Network Data and Statistical Models

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

By: Brian V. Carolan
Pub. Date: 2013
Access Date: November 16, 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: 185-212

Network Data and Statistical Models

Objectives

In this chapter you will learn about the different techniques that make use of simulations when statistically modeling network data. You will understand that these techniques vary according to the types of questions that are asked and whether the focus is on ties between actors in complete networks, individual attributes, or relations within and between groups. These models are important, as they enable you to employ social network analysis in ways that move beyond description toward explanation, an important goal of social science. Without getting into the technical aspects of each modeling strategy, you will begin to appreciate the ways in which statistical models can be used with network data to address questions that more completely reflect the complexity of educational phenomena. In addition, you will be introduced to a number of studies in the field of education that make use of these models.

Connecting Questions to Statistical Models

To address the types of questions asked at the beginning of the previous chapter, we need models that combine network (relational) data and individual attribute data and/or dyadic attributes. The statistical underpinnings of these advanced models were introduced in the previous chapter, and the models presented in this one reflect the most sophisticated and interesting strands of contemporary social network analyses. Following the lead of van Duijn and Huisman (2011), these models are categorized by the questions they are designed to address. Three different analytical emphases are presented: (1) relationship-level models that focus on the ties between actors in complete networks, (2) models that predict individual actors' attributes, and (3) actor-level models that emphasize the differences within and among groups of actors within a complete network.

The first three questions used to introduce the previous chapter can be answered with models that explain and predict the presence or value of a network's ties. To do so, these models require additional information on the network's relationships or actors, if available. The fourth question predicts an individual outcome (trust) from network data. The fifth and final question is answered with models that identify groups of actors who have the same probability distribution of ties to other actors (known as "stochastically" equivalent actors). Many of these models can—and often do—make use of additional covariate information. These different modeling approaches can be thought of as analyses that make use of the same relational data, but with a different emphasis. In the first (models for ties), the focal variable is the observed relationship between any two of the network's actors, which may be explained by attributes. In the third, for example, the focal variable is the actor's group membership expressed as an unobserved, latent variable whose value is the result of the observed ties among actors (i.e., a group assignment derived from one of the algorithms discussed in Chapter 5) and any other actor characteristics.

After briefly introducing the data to be used in this chapter and revisiting the modeling strategy used to analyze egocentric network data, this chapter's next three sections introduce some of the more cutting-edge approaches to the statistical analysis of social network data and show how these models can be used to address the five questions asked at the start of the previous chapter. First, the statistical analysis of ties within complete networks will be discussed, with a focus on how dyadic ties within a complete relational network can be modeled from both structural and attribute variables. Second, the chapter examines how actors' individual attributes (e.g., trust) can be modeled from both attribute (e.g., efficacy) and relational data (e.g., collaboration). Third, the chapter turns to a set of procedures that are used to model ties between and within groups of actors. The chapter ends with a brief review of recent advances in the statistical modeling of two-mode networks (bipartite graphs, which are networks in which the rows and columns consist of different actors) and an introduction to a handful of advanced approaches, including agent-based modeling and actor-oriented models.

Description of School Leaders Example Data

These ideas and tools will be demonstrated using a portion of the School Leaders data set, particularly observations in years 1 and 3 of the "collaboration" and "confidential help" networks. Collaboration is a directed valued network measured on five-point scale ranging from 0 to 4, with higher values indicating more frequent collaborations (1–2 times/week). Similarly, the confidential help network is a directed, valued network measured on the same scale. Complete network data are available on all 43 school leaders drawn from two large school districts. In addition to these relational data, the survey instrument in year 1 captured attribute data, including composite scores from a set of items related to efficacy (19 items) and trust (8 items), as well as information about leaders' gender and whether they worked at the district or school level.

These data are summarized in Table 9.1. For most of the analyses that follow, the collaboration and confidential help network data were dichotomized, with values originally coded as a 3 and 4 recoded as 1s, indicating the presence of a directed tie for both relations, zero otherwise. These dichotomized relations are reported in Table 9.1. A few of the analyses that follow rely on the original ordinal scores that range from 0 to 4. Measures are calculated on the complete network (Table 9.1, total column) and on the two subnetworks, defined by whether a leader works at the school or district level.

For the collaboration networks, the mean number of ties sent and received in year 1 is 1.72 and increases to 8.65 in year 3. This pattern is the same whether you work at the school or district level. For example, district-level leaders send and receive 2.00 and 2.06 collaboration ties in year 1, and these numbers increase to 12.72 and 9.78, respectively. Compared to leaders at the school level, district-level leaders send and receive, on average, more collaboration ties; for example, the average number of collaboration ties received (in-degree) by school-level leaders in year 3, nearly 2.00 ties less than district-level leaders. In addition, in year 3, district-level leaders have a higher standard deviation, suggesting that several leaders send or receive either very few or very many collaboration ties.

A similar pattern is evident for the confidential-exchange networks. On average, leaders send and receive less than 1.00 tie in year 1. Two years after the reform initiative, however, these same leaders send and receive, on average, 3.58 ties. While the average numbers of confidential ties sent and received by school- and district-level leaders are similar in year 1, district-level leaders send and receive about 2.00 more confidential ties than school-level leaders in year 3. This suggests that the reform initiative is related to the frequency with which leaders exchange confidential information and that this relation is especially salient for those working at the district level.

The bottom rows of Table 9.1 report on three attributes collected on 43 leaders in year 1. The trust scores are a composite derived from a set of eight items originally scored on a seven-point Likert scale modified from Tschannen-Moran and Hoy's (2003) instrument. Higher scores reflect high levels of trust in the leaders' schools and districts. School- and district-level leaders report similar levels of trust (4.90 versus 4.60), with school-level leaders having slightly more variation as evidenced by a higher standard deviation. The efficacy scores, derived from a set of 18 items from Tschannen-Moran and Gareis' Principal Efficacy Scale (2004), show that school-level leaders, on average, have scores that are about one standard unit higher than district-level leaders. Finally, 48% of the 25 school-level leaders are male compared to 39% of the 18 district-level leaders.

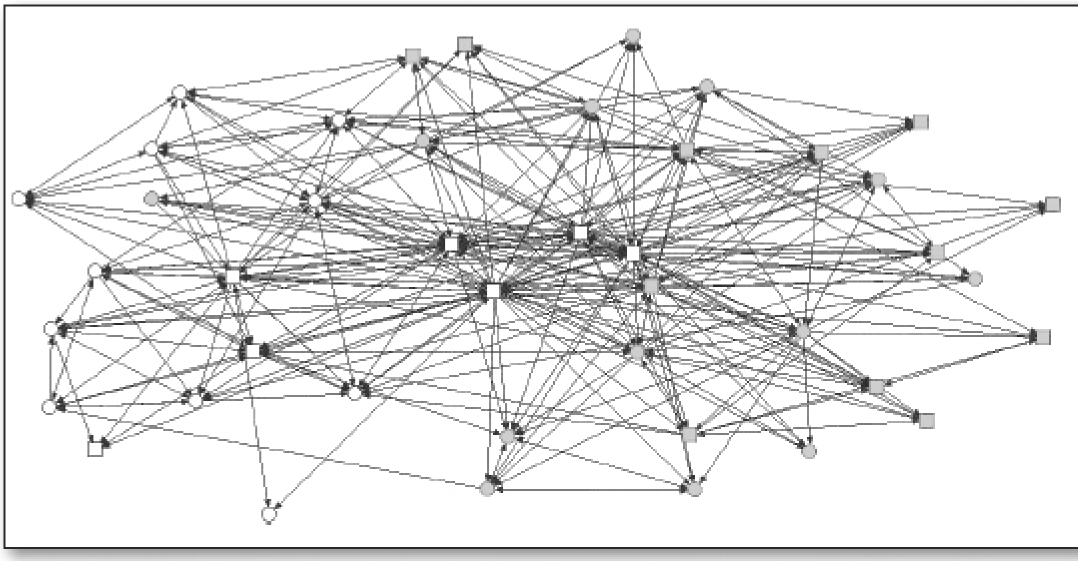
Table 9.1 Summary Statistics of the School Leaders' Collaboration and Confidential Exchanges Networks (Years 1 and 3) and Attribute Data.

	School-level M (SD)	*District-level M (SD)*	*Total M (SD)*
Collaboration year 1			
Out-degree	1.52 (0.94)	2.00 (0.94)	1.72 (0.98)
In-degree	1.48 (1.09)	2.06 (1.73)	1.72 (1.40)
Collaboration year 3			
Out-degree	5.72 (5.93)	12.72 (11.84)	8.65 (9.44)
In-degree	7.84 (1.63)	9.78 (4.21)	8.65 (3.10)
Confidential exchanges year 1			
Out-degree	0.60 (0.71)	0.56 (0.78)	0.58 (.073)
In-degree	0.44 (0.58)	0.78 (1.17)	0.58 (0.88)
Confidential exchanges year 3			
Out-degree	2.72 (3.08)	4.61 (5.59)	3.51 (4.35)
In-degree	2.64 (1.68)	4.72 (2.16)	3.51 (2.14)
Trust	4.90 (0.72)	4.60 (0.67)	4.77 (0.71)
Efficacy	7.01 (0.96)	6.12 (1.11)	6.64 (1.11)
Male	.48	.39	.44
N	25	18	43

Note: SD not reported for the male indicator variable.

The directed and dichotomous collaboration network in year 3 is shown in Figure 9.1. Each directed line represents a collaboration tie. At the center of the graph are four male district-level leaders that send and receive a large number of ties. Given that males constitute 44% of the sample, this "central" location is noteworthy. To confirm what this graph is showing, you could easily calculate one of the centrality measures presented in Chapter 5. The right side of the graph consists mainly of school-level leaders (colored nodes) that primarily, but not exclusively, collaborate among themselves, while the left side of the graph shows a similar pattern among district-level leaders. Not one of the leaders is isolated, indicating that the entire graph constitutes one weak component. This is a different picture from the same network in year 1, where there were three isolates (not presented). Using the information presented in Table 9.1 and shown in Figure 9.1, we can ask a number of questions that consider, for example, whether a certain configuration is expected and "normal," how one relation is associated with another among this same set of actors, or whether one network structure could be thought of as "better" than another. The rest of this chapter is dedicated to showing how a statistical approach can be used to address the issues.

Figure 9.1 Directed and Dichotomized Graph of the School Leaders Collaboration Network Year 3. Square nodes are males and clear nodes are district-level leaders. The right side of the graph consists mainly of school-level leaders (colored nodes) that primarily, but not exclusively, collaborate among themselves, while the left side of the graph shows a similar pattern among district-level leaders.



Statistical Models Using Egocentric Network Data

Before turning to these issues, it is worth reiterating a couple important points when using egocentric network data to make a statistical inference. First, recall that egocentric network data can be collected as part of a complete network study (as is often the case) or (ideally) drawn from a random sample of egos from some larger target population. Regardless of how they are sampled and collected, egocentric network data reflect a focal ego actor's (usually the respondent's) network from the perspective of ego. When drawn from samples, egocentric network data have to be transformed prior to analysis. The key is to transform egocentric network data into a dyadic format, in which each observation (case) in the data file consists of ego and one network alter, as was described in the previous chapter. Therefore, there will be one row in the data file for each one of ego's named alters. This transformation is done solely to facilitate statistical analysis.

Second, once the data are converted into this dyadic format, you can create various network measures that capture, for example, the size and diversity in one's ego network. However, the statistical analysis of these data requires that you control for the nonindependence of the observations. Having the data in this dyadic format facilitates this. One of the foundational assumptions of statistical analysis is that observations are independent. By definition, network data violate this assumption, with each ego nested in one or more dyads. Another level of clustering could be introduced if the design is longitudinal. Addressing these issues is simple enough if researchers employ hierarchical linear models or random effects models that adjust for the nonindependence of the data. However, such multilevel models (discussed at the end of this chapter) can become quite complex, and you should not employ these models without being familiar with multilevel regression techniques in general.

Modeling Ties in Complete Networks

An interesting question for the School Leaders data presented in Table 9.1 concerns whether collaboration among the school leaders increased or decreased from years 1 to 3, after taking into account each leader's overall trust in their administrative colleagues and whether they worked at the school or district level. Such a question implies modeling one complete network (collaboration at year 3) as well as attributes of each leader (trust and district/school level). There are several types of statistical models for modeling these complete network data with two distinctions among them: (1) the measurement level of the relational variable (binary or valued) and (2) and the statistical modeling tradition used to make the inference.

This section reviews a few methods used to address questions such as the one above. The first is QAP (quadratic assignment procedure), which is based on regression models and permutation tests for valued (i.e., continuous) relational variables (see Dekker, Krackhardt, & Snijders, 2007 for a review of its development). Second is a set of related models (p_1 and p^*) that predict dichotomous (nonvalued) dyadic relations among actors in a directed graph. These models focus on the ties between each possible dyad within a network, predicting its four possible outcomes (null, mutual, and two asymmetric states). In addition to these there are several other procedures that can be used to model ties in complete networks (see van Duijn & Huisman, 2011 for a review).

QAP/MR-QAP

When examining the relationship between two (or more) complete networks, you have to consider the dependence inherent in the data due to the fact that actors send and receive multiple ties. This results in the observed outcomes of the relational variable being nonindependent. You could use standard ordinary least square regression/correlation models to estimate the relationship between networks, with an implicit (and incorrect) assumption that there is independence within and between each of the network's dyads. As noted earlier, this would result in coefficients with grossly underestimated standard errors (and, consequently, artificially high test statistics). The quadratic assignment procedure developed by Hubert (1987) and Krackhardt (1987b) tests the null hypothesis of no correlation between the two networks and adjusts for this dependence between networks by repeatedly permuting the order of rows and columns of one of the networks while keeping the other network intact. The resulting sample of product–moment correlations provides the distribution of the correlation coefficient under the null hypothesis to which the actual observed correlation can be compared. This is similar to the idea discussed above of using repeated simulations—thousands of them—to determine probability. This procedure can be extended to examine the relationship between two networks, X and Y, while controlling for a third, Z. However, this becomes more complex because of the dependency of the two networks, X and Z, which is similar to the idea of multicollinearity in multiple regression. These procedures are referred to as MR-QAP (reviewed in Dekker, Krackhardt, & Snijders, 2007).

How can this procedure be applied to the School Leaders data? It is useful, for example, in determining whether there is a relationship between collaboration among school leaders in year 1 and how often they engage in confidential exchanges in year 3. Or you could examine the relationship between how frequently school leaders turn to each other to discuss issues of a confidential nature in year 1 and collaboration in year 3. The results of the QAP correlations between these four social networks with ordinal variables (recall that the ties between leaders originally range from 0–4) are presented in Table 9.2. These coefficients can be interpreted like your standard Pearson's product–moment correlations, with absolute values closer to 1 indicating stronger relations.

This correlation matrix reveals a few interesting relationships, two of which are relevant given the above questions. For example, the correlations are highest for the collaboration and confidential-exchange networks in year 1 (r = 0.56, $p < .001$) and collaboration and confidential-exchange networks in year 3 (r = 0.66, $p < .001$). Both of these relationships are somewhat predictable. Those who engage in one type of favorable exchange are also likely to engage in another at a similar point in time. On the other hand, what is interesting is that having one type of tie in year 1 is only weakly related to the same tie in year 3. For example, the correlation between collaboration networks in years 1 and 3 is 0.14, ($p < .001$). This type of analysis, while interesting in and of itself, is most useful when it informs multivariate models that include covariates that are not too strongly correlated.

Table 9.2 QAP Correlations for the School Leaders Network Data. Based on 5,000 permutations. The QAP correlations are highest for the collaboration and confidential exchange networks in year 1 (*r* = .56, *p* < .001) and collaboration and confidential exchange networks in year 3 (*r* = .66, *p* < .001).

	1	*2*	*3*	*4*
1. Collaboration year 1	–	.14*	.56*	.19*
2. Collaboration year 3		–	.12*	.66*
3. Confidential exchanges year 1			–	.18*
4. Confidential exchanges year 3				–

* = $p < .001$

Therefore, this analysis can be extended in order to predict collaboration in year 3 from the collaboration and confidential-exchange networks in year 1 networks. The question, therefore, is whether school leaders prefer to collaborate with those with whom they have collaborated in the past or with those that they have turned to discuss confidential issues. This is analogous to having two independent variables, with the obvious catch that these predictor variables are complete networks. Therefore, this would require a MR-QAP procedure that controls for the effect of the model's second predictor.

In Table 9.3, the results of an MR-QAP analysis are presented with the collaboration network in year 3 as the outcome variable. Model 1 starts with collaboration in year 1 as the sole predictor, and then Model 2 incorporates confidential exchanges year 1 as an additional predictor. Model 2, therefore, shows which of the two predictors has the stronger relationship with the collaboration network in year 3. When confidential exchanges in year 1 are added to the model, the relationship between collaboration in years 1 and 3 is reduced and no longer significant. The relationship of confidential exchanges in year 1 to the outcome is strong, as is demonstrated by the large increase in adjusted-R^2, the proportion of variance explained by the model. Therefore, while collaboration in year 1 does not significantly predict collaboration in year 3, confidential exchanges in year 1 does. From these results, you can intuit that collaboration among school leaders provides an important foundation for more sensitive, perhaps even deeper, relations (e.g., confidential exchanges) at a later point in time.

Table 9.3 MR-QAP Analyses Predicting School Leaders' Collaboration Network Year 3. Based on 2,000 permutations.

	Model 1	*Model 2*
Intercept	1.40	.90
Collaboration year 1	.24* (.71)	.02 (.41)
Confidential exchanges year 1		.77* (.57)
R^2 adjusted	.20	.43

* $p < .05$. Standard errors in parentheses.

*P*1 and *P** (*P*-Star)

The next two procedures also model ties in complete networks but do so in a manner that explains the presence (or absence) of dyadic ties as a function of individual-and/or network-level characteristics. So, while QAP and MR-QAP procedures control for network structure through permutations, these next two models attempt to explain it. For example, these models can address questions such as whether gender or some other individual attribute predicts confidential exchanges between school leaders, or does some previous relation have a stronger effect? Both procedures focus on the four different possible outcomes of a dyad in a directed and binary network (therefore, the relational data are nonvalued, 1s or 0s). In this respect, these models are closely related to logistic regression in that they analyze a dichotomous dependent variable (1/0) that is assumed to follow a binomial distribution. While both seek to predict the dyadic relations among actor pairs, the *p** model is able to do so while using relational attributes of each actor and of the network as a whole.

The key difference between them rests with the assumptions of the dependence graphs that are used in the simulations. The dependence graph is the simulated network that is used as the basis for comparison and specifically indicates the nonindependence that exists between a network's actors (Valente, 2010). Technical discussions of this distinction are available in Wasserman and Robins (2005) and Wasserman and Faust (1994). Briefly, the structure of the *p*1 dependence graph assumes that all dyads are statistically independent (Holland & Leinhardt, 1981), whereas the *p** model operates from a more realistic assumption that dyads are conditionally dependent. It is for this latter reason that *p** models are preferred. In addition, the *p** model can also include additional global features of the graph such as tendencies toward transitivity and the variance across actors in the propensity to send and receive ties. Both models are some of the first to make use of the ERGM (Chapter 8), which provides the basis for comparing whether a network's observed structural properties occur more frequently than you could expect from chance alone.

While the technical aspects of estimating the *p*1 and *p** models are complicated, the interpretation of its individual estimates is fairly straightforward. The basic idea of the *p*1 model is to understand relations between pairs of actors as functions of an individual's relational attributes (actor's tendencies to send ties and to receive them) as well as key features of the graph (complete network) in which the two actors are embedded (e.g., the overall density and overall tendency toward reciprocity; Hanneman & Riddle, 2005). Similarly, the basic idea of the *p** is to understand these same relations but to include actor-level and network-level attributes in the model (Pattison & Wasserman, 1999).

How might these models be used make inferences about the social processes at work in the School Leaders data? Using directed and dichotomized relational data, a *p*1 model can be used, for example, to test whether school leaders tend to reciprocate relationship choices. In other words, if one school leader turns to another to discuss something confidential, is the latter likely to reciprocate? *P** models are often employed to take this further by including actor- and network-level covariates. In the School Leaders data, these would include an actor attribute variable such as male and a network-level attribute such as transitivity. Therefore, the latter, *p**, can test whether a leader's gender, efficacy score, and other complete graph properties predict a confidential exchange between two leaders. Table 9.4 presents the results of both models, derived from UCINET and PNet, respectively (both applications are reviewed in Chapter 12). The book's companion website also provides a step-by-step guide that helps you perform your own analysis using the *p** model.

The results of these models suggest several interesting social processes that may explain confidential exchanges between school leaders in year 3. The analysis of the School Leaders data with the *p*1 (Model 1, Table 9.4) results in a negative estimate for density, implying that the probability of a confidential exchange tie in year 3 is much smaller than 0.50. The reciprocity parameter, on the other hand, is 4.11, slightly larger in absolute value than the density parameter. When the reciprocity value is positive and large, as is the case here, this indicates a strong tendency for mutual ties (Wasserman & Faust, 1994). If this parameter estimate were negative, it would indicate that there are many nonreciprocated confidential exchanges in year 3.

Models 2 and 3 (Table 9.4) go further by including two covariates, one binary measure for male (1 = yes, 0 = no) and one continuous efficacy measure. Parameter and standard errors are provided. Those estimates that are noted as significant are those that are approximately twice its standard error in absolute value. The two significant estimates in both models are structural parameters; that is, they are directly calculated from the network itself. For example, in Model 2, it is evident that there is a strong tendency for confidential-exchange ties to be reciprocated indicating an absence of hierarchy among the school leaders. That same model shows that there is little tendency toward transitivity, which is unsurprising given that a confidential exchange is likely to occur between two people, not among three. Model 2 also shows that being a male is not significantly related to either sending or receiving a tie. Model 3 adds a second attribute, efficacy, which is significantly related to sending or receiving a tie. This suggests that school leaders who believe that they have the capacity to have an effect are more likely to send and receive confidential exchange ties. This last finding speaks to a larger point regarding the importance of efficacy, in general.

Table 9.4 Results of *P*1 and *PAnalyses for the Dichotomized and Directed School Leaders Confidential Exchanges Network Year 3.**

	P_1 Model 1	P^* Model 2	Model 3
Density	–3.68		
Reciprocity	4.11	3.02* (0.30)	2.95* (0.33)
Arc		–3.27* (0.17)	–2.72* (0.51)
Transitivity		0.13 (0.12)	0.10 (0.11)
Sender male		–0.01 (0.22)	0.09 (0.19)
Receiver male		–0.07 (0.21)	0.03 (0.20)
Sender efficacy			–0.22* (0.05)
Receiver efficacy			0.13* (0.07)

Notes: Standard errors are in parentheses. A * indicates that an estimate is more than twice the standard error in absolute value.

The development of both the *p*1 and *p** models, especially the latter, reflects an important shift in the analysis of social networks. These were some of the first models to move from the representation and description of social networks toward a statistical approach that emphasized the explanation of relational ties between actors. *P** models explicitly model the influence of relational ties on network and individual characteristics, and because of this, they present an array of opportunities for educational researchers to analyze numerous issues. However, this potential has yet to be fully appreciated, as there are few examples of these models being used in educational research. This is likely due to the absence of didactic texts with substantive illustrations (Knoke & Yang, 2008), which fall outside the scope of this cursory introduction. One of the first and still most accessible primers on *p** models is provided by Anderson, Wasserman, and Crouch (1999), in which they employ this model to predict friendship choices within one fourth-grade classroom and then extend the analysis across a set of three classes and incorporate a covariate for students' gender. While showing a strong tendency for dyads to exhibit mutuality, they also clearly outline the strengths and potentials of this still-underutilized modeling approach. Two other and equally accessible primers are provided by Robins (2011), who demonstrates these models in a step-by-step manner using a new application, PNet, designed explicitly for *p** models, and Prell (2012, Chapter 10), who discusses various aspects of these models, including their probability distributions, the role of network configurations, and parameter estimation.

Modeling Actors' Attributes

The previous section described models and methods for testing the likelihood of a relation between any two actors in a given network using both local (micro) and global (macro) network properties. This section shifts the analytical lens to predict an individual actor's outcome, whether it is an attribute variable (e.g., a student's test score) or a structural variable (e.g., a teacher's betweenness centrality score), using relational data. For example: does a teacher's gender predict his or her influence (as measured by degree centrality)? This question relates an attribute (gender) to a measure of the actor's location in a network (degree centrality). We might be interested in the relationship between two (or more) aspects of actors' locations. For example: How much of the variation in teachers' degree centrality can be explained by their network's size and the number of cliques that they belong to? We might even be interested in the relationship between two individual attributes among a set of actors who are connected in a network. For example, in a school classroom, is there an association between students' engagement and their academic achievement?

Each of these questions focuses on variables that describe individual actors. These variables may be either nonrelational attributes (e.g., efficacy or gender) or variables that describe some aspect of an individual's relational position or affiliation (e.g., betweenness or clique membership). In most cases, standard statistical tools for the analysis of variables can be applied to describe differences and associations. But, as noted by Hanneman and Riddle (2005), standard statistical tools for the analysis of these variables cannot be applied to inferential questions—hypothesis or significance tests—because the actors are not independent observations randomly drawn from some large population. Instead of applying the normal formulas (i.e., those built into statistical software packages and discussed in most introductory statistics texts), we need to use other methods to get more correct estimates of the reliability and stability of estimates (i.e., standard errors). The “boot-strapping” approach (estimating the variation of estimates of the parameter of interest from large numbers of random subsamples of actors) can be applied in some cases; in other cases, the idea of random permutation can be applied to generate correct standard errors. Fortunately, these approaches to calculating standard errors are incorporated into specialized programs that analyze social networks. There are three common techniques that are typically part of these programs and can be used to model individual actor attributes that are nonindependent (as is the case with social network data).

T-Tests

Suppose we thought that school leaders who worked at the district level were less likely to collaborate with colleagues than those leaders who worked at the school level (year 1). This hypothesis can be tested by comparing the average out-degree (number of dichotomized collaboration ties sent) of district- and school-level leaders. Since each individual leader is an observation, the data are located in a column (or, sometimes, a row) of one or more files: one column indicates whether the leader works at the district or school level and the other column indicates each leader's ego network size (as measured by out-degree). For this test, 10,000 trials (this researcher can choose the number of trials) are used to create the permutation-based sampling distribution of the difference between the two means (comparing the mean out-degree between district- and school-level leaders). For each of these trials, the scores on network size are randomly permuted (that is, randomly assigned to district- or school-level leaders, proportional to the number of each type.) The standard deviation of this distribution based on random trials becomes the estimated standard error for our test. Figure 9.2 shows the results generated by UCINET.

The output in Figure 9.2 first reports basic descriptive statistics for each group. The group numbers are assigned according to the order of the cases in the file containing the independent variable. In this example, the first actor was a district-level leader, so that became Group 1 and school-level leaders became Group 2. From this output, you can see that the average number of collaboration ties by sent by district-level leaders (M = 2.00, SD = 0.94) is only 0.48 units higher than the average number of collaboration ties sent by school-level leaders (M = 1.52, SD = 0.94). This would seem to support the hypothesis that district-level leaders are more likely to send collaboration ties than school-level leaders. But tests of statistical significance suggest that this conclusion is unwarranted.

Figure 9.2 UCINET Output of *T*-test Between School- and District-Level Leaders and the Number of Collaboration Ties (Out-Degree) Year 1. From this output, you can see that the average number of collaboration ties by sent by district-level leaders (*M* = 2.00, *SD* = 0.94) is only 0.48 units higher than the average number of collaboration ties sent by school-level leaders (*M* = 1.52, *SD* = 0.94). This difference in means is not statistically significant.

```
TOOLS>STATISTICS>T-TEST
----------------------------------------------------------------

Dependent variable: col 6
Independent variable: col 4
# of permutations: 10000
Random seed: 481



Basic statistics on each group.

                      1      2
                  Group  Group
                  ------ ------
    1      Mean   2.000  1.520
    2   Std Dev   0.943  0.943
    3       Sum  36.000 38.000
    4  Variance   0.889  0.890
    5       SSQ  88.000 80.000
    6     MCSSQ  16.000 22.240
    7  Euc Norm   9.381  8.944
    8   Minimum   0.000  0.000
    9   Maximum   3.000  3.000
   10  N of Obs  18.000 25.000
   11 N Missing  25.000 18.000



SIGNIFICANCE TESTS


     Difference        ...One-Tailed Tests...     Two-Tailed
       in Means    Group 1 > 2    Group 2 > 1           Test
 ============== ============== ============== ==============
          0.480          0.078          0.959         0.1199
```

Differences as large as 0.48 in favor of district-level leaders happen about 8% of the time in random trials. So you would be taking an unacceptable risk of being wrong if you concluded that the data were consistent with the research hypothesis. Therefore, the null hypothesis cannot be rejected.

About ANOVA

A similar procedure can be employed to test hypotheses about the means of three or more groups with one-way analysis of variance (ANOVA). Suppose we divided the school leaders into three groups based on each leader's composite trust score: low, medium, and high. Those assigned to the low category have trust scores equal to one standard deviation below the mean, and those in the high category have scores that equal one standard deviation above the mean. Membership in one of these three categories is a column vector, with each leader coded as a 1, 2, or 3 (low, medium, and high). Using the same column vector that reports the number of the school leader's collaboration ties sent, we can ask the following: Do a leader's collaboration ties vary by their level of trust? Figure 9.3 shows the results of a one-way analysis of variance using these two column vectors.

Figure 9.3 UCINET Output of ANOVA Between the School Leaders' Level of Trust (Three Groups) and the Number of Collaboration Ties (Out-Degree) Year 1. The overall differences among these means is not significant ($F_{(2, 40)}$ = 0.94 and *p* = .42).

```
TOOLS>STATISTICS>ANOVA
------------------------------------------------------------------------

Dependent variable: Col 6
Independent variable: Col 5
# of permutations: 5000
Random seed: 21934


        ANALYSIS OF VARIANCE

         Source             DF            SSQ   F-Statistic  Significance
 ============== ============== ============== ============== =============
      Treatment              2           1.83        0.9423        0.4179
          Error             40          38.82
          Total             42          40.65

R-Square/Eta-Square: 0.045
```

The differences among these means are not significant ($F_{(2, 40)}$ = 0.94 and *p* = .42). In addition, the output also shows that the differences among group means accounts for only about 5% of the total variance in collaboration ties sent in year 1 among the leaders. Therefore, we cannot reject that null hypothesis that the number of collaboration ties sent in year 1 does not vary by the level of leaders' trust.

Regression

Where the attributes of actors that we are interested in explaining or predicting are measured at the interval level and one or more of our predictors are also at the interval level, multiple linear regression is a common approach. This approach computes basic linear multiple regression statistics by ordinary least squares (OLS) and estimates standard errors and significance using the random permutations method for constructing sampling distributions of *R*-squared and slope coefficients. Using this method, we can ask the following question: How well does a leader's perceived level of trust in his or her colleagues predict the number of alters to whom the person sent a collaboration tie in year 1, controlling for gender and the level at which the person works (district vs. school)? This question requires three vectors of independent variables (trust score, an indicator for gender, and an indicator for level) and one dependent variable vector (collaboration year 1 out-degree).

The results of this regression are shown in Figure 9.4. The top portion of the output shows the model R^2, 0.09, indicating that slightly less than 10% of the variance in the number of collaboration ties sent in year 1 is explained by these three covariates; that is, how much change in the outcome is explained by this set of covariates. In addition, based on permutation tests, this model's R^2 is nonsignificant ($p = 0.27$, one-tailed). Turning to the model's covariates, trust has a positive but nonsignificant relationship to the outcome ($b = 0.17$, $p = 0.22$), as does the indicator for male ($b = 0.26$, $p = 0.20$). The indicator for whether one works at the school level is negative and also nonsignificant ($b = -0.55$, $p = 0.96$). Therefore, controlling for gender and the level at which they work, a leader's level of trust in colleagues does not significantly predict the number of alters to whom that leader sent a collaboration tie in year 1. The interpretation of this multiple regression is the same as if you were running a multiple regression with non-network data. The coefficients are generated by standard ordinary least squares linear modeling techniques and are based on comparing scores on independent and dependent attributes of individual actors. What differs here is the recognition that the actors, in this case school leaders, are not independent, so that estimation of standard errors by simulation rather than by standard formula is necessary (Hanneman & Riddle, 2005), a point emphasized in Chapter 8.

The *t*-test, ANOVA, and regression approaches discussed in this section are all calculated at the micro or individual-actor level. Variables that are analyzed as independent and dependent may be either relational or nonrelational. That is, we could be interested in predicting and testing hypotheses about actors’ nonrelational attributes (e.g., trust) using a mix of relational (e.g., ego network size) and nonrelational (e.g., gender) attributes. We could also be interested in predicting a relational attribute of actors (e.g., centrality) using a mix of relational and nonrelational independent variables.

Figure 9.4 UCINET Output of Regression Analysis Predicting the Number of Collaboration Ties (Out-Degree) Year 1. These results show that controlling for gender and the level at which a leader work (district-level is the intercept), a leader's level of trust in his or her colleagues does not significantly predict the number of alters to whom the leader sent a collaboration tie in year 1.

```
REGRESSION
----------------------------------------------------------------------------

Dependent variable: (column 6)
Independent variables: (columns 2 3 4)
# of permutations: 1000
Random seed: 982

NOTE: All probabilities based on randomization tests.


MODEL FIT

        Adjusted                One-Tailed
R-square R-square    F Value   Probability
-----------------    -------   -----------
   0.092    0.002      1.323         0.271


REGRESSION COEFFICIENTS

               Un-st'dized    St'dized   Proportion  Proportion  Proportion
 Independent   Coefficient    Coefficient As Large    As Small    As Extreme
 ----------- -------------- ----------- ----------- ----------- -----------
   Intercept       1.095301    0.000000       1.000       0.000       1.000
       TRUST       0.173922    0.126004       0.221       0.779       0.446
        MALE       0.263656    0.134663       0.196       0.804       0.376
      SCHOOL      -0.554212   -0.281197       0.959       0.041       0.085
```

The examples using the School Leaders data illustrate how relational and nonrelational attributes of actors can be analyzed using common statistical techniques. The key thing to remember, though, is that the observations are not independent (since all the actors are members of the same network). Because of this, direct estimation of the sampling distributions and resulting inferential statistics is needed—standard, basic statistical software (e.g., SPSS, Stata, or SAS) will not give correct estimates.

Modeling Groups of Actors

The final set of models and methods can be used to test whether groups of actors cluster together in ways that are consistent with one's expectations or can be understood given some other observed information. For example, to answer whether you can distinguish different groups of school leaders based on how often they collaborate, and if so, whether these groupings are related to the level at which they work (school versus district) requires you to evaluate whether the groupings are stochastically equivalent (i.e., they have the same probability distribution of ties to other actors in their grouping). The focus of these methods and models, therefore, is on groups of actors. Whereas the previous two sections focused on predicting dyadic ties between a network's actors or individual outcomes in complete networks, here the emphasis shifts to groups of actors and whether the actors in these groups share one or more attributes in common. Therefore, these models can be used to test hypotheses about the relations between and/or within groups of actors.

Among the several different ways in which groups of actors are modeled in a probabilistic manner, three are especially noteworthy and are labeled by the software in which they are implemented (van Duijn & Huisman, 2011). First among these is KliqueFinder (Frank, 1995, 1996, 2009), which focuses on identifying groups of actors within a network that have a higher probability of interacting with each other than with members of other groups. KliqueFinder identifies groups based on a simple idea: cohesion. Actors identified as members of the same group should have a higher probability to interact with each other than with actors from other groups. It is based on a model that is very similar to a $p1$ but goes further by incorporating a categorical actor covariate (group) in the form of a similarity index. No information other than the observed network is necessary, and the algorithm makes no distinction between the directions of ties.

Kliquefinder works in the following manner. Using iterative partitioning, actors are preassigned to a group, specifically a clique of three actors, and this assignment iteratively changes until the objective function does not improve. The objective function is defined to maximize the likelihood of within-group ties. You can also decide to preassign actors to different groups. Kliquefinder generates a graphical representation of the groups and the actors within them. This representation is obtained by using multidimensional scaling (MDS) on distances between actors and groups defined by the density of ties between them. KliqueFinder has been used widely used in educational research to identify groups of students, including by McFarland (2001), Crosnoe, Riegle-Crumb, Frank, Field, and Muller (2008), and Plank (2000). A useful and appropriate application of the KliqueFinder algorithm to the School Leaders data would allow you to ask whether groups of school leaders can be distinguished based on how frequently they collaborate and if these groups are related to whether one works at the school or district level, or some other nonrelational dimension.

A second model, ULTRAS, uses a different process to address questions of a similar nature (Schweinberger & Snijders, 2003). Like KliqueFinder, ULTRAS finds groupings of actors that have a higher density, but these groupings are nested, with higher-density groups nested in groups with lower density. These nested groupings are represented in a visual that resembles a geographic map with contour lines. The lines, of course, demarcate different groupings. The ULTRAS model can handle valued (continuous) or nonvalued (dichotomous) tie variables, but the data matrix must be symmetric.

ULTRAS is a model based on two assumptions. First, the observed network is generated by hierarchically nested latent transitive structures, expressed by ultrametrics. Second, the expected tie strength decreases with ultrametric distance. The distance between actors is measured by an ultrametric (latent) space defined for each pair relative to their distance to third actors, which implies a transitive structure. Ultrametric structures can be regarded as a mathematical expression of Mazur's proposition (1971, p. 308) that “friends are likely to agree and unlikely to disagree; close friends are very likely to agree and very unlikely to disagree.” A larger distance between actors translates into a lower probability of a tie. Like KliqueFinder, the computation behind this process is very complex; fortunately, the ULTRAS software that makes use of this model does all of the computation. Schweinberger and Snijders (2003) demonstrate how this algorithm works and the image it generates by performing a secondary analysis of the data gathered by Bernard, Killworth, and Sailer (1980), who studied the interactions among 58 students living in a fraternity at a West Virginia college for at least 3 months.

A third model, Latentnet, is the most flexible in that it can include actor and dyadic covariates when trying to find the optimal number of positions within a network. It can also handle directed data, a drawback of the previous two models. Using the School Leaders data, this model can be used to assess, for example, whether the level at which a school leader works (school versus district) influences the groups within which the leader most frequently collaborates. The covariate indicating level at which one works can be used either to help generate the clusters or after the fact to assess the clustering.

A didactic example of this model's application is provided by Handcock, Raftery, and Tantrum (2007), who demonstrate its utility on friendship data from 69 students in one school (Grades 7–12). Using these data, which were extracted from the National Longitudinal Study of Adolescent Health (Harris et al., 2009), Handcock, Raftery, and Tantrum apply the Latentnet model to demonstrate a strong tendency for students to form ties with others in their own grade, as the clusters line up well but not perfectly with the grades. In addition, they also reveal a subtle tendency for the within-grade cohesion to weaken as students move up from one grade to another in the school, from the tightly linked seventh graders to the more loosely tied students in the top three grades who associate more easily with students in grades other than their own. Like the other two models just presented, it also estimates the cluster to which each actor belongs. These estimates are probabilistic and estimate the likelihood of each student belonging to each cluster. Finally, Latentnet is an *R* package, which is an extensive suite of open-sourced statistical software.

KliqueFinder, ULTRAS, and Latentnet are just three of the ways in which groups of actors are modeled. This section focused on a few basic ways in which groups of actors can be identified and compared against a simulated distribution, which enables you to assess the likelihood (the “fit”) of an observed group structure. Because these three models define differently what it means to be stochastically equivalent (actors that have the same probability distribution of ties to the other actors), they will more than likely return varied solutions. Therefore, you must choose a model that best matches what you think constitutes group membership given the empirical data with which you are working. A more thorough treatment of these different models and the choices you must make as an analyst can be found in Wasserman and Faust (1994) and Scott (2000).

The next two sections on (1) models for two-mode networks and (2) other advanced models provide a brief and nontechnical introduction to several areas in which the statistical approach to social network analysis has made rapid advances in ways that will surely influence network studies in and around education. Given the complexity of these topics, the aim of this introduction is to communicate the importance of these advances in relation to the type of research questions they can address.

Modeling Two-Mode Networks

Chapter 3 introduced a special kind of matrix, one in which the rows and columns represented different entities. These networks are referred to as having two modes and typically reflect the relationship between actors and events, that is, whether there is a relationship between two actors that share the same affiliation or whether there is a relationship between two events because they share the same actors. For example, two students are affiliated because they are in the same class, or two classes are affiliated because they share similar students. Whether the analytical focus is ultimately on actors or events, these types of data are first organized in a matrix in which the rows (typically) represent actors and the columns represent events.

The most common way to analyze these data is to convert them so the two-mode data are collapsed into a single-mode adjacency matrix. This means that if Student A and Student B are in the same math class, then Student A and B are adjacent (where students [actors] are one mode and math classes [events] are another mode). This is what is typically done with most affiliation networks (e.g., see Carolan, 2008a), a transformation that is easily done through general social network software applications. But recent advances have sought to preserve the duality of actors and events and directly analyze the two-mode network (in graph theoretic terms, these are referred to as “bipartite graphs”) without converting it to a single-mode adjacency matrix. The key to understanding the concept of duality is that links between units of the same mode (either actors or events) must pass through units in the other mode (Breiger, 1974). The consequence of this duality is that important structural features of the relations between the elements of one mode can only be completely understood if you simultaneously consider the way in which these same elements form relations among the elements of the other mode (Field, Frank, Schiller, Riegle-Crumb, & Muller, 2006). For example, the relationships between courses (events) that students (actors) take can only be fully understood in terms of the specific students that take those courses. Therefore, when collapsing an affiliation network into an adjacency matrix, you “lose” information that is essential to understanding how people intersect with various social collectivities.

A number of techniques have been developed that retain this duality and do not require the data to be converted into a single-mode graph prior to analysis. These techniques, however, are not “statistical” in the sense that they generally do not make use of simulations for the purposes of hypothesis testing and inferences. Rather, they reflect advances in mathematical approach that take into account that the observed network is not just two mode (bipartite) by happenstance but is so by design (Borgatti & Halgin, 2011). In practice, this requires that measures designed for adjacency matrices be adjusted by applying a *post hoc* normalization. This is the strategy used when applying typical centrality measures to affiliation data (e.g., degree, closeness, betweenness, and eigenvector).

In other cases, however, a totally different approach must be constructed. For example, when trying to find cohesive subgroups in a two-mode network, the bi-clique measure has been developed, which is essentially what a clique is to an ordinary single-mode network but adapted for a two-mode network (Borgatti & Everett, 1997). When conducting a positional analysis on a two-mode network, algorithms have been developed to account for the bipartite nature of the graph (reviewed in Borgatti & Everett, 1992b) and identify similar actors or events based on either regular or structural equivalence. Finally, Field and colleagues (2006) have adapted Frank's Kliquefinder routine (1995) to identify positions of similar actors and events. This algorithm also incorporates Monte Carlo simulations—a popular class of computational algorithms that rely on repeated random sampling to compute their results—to evaluate the internal validity of identified positions and then apply this same algorithm to simulated data with optimal positions to evaluate to what extent and under what circumstances the algorithm recovers the optimal positions. Thus, this last technique combines a number of advances in the mathematical and statistical modeling of two-mode social network data.

Advanced Models

Most of the recent advances in the statistical analysis of network data address hypothesis-driven research questions about how networks emerge and change and what effects they have on varied outcomes. It should be evident from this chapter that things get statistically complex very quickly, as the nonindependence of observation requires that probability distributions be constructed through simulations (e.g., ERGMs). While complex, these simulations are increasingly necessary, as they permit inferences to be drawn and uncertainty to be modeled. In addition to the methods and measures described thus far, there are three other areas in which computational and statistical advances have allowed researchers to employ the network perspective in ways that were not possible a short time ago (Daly, 2010).

Agent-Based Models (ABMs)

The first of these recent advances relies on simulation to explore systemic implications of rules for behavior and interaction assigned to a set of actors (Frank, Kim, & Belman, 2010). Based on utility functions, these computational agent-based models (ABMs) are becoming important tools for understanding human–environment interactions. Maroulis and colleagues (2010) argue that this modeling process can help educational researchers show how individual actions aggregate into macro-level outcomes, an approach that can help integrate insights from different types of research and better inform educational policy. Such an approach, they continue, can establish not only what works but also how and why it works.

For example, why would one teacher help another teacher? To address this question, a computational agent-based model would work like this. You preselect rules for behavior and interaction among a set of nodes, which, in this instance, are the teachers. These rules (e.g., who interacts with whom) are then modified to examine how patterns of behavior and interactions emerge over multiple iterations. The results of these iterations are then compared against or used to predict what goes on in a social system (Daly, 2010). For example, consider school choice reform. Empirical research on programs that give households more choice is inconclusive, with methodological concerns arising for both observational and experimental studies (Goldhaber & Eide, 2003). To address these concerns, economists have used agent-based simulations to identify features likely to minimize “cream skimming” of top students by private schools in systems in which government-issued vouchers are used to pay for private schooling (Epple & Romano, 2008). In addition, researchers such as Frank, Kim, and Belman (2010) have used agent-based models to examine how teachers respond to external demands for change within the social organization of their schools.

Multilevel Models

Does collaboration between teachers vary by grade level? Such questions rely on relational and attribute data that are nested—pairs of teachers within grade levels within schools—that, by definition, violate the assumption of nonindependence. Therefore, they require a multilevel modeling strategy, which permits network analysts to examine relational data through a regression framework that properly adjusts for the dependency inherent to social network data as well as the "nestedness" of the dyads in higher levels (grade levels and schools). Therefore, these models also allow researchers to simultaneously examine micro and macro perspectives and decompose the variance attributed to these different levels.

These types of models were first developed and employed (Bryk & Raudenbush, 1992) to address the disparity between what theory suggests and what empirical reality shows (McFarland, Diehl, & Rawlings, 2011). Standard linear models, as noted earlier, assume independence among actors, but this clearly is not the case when, for example, studying students who are nested in classrooms, which, in turn, are nested in schools. By allowing variance to be measured at multiple levels, multilevel modeling presents a method more in line with our understanding of how schools are actually structured. Similarly, these models can also take into account the dependencies among related actors. Such models are pervasive in egocentric network studies. In these types of studies, each ego contributes to multiple rows in the data set (dyadic data in which there is one row for each ego/pair). Therefore, multilevel models are necessary to cope with this nonindependence. In addition, these models can also incorporate more than two levels; for example, students who are nested in dyads and dyads that, in turn, are nested in classrooms. This is an example of a three-level model, which can also employ both fixed and random effects at each level. The important conceptual point to keep in mind is that these models properly adjust for the inherent dependencies in the data.

An exemplary education-related example of social network analysis that employs multilevel modeling is Penuel, Frank, Sun, Kim, and Singleton's (2013) study on the normative influences of reading policy on teaching practice in the early 2000s. Drawing on relational and attribute data collected over four time points from 131 teachers with direct responsibilities for reading instruction in 11 schools, this study reached two conclusions. First, teachers' practices did not conform exclusively to the new normative regime but rather depended on exposure to external professional development in reading instruction and on local norms of practice in their schools and collegial subgroups. Second, over time, subgroups' practices diverged with respect to teachers' implementation of skills-based reading instructional practices. To reach these conclusions, they employ a multilevel framework, with teachers (level 1) nested within subgroups (level 2), which, in turn, are nested within schools (level 3). To identify subgroups, they rely on the Kliquefinder algorithm discussed earlier in this chapter. Not only does this multilevel analytic strategy account for the dependencies among observations, it also enables them to decompose the variance in teachers' practices among schools, subgroups, and individuals. This last point ultimately permits them to examine how much change in the outcome is attributed to each analytic level and examine this change over time. This is an excellent example of using social network data in a multilevel framework.

Actor-Oriented Models

A third developing area of network research that makes use of advanced statistical tools focuses on the examination of network change over time. These models show how networks evolve and whether there are actor-level characteristics associated with this evolution. These models are increasingly important, as it is obvious—or, at least it should be—that social networks are dynamic by nature. Ties are established between actors, they may flourish and perhaps evolve into close relationships, and they can also dissolve quietly or suddenly (Snijders, van de Bunt, & Steglich, 2010). These relational changes may be considered the result of the structural positions of the actors within the network—for example, when friends of friends become friends—characteristics of the actors (actor covariates such as gender), characteristics of pairs of actors (dyadic covariates such as whether there is a tie), and residual random influences representing unexplained influences.

Snijders, van de Bunt, and Steglich (2010) give a tutorial introduction to these types of models. They refer to them as *stochastic actor-based models for network dynamics*, a family of models that have the purpose of representing network dynamics on the basis of observed longitudinal data, and evaluate these according to the paradigm of statistical inference. Whereas ABMs rely exclusively on simulation, actor-oriented models make use of observed longitudinal data collected at least two points in time on at least 20 actors with (ideally) complete network information. Current software programs (e.g., SIENA, STATNET, and PNET) require you to specify the parameters that they think govern how the network evolves from one time point to another and then generate a simulation of networks to determine whether imposing those rules will generate networks (dis)similar to the observed network at a later point in time (Valente, 2010). The challenge in creating these models is for you to specify the network's objective functions—the network's assumed behavioral and structural tendencies. The theoretical model that is being tested should guide these objective functions. These stochastic actor-based models allow users to test hypotheses about these tendencies toward reciprocity, homophily, and transitivity and to estimate parameters expressing their strengths while controlling for other tendencies (“confounders”).

These models are relatively new and are more complicated than many other statistical models. But there are several interesting examples that show the exciting potential of these models, two of which are noteworthy. First, using four waves of friendship data collected from 26 students in one Dutch classroom, Snijders, van de Bunt, and Steglich (2010) show how this model can be used to examine whether network influence processes play a role in the spread of delinquency among students. They find that there is evidence for delinquency-based friendship selection, expressed most clearly by their delinquency similarity measure, and for influence from pupils on the delinquent behavior of their friends, expressed best by the average similarity measure. The delinquent behavior does not seem to be influenced by sex.

Second, Daly and Finnigan (2010) used this modeling approach to examine three different relational networks (advice, knowledge, and innovation) among 49 school leaders in a district that was in the middle of a reform effort. Their results show that the district's leaders became more connected to one other over time around reform-related knowledge and advice relations, but these connections were still infrequent. On the other hand, the ties associated with innovation became even less frequent over time. This finding suggests that while leaders increased their infrequent interactions around the reform, they interacted less frequently around innovative practices. Furthermore, those that engaged in frequent, stable interactions around knowledge, advice, and innovation demonstrated little change over time. Both the Snijders, van de Bunt, and Steglich (2010) and Daly and Finnigan (2010) examples show how these cutting-edge models can be used to test the hypothesized mechanisms involved in the formation, maintenance, and dissolution of social ties.

Summary

Chapter 8 distinguished between the mathematical and statistical approaches to the study of social networks. The central difference between these two is that the mathematical approach is useful in describing properties of both egocentric and complete networks, whereas the statistical approach emphasizes the generalizability and reproducibility of analytical results. In addition, recent advances in the statistical approach have been extended to include models that change over time and evaluate the likelihood of one network configuration versus another. In many ways, these “new” statistical approaches are similar to the use of statistical inference with the standard actor-by-attribute data through which quantitative data are typically collected and analyzed. Likewise, the statistical approach to the study of social network is primarily concerned with the probability of generating a similar result in a different sample drawn from the same population. The catch is that the statistical approaches discussed in this chapter calculate this probability primarily through simulations, which construct a probability distribution against which the observed result can be compared. This provides the basis for significance testing using relational data that do not conform to your typical inferential assumption of independence. The development of ERGMs was a major breakthrough and provides the basis for most statistical approaches of network data.

This chapter extended the foundation established in Chapter 8 by presenting several procedures that examine 1) ties in complete networks; 2) individual actor attributes, or 3) groups of actors. Therefore, these various techniques can be used to address questions that are especially relevant for educational researchers. Relational and attribute data from the School Leaders data set were used to illustrate how these techniques can be used to answer questions that attend to these three areas. Following an introduction of these varied techniques, the chapter focused on advances that have been made in the analysis of two-mode data, with an emphasis on retaining the duality of actors and events. Finally, this chapter discussed three areas that are especially promising, as they combine several recent computational and statistical advances. These areas were discussed in light of how they have been applied to the study of education-related phenomena. While a complete discussion of these areas is beyond the scope of this introductory text, it is important to have an awareness of the new and exciting ways in which the conceptual and analytical tools associated with the social network perspective are being applied to important topics in and around education.

I Chapter Follow-Up

Using one of the studies mentioned throughout this chapter, identify the statistical model that was employed and evaluate whether this was the appropriate modeling choice.

If this same study were to use standard statistical models that assume independence among observations, how would this influence the study's results?

Assume you had network data from an entire high school student body (*N* = 250) and were interested in predicting a student's number of friends from covariates such as sex, grade level, and academic performance. What model would be most appropriate to test these relationships?

Critical Questions to Ask about Studies that Use Network Data and Statistical Models

What is being modeled: ties in complete networks, individual actor attributes, or groups of actors?

What procedure was used and what, if any, covariates were included in the models? At what level were these covariates measured: continuous (ratio or interval) or categorical?

Did the modeling strategy measure a network's change over time?

How were simulations used to create a probability distribution against which observed network parameters were compared?

What does it mean when the result from a statistical model using network data is reported as being “statistically significant”? How does this relate to the implied null hypothesis?

Essential Reading

Daly, A.Finnigan, K. S. The ebb and flow of social network ties between district leaders under high-stakes accountability. *American Educational Research Journal,*(2010).48(1),39–79.

Pittinsky, M.Carolan, B. V. Behavioral vs. cognitive classroom friendship networks: Do teacher perceptions agree with student reports? *Social Psychology of Education,*(2008).11,133–147.

van Duijn, M. A. J., & **Huisman, M.**(2011). Statistical models for ties and actors. In Edited by: **J.Scott** & **P. J.Carrington** (Eds.), The Sage handbook of social network analysis (pp. 459–483). Los Angeles: Sage Publications Ltd.

http://dx.doi.org/10.4135/9781452270104.n9