15

UNDERSTANDING GROUPS FROM A NETWORK PERSPECTIVE

Noshir S. Contractor

NORTHWESTERN UNIVERSITY

Chunke Su

UNIVERSITY OF TEXAS AT ARLINGTON

There is a long-standing, albeit modest, tradition of applying network approaches to the study of groups starting with the launch of the Group Networks Laboratory at MIT shortly after World War II by Alex Bavelas. However, many group researchers, especially those not familiar with network research methods, are often frustrated and challenged by this methodological approach. Thus the goal of this chapter is to offer readers a pragmatic guidebook based on our experiences applying a social network approach to studying groups.

A fundamental commitment to incorporate relational and structural explanations along with individual and group level factors distinguishes the social network approach from other perspectives on analyzing groups (Contractor, Wasserman, & Faust, 2006; Katz, Lazer, Arrow, & Contractor, 2004; Kilduff & Tsai, 2003; Monge & Contractor, 2003; Wellman, 1988). This emphasis on incorporating relational explanations implies that the social network approach examines the group from a multilevel perspective by spanning the individual level (member attributes, such as expertise and satisfaction), the dyadic level (information retrieval, trust), and the overall group level (group density, centralization). We begin this chapter with a brief discussion of important research questions that motivate the use of a social network approach to study critical group processes. Second, we share with you our experiences on collecting network data from groups. The third section describes the "sausage-making" process of manipulating, visualizing, and analyzing network data. We conclude this chapter by addressing the challenges and limitations of network research in its current state, as well as its future.

Social Network Approach to Studying Groups

A social network is defined as a collection of social entities (termed as *nodes*) that are connected by one or more types of relationships (termed as *ties*) (Scott, 2000; Wasserman & Faust, 1994). When a group is conceptualized as a network, the nodes typically include individual group members, and the network ties can be several types of relationships among group members: social communication, professional collaboration, trust, information retrieval and allocation, advice-seeking, perception of expertise, etc. Drawing upon our past research experiences, we suggest that a social network perspective would be particularly applicable and useful to examine four facets of group processes.

Group formation

We find ourselves increasingly participating in ad-hoc, distributed, virtual, and transient groups both at work and socially. Therefore, an increasingly important question is for us to understand why people form groups and how do those formation mechanisms influence the outcomes of these groups. Intuitively we can conjecture that our prior networks influence the groups we join and our experiences in these groups in turn shape our future networks. In the past half-decade, there has been some promising and intriguing research that have built on these intuitions by drawing on network approaches to address questions of group formation and assembly (Cummings & Kiesler, 2005; Guimerà, Uzzi, Spiro, & Amaral, 2005; Jones, Wuchty, & Uzzi, 2008).

Information retrieval and allocation

Propelled by the ongoing digital revolution, members of groups have unprecedented autonomy and choice in determining from whom (or from where) they can retrieve information or with whom they can share or to what repository they can allocate information. This unfettered ability does not imply that members engage in random acts of information retrieval and allocation. Instead, it underscores the importance of understanding the social motivations that explain these nonrandom behaviors. Research over the past decade has begun to uncover the motivations for information retrieval and allocation behaviors, and much of this research illustrates the ability of network approaches to address these questions (Casciaro & Lobo, 2005; Contractor & Monge, 2002; Cross & Borgatti, 2004; Palazzolo, 2005; Su & Contractor, 2011).

Leadership in groups

There is an increasing appreciation that in contemporary groups, leadership is more accurately investigated as an emergent phenomenon than formally designated. A social network approach is particularly desirable to study emergent and informal leadership, hidden from the formal organizational chart (Cross & Parker, 2004) or decentralized, transient, and shared (Burke, Fiore, & Salas, 2003). Therefore, not surprisingly, the complex interplay between members' positions in the network and their emergence in leadership roles has been the subject of growing interest in recent years (Balkundi & Harrison, 2006; Huffaker, 2010).

Outcomes of group processes

Finally, there is a growing awareness that social network approaches contribute additional explanatory variance in understanding the outcomes of group processes, such as performance and satisfaction. For instance, group members tend to be more satisfied with group work when actively retrieving information from others than passively receiving unsolicited information allocated from others in the network (Su, Huang, & Contractor, 2010). In a meta-analysis of 37 studies on naturally existent groups, Balkundi and Harrison (2006) concluded that groups with denser social networks among their members tended to achieve better performance and higher cohesiveness. At the group level, they found that groups that were central in their intergroup network tended to perform better as well (Balkundi & Harrison, 2006).

The four group phenomena summarized above illustrate why social network approaches have a growing relevance to the group research. Armed with this motivation, we next delve into the pragmatics of collecting network data.

Collecting Network Data from Groups

In this section, we share our experiences based upon a program of research over the past decade involving three large interdisciplinary projects investigating networks and groups. In the *first* project, we investigated how networks could help us better understand what motivated members of a team to retrieve or share expertise about certain topics with specific other members of the team. Our research, which investigated over two dozen teams from organizations in government, private, and public sector in the US and Europe, uncovered that members did not always go to those whom they identified as experts. There were other network motivations that explained their retrieval and allocation behaviors. In the *second* project, we have been investigating how networks can help us understand team formation in massively multiplayer online role-playing games. Our research, which is investigating thousands of online teams ranging in size from three to 70, indicate that decisions on whom to invite on to teams are driven both by social factors and the need to enlist members with specific skills. The networks among assembled teams have a systematic impact on the performance of these teams. Our *third* project is investigating the formation and leadership among virtual and co-located interdisciplinary research teams in the areas of nanoscience, translational science, and oncofertility. Here we find that co-authorship, citation, and prior collaboration have systematic but nonlinear impacts on teams' success in submitting successful proposals, publishing highly cited articles, or developing highly utilized software. Next, we describe the approaches we adopted and our experiences with the collection of network data.

Broadly speaking, we can collect two types of network data: *whole-network* (or census network) and *egocentric* network data (Marsden, 2005). Whole-network data refers to the complete set of data available from each and every member within the group. The egocentric network concerns the focal (or "ego") member's network connections with others. In many cases, the "ego" members are also asked to provide information about their perceptions of network ties that might exist among their contacts. Egocentric research and whole-network research require different methods of data collection (for detailed distinctions between the two designs, see Marsden, 2005). In general, greater insights can be obtained by collecting whole-network data. For instance, if one is interested in investigating the extent to which a centralized group will have higher or lower performance, it would be essential to collect whole-network data. However, if one is interested in the extent to which an individual's satisfaction with the team is explained by the satisfaction of other team members whom they trust, it would be sufficient to collect egocentric network data.

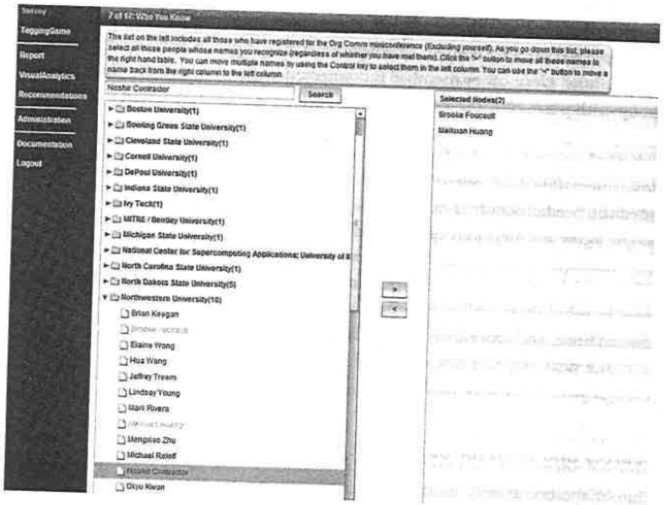
From a practical standpoint, when conducting research in small groups, it would be prudent to collect whole-network data. Collecting whole-network data in large groups is more time-consuming and challenging than collecting individual data or group data that is based on individual attributes only. When studying a group as a whole network, researchers need to collect data about each individual, as well as how each individual is connected with everyone else in the same group. Further, in some instances, we have collected cognitive social structure (CSS) data (Krackhardt, 1987a) where we ask group members not only how they are connected with every other member in the group, but also their perceptions of how every other member is connected with one another. Therefore, in a group with a size of n, the unit number of relational data to be collected amounts to n $(n - 1)$ (directional network) or n $(n - 1)/2$ (undirectional network). An empirical example can be found in a study we conducted at a city's public works department (Heald, Contractor, Koehly, & Wasserman, 1998), where we investigated the predictors of co-workers' perceptual congruence (the degree to which people agree on their perceptions of the organization's social network structure). Our findings showed that department employees' perceptual congruence was influenced by their similarities in formal organizational structure, gender and racial homophily, as well as emergent network ties such as social communication, acquaintance, and workflow relationships (Heald et al., 1998).

In the next two sections, we will discuss in detail the methods and procedures through which network data can be obtained from groups.

Sources and tools for data collection

One of the commonly used techniques to collect social network data is survey and questionnaire methods (Marsden, 2005). Traditionally, pen and paper-based surveys were widely used for respondents to mark and report their relationships with each other by thumbing through the hard copies. Since the inception of the World Wide Web, online network surveys have become increasingly popular and desirable. In recent years, there has been a rapid development of advanced web tools to collect social network data. Such web-based software is not only a data collection portal, but can also be a data visualizer and analytic tool. C-IKNOW (Cyber-infrastructure for Inquiring Knowledge Networks on the Web) is a tool that we have developed to support our research on networks in groups (Contractor, 2009a).

We decided to use online network survey tools over paper-based and traditional online non-network surveys for three reasons. First, respondents can enter their attribute and relational data via an interactive web interface. When answering network questions in the online survey, respondents are able to select their relational contacts by filtering the target person's attributes (e.g., one's organizational or group affiliation) or searching the target person's name (see Figure 15.1 for an illustration of such features in C-IKNOW). In subsequent questions, the online survey will only display those selected contacts rather than repeatedly displaying all respondents listed in the survey. This filtering mechanism is especially helpful and desirable when respondents are connected with only a small portion of a very large network. In addition, if a respondent communicates with a person who is not on the contact list in the survey, the respondent can add this person into the network question. Then other respondents have the option to log back into the survey and report their network connections to this person.



The list on the left includes all those who have registered for the Org Comm miniconference (Excluding yourself). As you go down this list, please select all those people whose names you recognize (regardless of whether you have met them). Click the ">" button to move all these names to the right hand table. You can move multiple names by using the Control key to select them in the left column. You can use the "<" button to move a name back from the right column to the left column.

Boston University(1)
Bowling Green State University(1)
Cleveland State University(1)
Cornell University(1)
DePaul University(1)
Indiana State University(1)
Ivy Tech(1)
MITRE / Bentley University(1)
Michigan State University(1)
National Center for Supercomputing Applications; University of Il
North Carolina State University(1)
North Dakota State University(5)
Northwestern University(10)
Brian Keegan
Brooke Foucault
Elaine Wong
Hua Wang
Jeffrey Treem
Lindsay Young
Mark Rivera
Meikuan Huang
Mengxiao Zhu
Michael Roloff
Noshir Contractor
Okyu Kwon

Brooke Foucault
Meikuan Huang

FIGURE 15.1 C-IKNOW interface: selecting relational contacts in the network.

Second, after the network data are collected, we have used C-IKNOW to provide respondents with network visualizations and algorithm-based recommendations based on the data they have provided. One critical challenge to network data collection is to obtain complete and quality data from group members. One novel solution to this challenge is to give respondents some potential payoff for providing complete and quality data. In one study we conducted at a large food and beverage products firm, we were interested in helping assess the effectiveness of distributed teams that were charged with the design of new food products. The design of the study required each member of the team to answer several questions about their own areas of expertise and their network relationships (such as prior or current collaboration) with other members of the team. Following the online survey, and with prior agreement from all group members, each member of the team was given the opportunity to log back into C-IKNOW and use it to identify who on the distributed team had expertise on a particular topic. Further, they could visualize how they might be directly or indirectly connected through network relationships with this individual. Clearly the quality of search results provided to the team member would depend on the quality of data entered by the members of the team. Thus it was in the team members' individual and collective interest to provide high-quality data as part of the survey. More generally, in our past experiences, the provision of network visualizations, access to metrics and recommendations incentivize respondents to provide accurate data. Of course, as mentioned in the above example, there needs to be a prior agreement about which responses provided by members will be shared among all members.

Third, respondents to network surveys are more likely to experience physical and mental fatigue due to the extended length and complexity of data inquiries. We designed the online survey to make the web interface as simple, visually appealing, and user-friendly as possible. The survey also prompts respondents to take breaks during the process and makes it easier for them to resume the survey from the point where they have left off.

The above online network survey tools are suitable for collecting whole-network data or egocentric network data. However, there are other computer programs specifically designed for collecting egocentric network data. For example, EgoNet (2011) is an open-source software for collecting egocentric network data developed by Christopher McCarty and his colleagues at the University of Florida. Researchers can set up network surveys by using EgoNet. In addition, EgoNet provides basic network metrics and data matrices which can be analyzed in other visual-analytic programs of social networks.

Although the survey method may be the easiest and most straightforward way to identify network connections in a group of individuals, it only collects subjective and self-reported data from the respondents. Usually these data are provided in retrospect and mediated by respondents' memory. In addition, the survey method is inadequate in capturing large-scale network data and in rich format. To overcome these limitations, we have joined with others who have a growing interest in exploring *observational* and *archival* approaches to collect more objective and dynamic network data in group settings.

First, researchers can acquire network data through observational coding. Here is a simple scenario: researchers join a group meeting of a software development team, and take notes on who is talking to whom as well as the frequency and length of each conversation. However, when studying larger, media-supported groups that may be geographically co-located or dispersed, there is a new generation of observational technologies that hold a lot of promise for capturing rich network data in teams. GroupScope (Poole et al., 2009) is an example of a new generation of an observational suite of tools that is being designed to include a multitude of high-definition video cameras, audio recorders, infrared sensors, accelerometers, RFIDs, and other observational instruments, to automate procedures to collect, annotate, and code large quantities of video and audio data generated by groups. Sandy Pentland at MIT, James Kitts at Columbia, and Tanzeem Choudhury at Dartmouth, and their colleagues have been at the forefront of developing these tools and utilizing them in their research (Wu, Waber, Aral, Brynjolfsson, & Pentland, 2008; Wyatt et al., 2011). Commodified versions of some of these observational technologies are now becoming available for use in research laboratories (Pentland, 2008).

An additional benefit of using observational technologies is to broaden our vision of what constitutes network data within groups. While survey methods typically provide data at one or a few points in time, observational technologies provide network data at intervals of a second or even less. These high-resolution network data, sometimes referred to as network event data, open up the possibility of developing new theories as well as methods for understanding the emergence and outcomes of group processes. The data collected from observational technologies may not be relational in nature. For example, observational technologies, such as RFIDs, might provide positional data for each group member. Additional processing would then be required to infer the distance between any two group members. Thus researchers need to utilize precise algorithms to identify network connections from observational data. For example, Mathur, Poole, Pena-Mora, Contractor, and Hasegawa-Johnson (2009) have developed an algorithm to map network linkages from video data combined with transcriptions of interactions within groups.

Another effective method is to collect *archival* data from groups by harvesting group members' digital traces from information systems that record members' behaviors (Lazer et al., 2009). For instance, as mentioned earlier, we are part of a large interdisciplinary effort to study teams within massively multiplayer online role-playing games (MMORPGs). MMORPGs serve as an excellent context in which to study how individuals (or more accurately their characters or avatars) with different skill sets need to come together in teams in order to accomplish

quests and raids to kill monsters or obtain resources. This study relies heavily on our getting access to anonymized server logs maintained by the developers of these games. These logs chronicle every single action (such as making a product or moving to a location), interaction (such as who is chatting with whom or broadcasting a message to a team), and transactions (such as buying a weapon or gifting a wardrobe item to another player) carried out in the game. These digital traces can be extremely large in size. Our team is working with the dataset that is over 40 terabytes! Here again we face challenges in developing algorithms to extract network relationships accurately from the corpus of digital traces. This research has provided new insights into why people choose to team up with specific other individuals and with what implications for their performance (Williams, Contractor, Poole, Srivastava, & Cai, 2011).

Finally, we have also explored understanding the dynamics of teams by harvesting archival network data from publicly available digital sources such as links among individuals' websites, bibliometric databases, and text-mining documents written by individuals. As more research in various fields of science and engineering in the past few years are being carried in teams, there has been a growing interest in understanding and enabling "team science" (Börner et al., 2010). This field of inquiry, sometimes referred to as the Science of Team Science, is particularly focused on trying to explain why some teams are more effective than others based on their affiliations (reflected, for instance, by links among their websites), prior collaborations (reflected, for instance, by co-authorship gleaned from bibliometric databases), and common interests (reflected, for instance, by similar use of concepts obtained by text mining their documents). This research illustrates the potential of leveraging external digital data that are available on the Internet such as bibliometric, web crawling, and text mining of transcripts.

Survey data collection procedures

Before the actual data collection procedures start, the boundary of the network to be studied needs to be defined (Marsden, 2005). That is, how do we decide who is considered as a member of the network? Monge and Contractor (1988) observed that this boundary could be based on space and time. They proposed three criteria for who to include based on the space dimension: common attributes such as membership and affiliations, participating in a particular type of relationship, and common activities. In the time dimension, they distinguished between cross-sectional and longitudinal network research. We have used two approaches to administer longitudinal network surveys. One is to aggregate the number of group interactions periodically (i.e., at equal time intervals). For instance, we have collected information retrieval network data among established teams in the online *World of Warcraft* game at three points in time at monthly intervals (Wotal, Green, Williams, & Contractor, 2006). This enables us to understand the emergence of experts in the team. A second approach is to collect data at specific stages or cycles of the group. For instance, we have used this approach to study the network structure of teams in online games when they first form and compare it to the teams' network structure at the conclusion of joint combat activities.

The above procedure is relatively straightforward if we are studying a formally assembled, well-defined, and pre-existing group. However, we need to be more cautious when we are examining ad-hoc, emergent, and bona-fide groups or where we are collecting network data to study group formation. Such groups have permeable boundaries, dynamic goals, and flexible group structures (Putnam & Stohl, 1996). Thus, it is more challenging to make a clear-cut decision on who should or should not be included in the network. Particularly if we are investigating the group creation process, we need to make sure to update the network formation continuously. Whenever a new member joins the group, we need to remind participants to provide their relational data with this new member in the survey. When someone exits the group, we may want to withdraw her or his relational data with others from the survey. However, we might want to retain information about those who exit the team if the research question is to understand why they may have left the team voluntarily or involuntarily.

Pre-survey interview

An additional strategy we have found useful to help researchers define the network boundary is to set up interviews with one or more members of the group (preferably the leaders or supervisors of the group) prior to data collection. Even when the network boundary has already been defined, the pre-survey interview can help verify the accuracy and validity of the determined boundary. In some cases it might reveal the need to include in the network some key individuals who are not technically on the team but have crucial interactions with the team.

Typically, we gather the following information about the group in the pre-survey interview: the name of each member (including individuals' nicknames if those are used widely at work), the basic structure of the group, the major tasks of the group, the primary knowledge areas required to complete group tasks (this is particularly applicable if we are studying information retrieval and allocation and want to ask questions about specific areas of expertise), the timeline of group work, the technology and resources utilized in the group (for multidimensional network research), the relative amount of interaction (e.g., daily, weekly, monthly, etc.) and other key information that might inform the design of specific questions. These meetings might also be used to discuss if and how much of the network data that is collected would be shared with the participants or their supervisors. As mentioned earlier, in some cases individuals' responses can be used to provide the team with a recommender system that can be used by the team to help them identify appropriate experts. While these systems will encourage respondents to provide more accurate data, it also means that some of their responses will be shared with other members of the team. These issues need to be negotiated prior to the design of the survey, and need to be included in any consent form that will be signed by respondents. Finally a pre-survey meeting is also helpful in enlisting a letter mailed from an influential individual in the organization encouraging participants to complete the survey.

As discussed elsewhere, unlike most other surveys, network surveys are only valid if the response rate is 100 percent or very close to it. This is because in a network survey, each respondent is providing information not only about themselves but also about every other person in the network. Therefore, a single missing respondent affects the data for all respondents.

Customization of survey

Based on the information garnered from the pre-survey interview, we create a survey customized for each group. If an online network survey is to be utilized, we need to add all participants' contact information to the survey database, such as their names, departmental affiliations, and email addresses. As mentioned previously, a major advantage of the online survey is to minimize the effort of each respondent. Therefore, it is always a good idea to pre-populate into the online survey as much information about respondents as we are able to collect in advance. This will give respondents a chance to review the material and modify it only if they identify certain errors. We then provide each participant with a unique login and password to access the survey online. Systems like C-IKNOW automate the procedure of emailing the URL of the survey to participants with their login information. Right after the opening welcome screen introducing the survey, participants must be presented with a consent form. The consent form should state clearly if some of their responses will be revealed in post-survey visual-analytics to group managers or all group members (e.g., the recommendation system mentioned earlier in this chapter).

Based on our own research experiences, for a group of 18 participants, a single network question would take a participant approximately five minutes to finish. Thus the entire survey may take a group member several hours to complete. As a result, we have on occasion given participants a generous time span (from days to weeks) to finish the survey at one's own pace. One advantage of online surveys is the ability for an individual to complete part of it at one time and then return to complete the rest at a later time. C-IKNOW provides the researcher with a dashboard indicating at any point in time how many questions have been completed by each respondent. Because the response rate is so crucial, we routinely and periodically send out reminders to participants encouraging them to complete their surveys.

The network survey is distinguished from most other surveys by its focus on relational questions. In a network survey, we are often interested in asking about different types of relationships between group members: interpersonal communication, trust, information retrieval and allocation, perception of expertise, advice-seeking, affect, collaboration, workflow, etc. (Krackhardt, 1992). This enables us to answer questions about how one relationship (such as trust) between two group members might influence a second relationship (such as information retrieval) between them. Later in this chapter we will discuss how to conduct these analyses.

After deciding on the specific relations one might want to measure, we next need to decide on the scales we use. We can use different scales and instruments to collect different kinds of network data: binary data (representing the presence and absence of certain relationships), continuous or valued data (differentiating degrees of certain relationships), and cognitive social structure data (which collects the perceptions of each respondent about relations among all members in the network).

The relational questions can be further categorized into sublevel context specific questions. For example, researchers can ask respondents to report their information retrieval behavior in a specific knowledge domain and repeat the same question for every other knowledge domain.

It is crucial to pay special attention to the importance and consequences of phrasing relational questions. For example, a question asking about "how often do you talk to your group members" would lead to a directional network of intragroup communication, whereas a question asking about "how often do you talk with your group members" results in a nondirectional network. This illustrates how what might appear to be a relatively trivial choice in wording can yield two very different kinds of network data. Further, relational questions can be phrased differently to elicit respondents' *desired* relationships or *actual* relationships. For example, the question "how often would you retrieve information from your group leader" might reflect a desired relationship and the question "how often did you retrieve information from your group leader" would imply an actual relationship. Likewise, there is a difference between *abstract* vs. *concrete* questions. An example of an abstract relational question would be "In a typical week, how often do you retrieve information from each group member?" A concrete question would be "In the past week, how often did you retrieve information from each member in the group?" An abstract question might be more appropriate if one is looking for general trends in the network that do not fluctuate dramatically. A concrete question might be more appropriate if one is looking for more accurate responses over a shorter (more recent) time frame.

Just as in non-network surveys, researchers should be cognizant of the challenges posed by confounding effects of social desirability. It is possible that respondents are giving researchers simply "what researchers want to hear" or "what make them look good" instead of genuinely reporting their perceptions and behaviors. Further, we have found that, in many cases, respondents are more hesitant to provide candid responses when answering network questions (especially on sensitive relationships such as personal affect and trust), because they are asked to reveal their perceptions of someone with whom they work or interact closely.

To alleviate these concerns, researchers should reiterate for the respondents the confidentiality and anonymity of survey data prior to data collection. This is a substantial challenge that is not always overcome. For instance, in a study we conducted among members on an emergency responders team, several participants refused to rate the extent to which they trusted other members because they did not feel "comfortable" sharing that information even with assurances of confidentiality.

Another strategy we have used to minimize socially desirable responses is to consider the ordering of questions carefully. For instance, it is not a good idea to ask respondents which members they retrieve information from on a particular topic, and follow that with another question asking them to rate the expertise of each member on the same topic. In order to reduce their cognitive dissonance, respondents might be tempted to rate highly the expertise of those from whom they reported retrieving information. This would contaminate the validity and reliability of the findings in the relationship between the two variables. One way to overcome such problems is to ensure that such questions are spaced far apart within the survey. We have also found it useful to encourage respondents to take breaks when answering a lengthy network survey. This helps mitigate respondent fatigue and improves overall completion rate.

If the data are collected online, it is critical that the survey portal directs respondents to where they have left off when they log back into the survey after taking the break. Finally, another strategy to reduce socially desirable responses is to make sure that we state the purpose of the research project to respondents in fairly general terms without making specific reference to research goals and research questions.

Administering data collection

To facilitate the data collection process and enhance the data quality, our first preference is to assemble all respondents in one room and administer the network survey physically in person. It is important to provide each respondent sufficient space so that they have privacy while responding to their surveys. Even if the network survey is to be completed online or in electronic forms, we have found that a face-to-face administration (in a room that has computers or where respondents bring their laptops) greatly increases the likelihood of collecting quality and complete data from participants. If a face-to-face meeting with respondents is not possible (as in the case with distributed teams), we have administered the online survey collectively at a scheduled time where we communicate with all respondents via conference call and stay on the line to answer any questions and offer clarifications. Even with the best of intentions, we have always encountered a few individuals who are unable to participate in a collective session. In such cases, we have invested the time and effort to schedule a one-on-one meeting (in person or via the phone) with individual participants to walk them through the survey questions. While these approaches might seem fairly labor-intensive, the critical importance of obtaining a high response rate justifies the return of investment.

At the beginning of the survey administration, in person or collectively, online or offline, we begin by briefly explaining the general goal of the research project and the procedures of survey completion. Next we ensure that all respondents have reviewed and approved the consent form. If the survey is administered online, respondents can indicate their willingness to "sign" a consent form by clicking on a button that says "Agree." Most of the respondents in our surveys have not previously encountered network questions. Since many of the group members are not accustomed to reporting relational data, it is crucial to clarify the privacy and confidentiality of the data they provide. This is particularly important if researchers are collecting sensitive relational data such as interpersonal trust, affect, or preferential choice. In cases where participants are given the option to access and review network visualizations or other network metrics following the survey, we make sure orally that they are aware of the relevant text in the consent form prior to them signing it.

Many of the strategies outlined above are motivated by the importance of a 100 percent response rate in network research. The 100 percent response rate refers to all respondents' completion of all survey questions, including individual-based and relational questions. In non-network research, there are standard protocols to deal with missing data without causing significant data loss both conceptually and statistically. However, as network data are relational in nature, if some participants do not respond, researchers would lose not only their data, but also the relational data they report about all other participants in the group. Hence, the lack of response by participants would render all their reported relationships with others incomplete and difficult to interpret. Clearly a small amount of missing data in network research would lead to considerable data loss. In addition, most social network analysis techniques do not have standard procedures for handling missing data. Any missing information in the data input would be invalid and cause errors in the analysis. Therefore, while a 100 percent response rate is not strictly required in traditional non-network research, researchers should make every effort to collect as complete network data as possible.

Of course, a 100 percent response rate cannot always be guaranteed in reality. Should researchers have to deal with missing network data, they can use some data manipulation methods to mitigate the negative impact of missing data on data analysis and interpretation, which will be discussed in a later section of this chapter.

Dealing with large samples

Groups vary in their sizes. In network terms, large samples could refer to either or both of the following situations: a network composed of one group with a large number of members, and a network composed of a large number of groups. Generally speaking, using the survey method to collect network data from large samples is extremely difficult and challenging. As the size of the network increases, the length of the relational questions in the survey expands considerably. Consequently, the data collection process becomes more time-consuming, and respondents are less willing and able to provide quality and complete data.

The large sample problem is especially salient in collecting whole-network data. One way to help resolve the large sample problem is to begin by focusing on only a small number of key members in the group. Then by using the respondent-driven sampling method (RDS; Heckathorn, 1997), researchers can invite these key members to identify other members with whom they are connected in a larger network. Further, the procedure can be repeated to survey even more members in the extended network. In this way, researchers can collect a large volume of network data without burdening every individual member for data input. In addition, the RDS approach is particularly helpful when investigating groups (such as "underground" groups) that might have a vested interest to remain concealed. In such cases, RDS provides an excellent strategy for researchers to develop trusting relationships with participants and utilize their network contacts to identify other members within the group.

The second method to help collect network data from large samples is to utilize diverse data collection methods and "mash" all the data into a coherent and meaningful data structure. When studying a large group, researchers can supplement survey methods with capturing group members' digital traces such as server log data and online behaviors (e.g., bibliometrics for scientific and research groups). These digitally generated data can be "mashed" with survey and observational data to provide a more comprehensive and enriched view of the group. For instance, in our study of groups in massively multiplayer online games, we invited players to complete an online survey. Since they logged into the game to complete their surveys, we were able to "mash" their survey responses with their online data obtained from server logs. In practice, the networks generated from each of these sources (surveys, server logs, bibliometrics) are stored in separate matrices where the rows and the columns represent the nodes in the network. We might conclude that each of these networks provides an important but incomplete indication of some underlying relationships (for instance, collaboration). In that case, "mashing" the three networks would imply adding the cell entries in the three matrices to generate a new network that might offer a more complete representation of the underlying relationships. The next section will discuss in detail how to mash and manipulate raw network data prior to data visualization and data analysis.

Manipulation of Network Data

After network data have been collected, researchers need to manipulate the raw network data to prepare them for visual-analytics. Manipulation typically refers to the process by which the "raw" network data are converted into the form which can be input directly into visualization and analysis software to explore specific research questions. Different visual-analytic tools may require different data inputs. Therefore, it is a common practice to manipulate the raw network data to make them suitable for a specific kind of analytic program. In some cases, researchers need to dichotomize the valued network data into binary data (either 1 or 0), because certain analytic programs or procedures require the input data to be binary. The cutoff value can be set to the mean score of the network, or the median score of the scales. It is worth noting that, in many instances, we are tempted to collect valued data from our respondents only to dichotomize them before conducting any analysis. If we can anticipate that the analysis we might want to conduct would only require binary data, we could have saved the respondent the additional effort in providing valued data. In other cases, researchers need to add multiple networks together, subtract one network from another, conduct cell by cell multiplications, and matrix multiplications.

We have used addition when we intended to combine two relations (such as advice and friendship) to generate a general measure of a "close" multiplex social tie. We have used subtraction where one of the relations we measured was the total amount of communication among team members, and the second relation we measured was the amount of task-related communication among team members. In this case, we subtract the latter from the former to generate a measure of non-task-related communication among team members.

We have used cell by cell multiplications when one of the relations measured was the extent to which each group member rated every other member's expertise on a topic, and the second relation measured was the frequency with which each member retrieved information on the same topic from every other member in the group. Cell by cell multiplication would provide a measure of the extent to which group members were retrieving information from those they considered knowledgeable.

Finally, we have used matrix multiplication in cases where we have group members reporting on their retrieval of information from multiple knowledge repositories. In this case the network is represented as a matrix where the rows refer to group members and the columns refer to knowledge repositories. In this so-called "bimodal" network, a cell entry of 1 in Row *i* and Column *j* indicates that the group member in Row *i* retrieved information from the knowledge repository in Column *j*. By multiplying this matrix with its transpose (where the matrix is flipped so that the rows now represent knowledge repositories and the columns represent the group members), we generate a new matrix where the rows and columns both represent group members and the cell entries represent the number of knowledge repositories from which they both retrieved information. The matrix algebra procedures to manipulate network data can be performed using a social network analysis software program such as UCINET (Borgatti, Everett, & Freeman, 2002) or by using R packages for Statnet (Handcock, Hunter, Butts, Goodreau, & Morris, 2003). The former program, which runs on Windows, is largely menu driven and has a low learning curve. The R packages for Statnet are Unix based and therefore platform independent. It requires a modest level of syntax writing and as a result has a relatively steep learning curve. But the tool has a great deal of flexibility and can be used on local computers as well as on higher performance cluster computing environments.

Dealing with incomplete network data

In practice, especially when collecting network data from a large group or a number of large groups, missing data are sometimes inevitable despite the best efforts of researchers. A government analyst who studies terrorist groups once remarked, "It is difficult to get terrorists to complete our network surveys."

When there is only an insignificant proportion of missing data, there are several remedies to minimize the loss of information in the data. Depending on the theoretical and conceptual nature of the variable being measured, we can recode the missing value into a new value that would be meaningful and valid for data analysis. For example, if group member A reports a friendship tie with member B in the survey, but member B does not respond to the survey, we might choose to infer that a friendship tie exists from member B to member A. The rationale for this type of recoding is based on the reciprocal nature of friendship relationships. We adopt this strategy when we believe that it is more likely for respondents to commit an error of omission rather than an error of commission.

In other cases, we have chosen to recode the missing value to 0 to signify the absence of a friendship from B to A. We adopt this strategy when we believe that it is more likely for respondents to overstate their friendship relationships, perhaps motivated by social desirability. If we have no plausible intuition about respondents' motivation to respond in a certain fashion, we have assigned a random value drawn from a distribution with the mean and standard deviation of all the values reported in the network. By doing so, we acknowledge that the missing values might in fact be in error but the errors are randomized across all missing values and hence would not systematically bias the results of any subsequent analysis.

Finally, if the recoding scheme is hard to justify based on one of the rationales outlined above, we have considered the option of removing the nonrespondent participants from the network. While the reduced network will now have complete data, we run the risk of excluding certain key members who did not respond but might have been the recipient of network ties from several respondents. In more than one of our studies we have identified key members who were "just too busy" to complete our network survey. Removing them from the network would clearly be counterproductive. Therefore the removal of missing data should be cautiously and deliberately used.

Visual-analytics of Network Data

After the raw network data have been manipulated, they are ready to be visualized and analyzed. The visual-analytics of network data are the ultimate instrument – a "macroscope" – to uncover the signature structures of network data. Broadly speaking, we undertake three tasks in this realm: visualization, descriptive metrics, and inferential statistics.

Visualization

Network visualization is a graphic illustration of the nodes and their linkages embodied in the network data. It serves as a visual aid to uncover the network structures, as well as a basic diagnostic tool to check the validity and accuracy of the network data. There are a number of network visualization programs that allow researchers to visualize their network data in customized layouts, such as different nodal sizes (to visualize continuous nodal attributes such as level of expertise), nodal colors (to visualize categorical nodal attributes such as areas of expertise), link widths (to visualize the strength of the network link), and network layout (to visualize clustering or other macro patterns in the network).

Huisman and Van Duijn (2005) provide a comprehensive and critical review of several network visualization tools. A recent addition to the suite of visualization tools is NodeXL (Hansen, Shneiderman, & Smith, 2010), which is a template for Excel 2007 and 2010 that lets you enter a network edge list, click a button, and see the network graph, all in the Excel window. Most network visualization programs offer limited analytic capabilities, but can import and export data to other network analytic programs to enable a seamless visual-analytic process.

We often conduct some network visualizations before network analytics to discern the basic network structure prior to data analysis. This procedure is also useful to make sure the data appear to be valid in light of all the manipulations we discussed earlier. But we also find considerable merit in utilizing visualization tools after conducting the analysis. The post-analytic visualization enables the incorporation into the visualization of metrics computed as part of the network analysis. For example, in NetDraw (Borgatti, 2002), a visualization tool that is built into UCINET, researchers can choose to display the size of the nodes in the network to represent degree centrality (the number of links connected to the node). The nodal color could indicate membership in a cluster identified by the network analysis, and the width of the link could represent the structural equivalence between two nodes (the structural equivalence is a network metric that indicates the extent to which the two nodes have similar patterns of interaction with all other nodes in the network).

In addition, network visualization can help researchers drill down into interesting facets or subregions of the network configuration and better understand the analytic results. For example, if a network analysis shows that a specific member has the highest betweenness centrality in the network, the visualization would be the most illustrative way to demonstrate the brokerage role of this member in the group (i.e., by connecting those members who are not directly connected with each other).

Descriptive metrics

Network scientists have developed a suite of descriptive metrics to analyze various properties of a social network at five distinctive levels: the individual, the dyad, the triad, the subgroup, and the global level (for a review, see Easley & Kleinberg, 2010; Wasserman & Faust, 1994). At each level, network analysis focuses on different descriptive metrics to measure different facets of network properties.

At the *individual* level, the key descriptive statistics include degree, betweenness, and closeness centralities of an individual member. Degree measures the extent to which a group member has a large number of direct network links. Betweenness measures the extent to which a group member connects group members who are not directly or weakly connected with one another. Closeness measures the extent to which a group member can reach all other group members via direct or indirect network links. Very often researchers will compute individual level network metrics and then use those as dependent or independent variables in non-network analytic procedures such as regression or ANOVA. For instance, researchers have used the degree centrality of an individual in the network as one of several variables to predict their level of leadership in the group (Huffaker, 2010).

At the *dyadic* level, the focus is placed on the relational properties of a pair of nodes in the network, such as reciprocity, redundancy, and structural equivalence. For example, we can measure the extent to which group members mutually seek advice from one another, by computing the ratio of the number of observed reciprocal ties (where A and B seek advice from one another) as a proportion of the number of possible reciprocal ties which is $n*(n-1)$ for a network of size n.

The *triadic* level focuses on metrics of three nodes and their relationships at a time, including transitivity and cyclicality. For instance, we can measure the extent to which if in a group, A trusts B and B trusts C, then A also trusts C. This can be computed by calculating the ratio of the number of observed transitive triads as a proportion of the total number of transitive triads. A researcher might posit that groups with higher levels of transitivity are more likely to experience higher levels of team identification.

At the *subgroup* level, the components and cliques metrics are calculated to measure the extent to which subgroups of individuals are cohesively connected in a network. For instance, a subgroup level analysis of the information retrieval network might reveal the presence of clear factions resulting in a fractured group where two sets of individuals only retrieve information from others within their own sets but not from the other set. A researcher might posit that groups that can be partitioned into factions based on their information retrieval network will underperform as compared to groups that are more cohesively connected.

Finally, the *global* level considers the network as a whole and examines the properties of the entire network such as density and network centralizations. For instance, a group would be highly centralized if one member is connected to all other members but the rest of the members are only connected to this one person who would then be the star of the network. The group would have low centralization if, for instance, each member of the network only has links with two other members. Some classic studies conducted by Bavelas (1948) over six decades ago have shown that highly centralized groups outperform decentralized groups on simple tasks. However, members in these highly centralized groups report on average lower levels of satisfaction than those in decentralized groups. As mentioned above, these descriptive metrics can be utilized to augment the illustrative power of network visualizations. Further, these metrics can be used as independent or dependent variables, or both, for other types of non-network analysis.

Inferential statistics

Visualizations and descriptive metrics are necessary but not sufficient tools to understand fully the antecedents and consequences of the structural signatures embedded in the network. For instance, descriptive network statistics discussed above can provide us with a measure of the extent to which there is reciprocity, transitivity, or centralization in the network. But what it does not provide is a statistically defensible measure of whether the observed reciprocity, transitivity, or centralization is significantly more than what we would expect by chance. This is where we turn to inferential statistics. The descriptive network statistics are analogous to measures of central tendency, such as the mean, in non-network analysis. Inferential network statistics are analogous to parametric tests such as the *t*-test or nonparametric tests such as the chi-square test in non-network analysis.

Unfortunately, most of the techniques used to compute inferential statistics in non-network analysis cannot be applied to network analysis. This is because a large proportion of inferential statistics used in non-network analysis make the assumption that the data are independently and identically distributed. But network data observations are not independent of one another. That is, the presence of a communication tie between individual A and B could conceivably impact the presence of a communication tie between individual A and some other individual C.

Most standard statistical analyses that focus on attributes of (rather than relations among) individuals are premised on the assumption that the data are drawn from a distribution where the observations are independent. For instance, the height of an individual A does not impact the height of an individual B. Thus, many of the standard statistical techniques used to analyze attribute social scientific data are not appropriate for analyzing network data. As a result, inferential statistics for network data are unable to use techniques that could violate the assumption of independence. Thus it is imperative to use specialized social network analytic techniques rather than traditional statistical methods for inferential hypothesis testing.

One common genre of hypotheses that is of considerable interest to group researchers is the extent to which one network relation among group members is positively or negatively associated with other network relations among group members. For example, we have examined the extent to which if A trusts B, A is more or less likely to retrieve information from B. In this case, we compute a simple correlation to measure the extent to which the trust link between two members in the network is accompanied by an information retrieval link.

Let us assume, that the correlation coefficient was 0.35. In order to test our hypothesis, we would need to establish if this value is significantly greater than 0. This would not be a problem in non-network analysis where the correlation coefficient would be accompanied with a *P* value indicating the likelihood that this value is greater than 0. However, since this correlation was computed on network data (which violate assumptions of independence), we cannot use the significance test provided with the correlation coefficient. Instead, we have to draw upon one of several specific analytic techniques developed by network statisticians.

Quadratic Assignment Procedure (QAP) is one popular technique to test the significance (Krackhardt, 1987b). To assess the relationships between more than two networks, researchers can use Multiple Regression Quadratic Assignment Procedure (MRQAP) (see e.g., Doerfel & Barnett, 1999; Krackhardt, 1988). QAP and MRQAP, both of which are available in network analysis software programs such as UCINET and StatNet, are reasonable approaches to test hypotheses about similarity among two or more network relations. But what if we are interested in assessing the extent to which there is a higher than expected level of transitivity in a single network (such as the advice network discussed earlier)? Or, what if we are interested in hypothesizing that an information-retrieval relation from group member A to B is explained not only by the extent to which A trusts B, but also the extent to which A trusts other members in the group who in turn trust B?

Recent developments in ERGM (Exponential Random Graph Modeling) analysis (also known as *p** analysis) provide a promising framework to test complex network hypotheses such as these (Frank & Strauss, 1986; Robins & Pattison, 2005; Wasserman & Pattison, 1996). In essence, ERGM/*p** analyses test the likelihood for the theoretically hypothesized structural properties to occur in the observed network (Robins, Pattison, Kalish, & Lusher, 2007). Researchers (Robins et al., 2007; Shumate & Palazzolo, 2010) have described how ERGM/*p** analyses can be used to uncover structural signatures in the observed network, thus reflecting the underlying social processes that generate such network structures.

Limitations, Challenges, and the Future

As we close this chapter, we must acknowledge the presence of a very large "elephant in the room" – the ethical challenges of network approaches. Unlike non-network research, a network study can never be truly anonymous. It makes little sense for respondents to report who they communicate with if we cannot establish the identity of the respondents! As a result, group researchers utilizing network approaches bear an additional burden in terms of meeting ethical standards. Indeed, over the years, proposals for network research projects have been met with more than their share of skepticism by members of Institutional Review Boards (IRBs) who have a commitment to protect human subjects. It is therefore not surprising that strategies to meet the requirements of the IRB have been the topic of discussion at several gatherings of researchers interested in network approaches.

While anonymity can never be upheld for the reasons outlined above, network researchers also bear a special burden in meeting requirements of confidentiality. That is, what results might one share with the respondents which would both be perceived as useful and not violating confidentiality? Ironically, the problems of confidentiality are greatest in small groups.

Consider the case where a faculty member conducted a "confidential" network analysis among a dozen students as part of a graduate seminar discussing network methods. Prior to taking a break during the three-hour seminar, students were asked to complete a confidential network survey listing whom they considered as their friends in the class. During the break, the faculty member drew a network visualization of the friendship network on the board without including the names of any of the students. As students returned from the break, they began to discuss the visualization on the board. They were drawn to the fact that one node in the network had listed all the other nodes in the network as friends, but none of the other nodes had reported a friendship link to this node. Even as the students reconvened for the second part of the seminar, many had made educated guesses about the identity of this node. There was no discussion specifically about the identity of the node during the seminar. However, for reasons that may or may not have been triggered by this event, the student did not return to class in the following days and ended up dropping out of the graduate program. This case illustrates how, despite the best of intentions and safeguards, network approaches still require the researcher to be extremely vigilant about potential ethical breaches.

A second potential ethical "landmine" deals with the use of archival digital trace data. Consider the case of the data we analyzed from a MMORPG. One of the games we are investigating is EverQuest II developed by Sony Online Entertainment. We presented a paper based on some of our results at a recent annual meeting of the American Association for the Advancement of Science (AAAS). In our presentation at the AAAS, we had indicated that all the data we were provided was anonymized and that we did not have access to any content of the chat. Given the wide audience for this meeting, it was not surprising that the findings were picked up by the popular press and by the blogosphere. Some of the stories reporting our findings failed to mention that the data were anonymized and did not include the content of the chat. We were in for an unpleasant surprise when we discovered that these stories had created quite a furor on some of the forums frequented by EverQuest II players who were understandably irate that Sony Online Entertainment might have contributed to an ethical breach by sharing with our research team personal and private information about the players without their permission. Even though no ethical breach was conducted, it heightened both Sony Online Entertainment's and our research team's sensitivity about the player's concerns.

Armed with this greater appreciation of the ethical challenges, we hope this chapter has illustrated why and how we have found network methodology to be an important "arrow" in the quiver of tools to advance our understanding of groups. The use of network approaches to study groups is by no means a recent phenomenon. As we mentioned at the start of this chapter, Alex Bavelas founded the Small Group Networks Laboratory at MIT shortly after World War II. After an initial flurry of activity, network approaches to the study of groups languished for several decades. But in the past decade, there has been a resurgence of interest in the use of network approaches to studying groups.

There are at least four reasons that explain this renewed interest. *First*, there is a much greater intellectual interest spurred by the societal appreciation of the role of networks as the primordial soup from which groups emerge. The trend from formal, long-term and heavily structured teams towards more agile, distributed, and ad-hoc teams in the contemporary workforce have underscored the role of networks. *Second*, the increasing prevalence of digital traces makes it much easier to capture copious amounts of network data through observational and archive methods. This mitigates one of the perennial challenges of social network approaches that rely heavily on labor-intensive (in particular, respondent-intensive) network surveys. *Third*, the recent methodological development in inferential statistics for network data outlined earlier in this chapter have finally enabled network researchers to augment exploratory network analysis (based on descriptive statistics) with the ability to test complex network hypotheses using confirmatory network analysis. *Finally*, recent developments in computational infrastructure from the desktop all the way to petascale computing and cloud computing have been crucial enablers in conducting network analyses.

As mentioned earlier, network analyses cannot rely on many of the standard statistical techniques developed for non-network data where one can make the assumption of independent observations. Instead, the statistical techniques developed specifically for network data tend to be very computationally intensive. It is not uncommon for sophisticated statistical models analyzing networks in relatively small groups to take up to an hour on the state of the art desktop machine. It is therefore not surprising that many of us have begun to exploit high-performance computing to conduct network analyses.

These four developments – renewed intellectual interest, new forms of digital data, recent developments in network methods, and advances in computational capabilities – argue well for the utilization of network approaches to advance contemporary group research. These developments also hold the promise for helping reconceptualize our notion of the group. Traditionally, the nodes in social network research are restricted to human members only, given its focus on "social" structures as opposed to impersonal networks such as the computer network or electric power grids. In empirical group research, the social network approach has been employed to examine the information retrieval relationship among group members (Palazzolo, 2005), the structures of information sharing and their effects on group member satisfaction (Su, Huang, & Contractor, 2010), and the effects of network structures on group performance (Rulke & Galaskiewicz, 2000). However, given the fast advancement of new media and web-based information technologies, there is an increasing demand for considering the "nonhuman" nodes when studying groups as social networks (Contractor, Monge, & Leonardi, 2011; Hollingshead & Contractor, 2002; Su & Contractor, 2011).

In recent years, we have begun to witness a transformation in our conceptualization of group to include not only human members but also digital agents such as Web 2.0, the Semantic Web (Shadbolt, Hall, & Berners-Lee, 2006) and Cyberinfrastructure (Atkins, 2003). The integration of both human and nonhuman nodes in the network inspires and requires researchers to conceptualize networks in a new way: as multidimensional networks. Contractor (2009b) defines a multidimensional network as a collection of multiple types of nodes together with multiple types of network ties among them. The nodes in a multidimensional network are "resources", including people, documents, datasets, analytic tools, instruments, concepts, and keywords (Hollingshead & Contractor, 2002).

The network ties represent different types of relationships between people and people, people and nonhuman nodes, and amongst nonhuman nodes themselves (Contractor, 2009b). For example, a multidimensional network of a software development team could include team members collaborating with each other, team members writing and publishing codes on the team intranet, the intranet reporting debugging procedures of the software, and the software being tested by different tools and by different members (Poole & Contractor, 2011). The inclusion of nonhuman nodes in the multidimensional network makes the data collection process even more complicated and time consuming. For example, in the human-only information retrieval network, researchers may only need to know "who retrieves information from whom" in the group. However, in the multidimensional network, it is important to collect data on "who is retrieving information from whom and/or which data repository." In short, the multiplicity of nodal types and their relationships in the multidimensional network demand some creative and innovative approaches to collection and collation of network data. We expect that in the future network approaches will be increasingly influential in helping us understand and enable groups conceptualized from this multidimensional perspective.

Acknowledgments

The authors wish to acknowledge the following funding sources that contributed to the development of materials presented in this chapter: Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053; Air Force Research Laboratory FA8650-10-C-7010 and National Science Foundation Grants: BCS- 0940851, CNS-1010904, OCI-0904356, IIS-0838564, IIS-0841583. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory, Air Force Research Laboratory, the National Science Foundation or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

Atkins, D. (2003). *Revolutionizing science and engineering through cyberinfrastructure.* Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure.

Balkundi, P., & Harrison, D. A. (2006). Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. *Academy of Management Journal, 49*(1), 49–68.

Bavelas, A. (1948). A mathematical model for group structure. *Applied Anthropology*, 7, 16–30.

Borgatti, S. (2002). *NetDraw: Graph visualization software.* Harvard, MA: Analytic Technologies.

Borgatti, S., Everett, M. G., & Freeman, L. C. (2002). *Ucinet 6 for Windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

Börner, K., Contractor, N., Falk-Krzesinski, H. J., Fiore, S. M., Hall, K. L., Keyton, J., et al. (2010). A multi-level systems perspective for the science of team science. *Science Translational Medicine.* 2:49cm24.

Burke, C. S., Fiore, S. M., & Salas, E. (2003). The role of shared cognition in enabling shared leadership and team adaptability. In C. L. Pearce & J. A. Conger (Eds.), *Shared leadership: Reframing the hows and whys of leadership* (pp. 103–121). Thousand Oaks, CA: Sage.

Casciaro, T., & Lobo, M. S. (2005). Competent jerks, lovable fools, and the formation of social networks. *Harvard Business Review, 83*(6), 92–100.

Contractor, N. (2009a). *C-IKNOW: Cyber-infrastructure for inquiring knowledge networks on the Web*. Evanston, IL: Science of Networks in Communities (SONIC), Northwestern University.

Contractor, N. (2009b). The emergence of multidimensional networks. *Journal of Computer-Mediated Communication, 14*, 743–747.

Contractor, N. S., & Monge, P. R. (2002). Managing knowledge networks. *Management Communication Quarterly, 16*(2), 249–259.

Contractor, N., Monge, P., & Leonardi, P. (2011). Multidimensional networks and the dynamics of sociomateriality: Bringing technology inside the network. *International Journal of Communication, 5*, 1–20.

Contractor, N., Wasserman, S., & Faust, K. (2006). Testing multi-theoretical, multilevel hypotheses about networks: An analytic framework and empirical example. *Academy of Management Review, 31*(3), 681–703.

Cross, R., & Borgatti, S. (2004). The ties that share: Relational characteristics that facilitate information seeking. In M. H. Huysman & V. Wulf (Eds.), *Social capital and information technology* (pp. 137–161). Boston, MA: MIT Press.

Cross, R., & Parker, A. (2004). *The hidden power of social networks*. Boston, MA: Harvard Business School Press.

Cummings, J. N., & Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Study of Science, 35*(5), 703–722.

Doerfel, M. L., & Barnett, G. A. (1999). A semantic network analysis of the international communication association. *Human Communication Research, 25*, 589–603.

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. New York: Cambridge University Press.

EgoNet (2011). EgoNet software. University of Florida. Available online at: http://sourceforge.net/projects/egonet/.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association, 81*(395), 832–842.

Guimerà, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science, 308*, 697–702.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2003). *Statnet: An R package for the statistical modeling of social networks*. http://csde.washington.edu/statnet/.

Hansen, D., Shneiderman, B., & Smih, M. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. New York: Morgan-Kaufmann.

Heald, M., Contractor, N., Koehly, L. M., & Wasserman, S. (1998). Formal and emergent predictors of coworkers' perceptual congruence on an organization's social structure. *Human Communication Research, 24*, 536–563.

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems, 44*(2).

Hollingshead, A. B., & Contractor, N. S. (2002). New media and organizing at the group level. In L. Lievrouw & S. Livingstone (Eds.), *Handbook of new media* (pp. 221–235). London: Sage.

Huffaker, D. (2010). Dimensions of leadership and social influence in online communities. *Human Communication Research, 36*(4), 593–617.

Huisman, M., & Van Duijn, M. A. J. (2005). Software for social network analysis. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 270–316). New York: Cambridge University Press.

Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science, 322*, 1259–1262.

Katz, N., Lazer, D., Arrow, H., & Contractor, N. (2004). Network theory and small groups. *Small Group Research, 35*(3), 307–332.

Kilduff, M., & Tsai, W. (2003). *Social networks and organizations*. London: Sage.

Krackhardt, D. (1987a). Cognitive social structures. *Social Networks, 9*, 104–134.

Krackhardt D. (1987b). QAP partialling as a test of spuriousness. *Social Networks, 9*, 171–186.

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analyses of dyadic data. *Social Networks, 10*, 359–382.

Krackhardt, D. (1992). The strength of strong ties: The importance of philos in organizations. In N. Nohria & R. Eccles (Eds.), *Networks and organizations: Structure, form and action* (pp. 216–239). Boston, MA: Harvard Business School Press.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., et al. (2009). Social science: Computational social science. *Science, 323*(5915), 721–723.

Marsden, P.V. (2005). Recent developments in network measurement. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 8–30). New York: Cambridge University Press.

Mathur, S., Poole, M. S., Pena-Mora, F., Contractor, N., & Hasegawa-Johnson, M. (2009). *Detecting interaction links in a collaborating group using manually annotated data*. Working paper, National Center for Supercomputing Applications, University of Illinois Urbana-Champaign, Urbana, IL.

Monge, P., & Contractor, N. (2003). *Theories of communication networks*. New York: Oxford University Press.

Monge, P. R., & Contractor, N. S. (1988). Measurement techniques for the study of communication networks. In C. Tardy (Ed.), *A handbook for the study of human communication: Methods and instruments for observing, measuring, and assessing communication processes* (pp. 107–138). Norwood, NJ: Ablex.

Palazzolo, E. (2005). Organizing for information retrieval in transactive memory systems. *Communication Research, 32*(6), 726–761.

Pentland, A., (2008) *Honest signals: How they shape our world*. Cambridge, MA: MIT Press,

Poole, M. S., Bajcsy, P., Contractor, N., Espelage, D., Fleck, M., Forsyth, D., et al. (2009). *GroupScope: Instrumenting research on interaction networks in complex social contexts*. Working paper, National Center for Supercomputing Applications, University of Illinois Urbana-Champaign, Urbana, IL.

Poole, M. S., & Contractor, N. S. (2011). Conceptualizing the multiteam system as a system of networked groups. In Zaccaro, S. J. Marks, M. A., & L. A. DeChurch (Eds.), *Multiteam systems: An organizational form for dynamic and complex environments* (pp. 193-224). Routledge Academic.

Putnam, L., & Stohl, C. (1996). Bona fide groups: An alternative perspective for communication and small group decision making. In R. Y. Hirokawa & M. S. Poole (Eds.), *Communication and group decision making* (pp. 179–214). Thousand Oaks, CA: Sage.

Robins, G., & Pattison, P. (2005). Interdependencies and social processes: Dependence graphs and generalized dependence structures. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 192–213). New York: Cambridge University Press.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks, 29*(2), 173–191.

Rulke, D. L., & Galaskiewicz, J. (2000). Distribution of knowledge, group network structure, and group performance. *Management Science, 46*(5), 612–625.

Scott, J. (2000). *Social network analysis: A handbook* (2nd ed.). Newbury Park, CA: Sage.

Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The semantic web revisited. *IEEE Intelligent Systems, 21*(3), 96–101.

Shumate, M., & Palazzolo, E. T. (2010). Exponential random graph (p*) models as a method for social network analysis in communication research. *Communication Methods and Measures, 4*(4), 341–371.

Su, C., & Contractor, N. (2011). A multidimensional network approach to studying team members' information seeking from human and digital knowledge sources in consulting firms. *Journal of the American Society for Information Science and Technology, 62*(7), 1257–1275.

Su, C., Huang, M., & Contractor, N. (2010). Understanding the structures, antecedents and outcomes of organisational learning and knowledge transfer: A multi-theoretical and multilevel network analysis. *European Journal of International Management, 4*(6), 576–601.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York: Cambridge University Press.

Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: An introduction to Markov graphs and p*. *Psychometrika, 61*(3), 401–425.

Wellman, B. (1988). Structural analysis: From method and metaphor to theory and substance. In B. Wellman & S. Berkowitz (Eds.), *Social structure: A network approach* (pp. 19–61). Cambridge: Cambridge University Press.

Williams, D., Contractor, N., Poole, M. S., Srivastava, J., & Cai, D. (2011). The virtual worlds exploratorium: Using large-scale data and computational techniques for communication research. *Communication Methods and Measures, 5*(2), 163–180.

Wotal, B., Green, H., Williams, D., & Contractor, N. (2006). *WoW!: The dynamics of knowledge networks in Massively Multiplayer Online Role Playing Games (MMORPG)*. Paper presented at the Annual Social Network Sunbelt Conference, Vancouver, Canada.

Wu, L., Waber, B., Aral, S., Brynjolfsson, E., & Pentland, A. (2008) Mining face-to-face interaction networks using sociometric badges: predicting productivity in an IT configuration task. *Proceedings of the International Conference on Information Systems*, Paris, France. December 14–17, 2008.

Wyatt, D., Choudhury, T, Bilmes, J., & Kitts. J. A. (2011). Inferring collocation and conversational networks using privacy-sensitive audio. *ACM Transactions on Intelligent Systems and Technology*, 2:1.

16

ANALYZING GROUP DATA

Deborah A. Kashy

MICHIGAN STATE UNIVERSITY

Nao Hagiwara

WAYNE STATE UNIVERSITY

Human beings are inherently social animals, and most of what we do involves participation in groups. Starting at birth, we are members of families; as we grow, we become members of friendship groups, classrooms, sports teams, to name a few; in adulthood we often become members of new family groups as well as work groups, religious groups, and so on. In each of the groups to which we belong, our behavior is likely to be influenced by other group members. Even young babies show evidence of this interconnectedness. For example, babies who receive consistent warmth and attention from their parents tend to develop into secure adults. In school, children's learning can be affected by other students in the class as well as by their teachers. For example, having a very disruptive child in the class may negatively impact all students' learning, but having a highly motivated teacher can raise all of his or her students' learning. Likewise, in adulthood, people's productivity in the workplace can be affected by the quality of their relationships with co-workers, the leadership style of their manager, or even the broader corporate culture.

In each of these examples, the groups clearly vary on many attributes, such as group size and structure, but the fundamental aspect of all of them is that the outcomes for members of these groups are linked. In some cases these links may reflect actual interpersonal influence, but in other cases the similarity in outcomes may result because group members share the same environment (Kenny, Mannetti, Pierro, Livi, & Kashy, 2002). We conceptualize these links broadly as *nonindependence*: The degree to which outcomes for persons who are in the same group are more similar (or dissimilar) to one another than are outcomes for persons from two different groups. The goal of this chapter is twofold: (a) to highlight the data analytic challenges and opportunities involved in conducting research with small groups; and (b) to introduce three data analytic models that