and attitudes. The social network approach presented throughout this text gives you a rigorous and flexible means through which the importance and influence of these relations can be studied in a variety of empirical settings. This focus on relationships, as opposed to attributes, will foster your ability to develop a more complete, rich understanding of the ways in which networks and people are connected and evolve. To that end, I hope this book has given you the insight and tools needed to answer the question posed at the end of Chapter 1: *How does an understanding of social networks help you make sense of educational opportunities and outcomes at the individual and aggregate levels?*

http://dx.doi.org/10.4135/9781452270104.n12