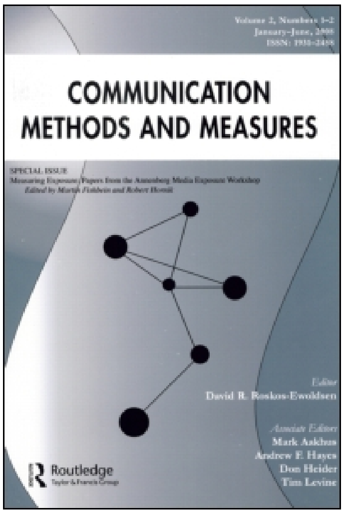

Communication Methods and Measures
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653633

Exponential Random Graph (p^*) Models as a Method for Social Network Analysis in Communication Research

Michelle Shumate[a]; Edward T. Palazzolo[b]
[a] University of Illinois, [b] Arizona State University,

Online publication date: 03 December 2010

To cite this Article Shumate, Michelle and Palazzolo, Edward 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

To link to this Article: DOI: 10.1080/19312458.2010.527869

URL: http://dx.doi.org/10.1080/19312458.2010.527869

Communication Methods and Measures, *4*(4), 341–371, 2010

ISSN: 1931-2458 print / 1931-2466 online
DOI: 10.1080/19312458.2010.527869

Exponential Random Graph (p^*) Models as a Method for Social Network Analysis in Communication Research

Michelle Shumate
University of Illinois

Edward T. Palazzolo
Arizona State University

Since the 1970s, communication researchers have utilized social network analysis to understand mass, health, organizational, and interpersonal communication. This article introduces communication researchers to a new class of social network analysis methods, exponential random graph (p^*) models. This new method represents the latest advancement in social network methodology and will enhance the trajectory of social network research in the communication discipline. The benefits of this class of models include allowing for the simultaneous estimation of attribute and structural parameters, accounting for the interdependent nature of network data, and retaining the complexity of network observations throughout the analysis. An example analysis using data from Shumate, Fulk, and Monge (2005) is provided to illustrate the potentials of exponential random graph modeling. Five different social network software programs capable of the analysis discussed in this article are introduced with regard to their respective benefits. Finally, a brief tutorial based on data from Palazzolo (2005) is given on how to conduct an ERGM analysis using the PNET software program.

Social networking technology, the emergence of Web 2.0, and the network society exemplify the ways networks have infiltrated public discourse. However, social networks are nothing new to communication researchers, who have examined social networks for more than 30 years. The purpose of this article is to introduce a

Correspondence should be addressed to Michelle Shumate, Ph.D., Communication, University of Illinois, 244 Lincoln Hall, 702 South Wright Street, Urbana, IL 61801. E-mail: shumate@illinois.edu

contemporary method for social network analysis that will allow communication researchers to test their network-based research hypotheses.

This article is organized as follows. First, we provide a brief overview of communication networks and social network analysis. Second, we review the social network analysis techniques utilized by communication researchers. We discuss the limitations of earlier social network methods utilized and why those previous methods are inadequate for contemporary research agendas. Following that review, we present an alternative methodology, a family of statistical models referred to as exponential random graph models (ERGMs), or p^* models (Wasserman & Pattison, 1999), as a means of overcoming prior challenges. Fourth, we discuss what is possible with ERGMs and demonstrate this method utilizing data from Shumate, Fulk, and Monge (2005). Fifth, we present five software packages for ERGM analyses. Sixth, we provide a brief tutorial on how to use PNET to conduct ERGM analyses with data from Palazzolo (2005).

COMMUNICATION NETWORKS AND SOCIAL NETWORK ANALYSIS

A social network is composed of a set of nodes and one or more relations among the nodes. Nodes are specific entities. Nodes in communication social network research have included individuals (e.g., Corman, 1990; MacDonald, 1976; Yuan & Gay, 2006), newsgroups (e.g., Choi & Danowski, 2002; Kang & Choi, 1999), organizations (e.g., Doerfel & Taylor, 2004; Shumate et al., 2005), journals (e.g., Reeves & Borgman, 1983; Rice, Borgman, & Reeves, 1988; So, 1988), knowledge concepts (e.g., Carley & Kaufer, 1993), groups (e.g., Barnett & Danowski, 1992; Doerfel & Barnett, 1999), television news sources (e.g., Reese, Grant, & Danielian, 1994), nation-states (e.g., Rogers & Antola, 1985), and characters on television shows (e.g., Fine, 1981; Livingstone, 1987). *Relations* are the links between the pairs of nodes. In communication research, relations have included friendship (e.g., Beinstein, 1977), perceived communication (e.g., Danowski, 1980), observed communication (e.g., Sykes, 1983), agreement (e.g., Stohl, 1993), membership (e.g., Barnett & Danowski, 1992), trust (e.g., Prell, 2003), cooperation (Doerfel & Taylor, 2004), telecommunications traffic (e.g., Monge & Matei, 2004), citations (e.g., Rice, Borgman, & Reeves, 1988), and hyperlinks (e.g., Shumate & Dewitt, 2008; Shumate & Lipp, 2008; Tateo, 2005).

While the nodes and relations of these social networks vary, the analysis techniques used to test hypotheses and answer research questions do not depend upon the type of nodes or relations examined. *Social network analysis* is the method used to measure and analyze social networks. Wasserman and Faust's (1994) book is the most commonly cited reference for both general and detailed information about social network analysis. They include considerable descriptions of all the relevant terms in social network analysis, as well as providing the mathematical formulas necessary to calculate various network statistics. Therefore, we refer to the reader to Wasserman and Faust's book as opposed to repeating such information here. Also, we refer the reader to Scott's (2000) handbook on social network analysis for, what some have said to be, a more accessible overview of social network analysis. Table 1 provides a list of common social network analysis terms and definitions used throughout this article.

TABLE 1
Key Terms and Definitions

Term	*Definition*
	Network Analysis Terms (Wasserman & Faust, 1994)
Affiliation network	A network in which common membership is the relation
Attribute	Property of a node which is independent of the connections to other nodes such as one's gender
Betweenness centrality	A score indicating the degree to which a node is on the shortest paths between actors
Blockmodel	A model resulting from the partition of actors into sets and analysis of the relations within and between sets of actors
Bonacich's centrality	A rank measure of prestige or power in the network determined by both the extensiveness of relations in the network and the relative prestige of actors to which a node has relations
Clique analysis	An analysis of subsets of nodes determined by their relations to one another and not to other sets of nodes
Closeness centrality	The total distance a node is from all other nodes in a network
Communication network	A network in which communication is the relation
Directed network	A network in which not all relations between nodes need be symmetric
Degree centrality	The number of relations which a node has; Divided into outdegree centrality, or the number of relations a node is sending to other nodes, and indegree centrality, or the number of relations a node is receiving from other nodes
Density	In a binary network, the ratio of present relations to the number of possible relations; In a valued network, the average strength of relations
MDS	Multi-dimensional scaling; An analysis technique which represents how proximate, similar, and dissimilar a set of nodes is to one another
MRQAP	Multiple Regression QAP; an extension of QAP to multiple network relations, similar to the multiple regression in that it allows estimation of multiple independent relations to a dependent relation.
Nodes	The representation of actors in a network
QAP	Quadratic Assignment Procedure; A correlation between network relations that takes into account the dependence of relations in the estimation of significance

(Continued)

TABLE 1
(continued)

Term	*Definition*
Relations	The lines or ties between actors
Semantic network	A network in which agreement is the relation
Structural holes	A gap between two groups in the network
ERGM Terms (Robins et al., 2009)	
Arc	A tie from one node to another in a directed network
Choice	A tie from one node to another
Edge	A tie from one node to another in a non-directed network
ERGM	Exponential random graph model; a family of statistical analyses of networks
Global clustering	A measure used in goodness-of-fit estimation which describes how well the model accounts for different types of *k*-triangles in the observed network
MCMC maximum likelihood estimation	A type of estimation in ERGMs that relies on computer simulations to produce a distribution of networks implied by the parameter estimates
Pseudolikelihood maximum likelihood estimation	A type of estimation in ERGMs in which the product of the log-odds ratio, or logit, of each probability for each tie being observed or not observed within a random network is computed
Reciprocity	The mutuality of relations
SD of degree distribution	A measure used in goodness-of-fit estimation which describes how well the model accounts for the standard deviation of the degree distribution in the observed network
Skew of degree distribution	A measure used in goodness-of-fit estimates which describes how well the model accounts for the skew of degree distribution in the observed network
Star	A structure in which a set of nodes have relations with a central node but not to one another; variations of star parameters include in-stars, out-stars, and mixed-stars, and the parameter is scalable into various *k*-stars, meaning the number of nodes can vary
Triangle	A network structure in which three nodes are connected to one another. Triangle structures include cyclical and transitive triads and the parameter is scalable into various *k*-triangles

In the past 30 years, communication researchers have used a variety of social network analysis techniques. These have included the classification of nodes into roles in the network (Rice & Love, 1987), measuring node connectedness (Feeley, 2000; Kang & Choi, 1999), visualization of networks (Reeves & Borgman, 1983), the examination of ego-network heterogeneity (Yum, 1982), the use of log-linear analysis to predict the frequency of interaction between each dyad (Sykes, 1983), the examination of changes in the number of transitive triads over time (Danowksi & Edison-Swift, 1985), quadratic assignment procedure

(Doerfel & Barnett, 1999; Stohl, 1993), structural holes (Doerfel & Taylor, 2004; Taylor & Doerfel, 2003), multiple regression quadratic assignment procedure (Shumate et al., 2005; Yuan & Gay, 2006), and ERGMs/p^* (Monge & Matei, 2004; Palazzolo, 2005; Shumate & Dewitt, 2008; Shumate & Lipp, 2008). The increasing complexity of research questions and hypotheses of interest for communication scholars has propelled these researchers to use and amend many social network analysis methods. However, the limitations of many social network analysis methods available and used in communication may have been a stumbling block to further theoretical and empirical research examining complex network phenomenon.

LIMITATIONS OF EARLIER METHODS

Prior to the introduction of ERGMs, methods were not available to deal appropriately with predictions of network ties by attributes (i.e., properties of nodes) and other relations without (a) reducing the data available or (b) ignoring issues that arise with interdependent observations. As such, many researchers over the past three decades either sacrificed the richness of social network data or ignored the interdependency of observations. For example, Beinstein (1977) examined the influence of friendship networks on the saliency and the influence of various information sources. In order to compare individuals, she computed the density of the friendship networks of forty respondents and classified the respondents as having either a high or low density in their friendship network. Then, the saliency and influence of several information sources were compared between groups. Thus, Beinstein took the rich information of the friendship networks of 40 individuals and classified these individuals as residing in one of two groups based upon the mean density. This is not to critique Beinstein's work, which in 1977 reflected the state of the social network analysis methods, but rather to highlight the way in which data reduction compromises the richness of network data. The richness of social network data is reduced when centrality, density, heterogeneity, or cluster analysis/multidimensional scaling are used in further analysis.

However, not all communication researchers reduced multiple observations down to a single or a few scores. Another limitation of prior communication research that uses social network methods is the use of parametric statistics. Most parametric statistics utilized rely on the foundational statistical assumption of independence of observations. However, communication researchers have largely ignored the ways in which centrality scores of individuals in a network may be interdependent when subjecting such scores to further analyses such as regression. For example, in an analysis of soap opera conversations, Fine (1981) treated each conversation dyad as an independent observation. Thus, instead of reducing data, Fine ignores the interdependency of the data coded, namely that some characters are likely to have multiple conversations and that conversations among some characters lead to conversations among others. Dyads within a network are not independent observations. Network structures are interdependent and treating dyads as independent observations violates the assumptions of many statistics used by communication researchers.

As another example, Feeley (2000) creates three logistic regressions to examine the role of centrality on turnover in workplace. However, each actor's centrality scores are related to the centrality scores of others in their workplace network. In considering a personal network of acquaintances, the interdependence becomes clear. Each communication interaction, and the amount of time available for it, is related to the communication with others within a network of acquaintances. For example, the amount of time one communicates with students during advising week is related to the amount of time one has to communicate with colleagues via e-mail. Therefore, the use of parametric statistics that assume independent observations is not appropriate for social network data.

However, these problems can be understood by limitations that were common in social network analysis methods before 2000. First, most methods allowed for a limited number of attributes to be examined simultaneously with relations. One might use a social analysis technique such as blockmodeling to examine the patterns of ties within and between sets of actors with particular attributes, but multiple attributes makes such a task unwieldy. Second, traditional network analysis has allowed only a few relations to be examined at a time. Until Multiple Regression Quadratic Assignment Procedure (MRQAP) was introduced, Quadratic Assignment Procedure (QAP) allowed only two relations to be examined simultaneously. Neither MRQAP nor QAP allowed the examination of relations simultaneously with attributes, without some data manipulation.

Finally, communication researchers often examined multiple networks of small sets of actors in order to gain enough power to find significance in traditional statistics. Methods that took advantage of the power of multiple interdependent observations did not exist. ERGMs present a solution to many of the problems that exist in social network analysis techniques used by communication researchers. In the next section, we explain ERGMs in detail including the types of variables that can be estimated with ERGMs for network analysis and the advantages of using ERGMs for research.

EXPONENTIAL RANDOM GRAPH (p^*) MODELS

Recently social network researchers in communication, computer science, physics, psychology, and sociology turned to a new class of social network statistics that addresses the problems raised in traditional social network analysis methods. ERGMs are well suited for the interdependencies present in network data (Pattison & Wasserman, 1999; Robins, Pattison, Kalish, & Lusher, 2007; Wasserman & Pattison, 1996). Although initially developed in the 1990s (Pattison & Wasserman, 1999; Wasserman & Pattison, 1996), ERGMs have gained much popularity among social network analysis researchers only within the last few years. Such a delay is not surprising in that ERGM analysis necessitated programming in advanced statistical packages, such as MATLAB, to perform the extensive calculations required at each step of analysis. Fortunately, these models now can be estimated using software easily accessible to communication researchers who may lack experience in advanced statistical programming. We describe several of these programs later in this article. Last, and perhaps most excitingly, "we can now reliably estimate parameters that describe established human social processes" (Robins & Morris, 2007b, p. 169). Further, we can use this type of social network analysis to make informed inferences about competing theoretical processes (Robins & Morris, 2007b), thereby further developing communication theories. With accessible software tools and an increase in research utilizing ERGM analyses appearing in journals, other researchers will discover and undertake the types of research that only now is possible.

In this section, we explain the principle concepts of ERGMs. First, we identify the purpose and usefulness of ERGMs. Second, we address the various types of interdependent variables whose effects can be estimated. After that, we present the two primary types of estimation used in ERGM analysis. Then we provide an explanation of the goodness-of-fit statistics for evaluating these two types of estimation. In the last part of this section, we discuss the advantages of ERGMs for communication research. For a more detailed explanation regarding the origins of ERGMs, we direct the reader to the two seminal pieces: Wasserman and Pattison (1996) and Pattison and Wasserman (1999). For more information about recent work on ERGMs, please see the special section of *Social Networks* (Robins & Morris, 2007a). Specifically, the article by Robins, Pattison, Kalish, and Lusher (2007) provides an excellent overview and outline of ERGM analysis and Robins, Snijders, Wang, Handcock, and Pattison (2007) provide an overview of the most recent developments in exponential random graph modeling.

Network Structures

ERGMs can be used to assess the statistical likelihood of specific network configurations, referred to as structures. While it may seem obvious that every network would contain *some* network structures, ERGM analysis examines the prevalence of the network structures *above* what would occur by chance alone. Imagine each possibility for each tie in a network to be similar to flipping a coin. If one flips a coin 80 times, one would expect that approximately 40 times the coin will land heads-up, so one would say a coin has a 50% chance of landing heads-up. Similarly, if a tie between two nodes is more likely than 50% of the time to appear in a network, then it appears more often than expected by chance. ERGMs are used to test if one or more structures are more prevalent in the network than would occur by chance alone.

Therefore, if the ties of a network structure appear more frequently than by chance, then the structure has a positive and significant propensity to be present within that network. If the tie appears less frequently than one would expect by chance, then the structure has a negative and significant propensity to exist in the network. While overly simplistic, this illustration highlights the basic principle of ERGMs, which compares the propensity of a network structure to the propensity that would occur by chance alone. Wasserman and Pattison (1996; Frank & Strauss, 1986) more precisely express the mathematical form of ERGMs as

$$P(\mathbf{X} = \mathbf{x}) = \frac{\exp\left\{\theta' z\,(x)\right\}}{\kappa\,(\theta)} \tag{1}$$

where *P(x)* indicates the probability of a given network, θ indicates a vector of model parameters, *z(x)* is a vector of network statistics, and κ is a normalizing function to ensure proper probability distribution across random networks (i.e., provides the comparison to a distribution of networks generated by random chance). The vector of model parameters, θ, contains the parameters of interest for estimation. The remainder of this section focuses on the model parameters, θ, available for estimation and discusses them in order of increasing complexity.

Types of Model Parameters for Estimation

The most basic parameter is the tie from one node to another. This parameter is commonly referred to as *arc*, *choice*, or *edge* and is an estimation of the probability of ties across the network as a whole. The estimation of an arc parameter is used in both directed and nondirected networks. In directed networks, the direction of the tie between nodes matters, and asymmetric connections are allowed (e.g., if Isabelle links to Jack, then Jack may or may not link to Isabelle), as is the case with e-mailing a prospective date. Nondirected networks have relations that are assumed to have no direction (e.g., a tie from Isabelle to Jack is equivalent to a tie from Jack to Isabelle), such as the relation for siblings. Estimating the *arc* parameter is useful for communication researchers in two different ways. First, in large communication networks such as the Internet, the arc parameter is an important control for the sparseness of the network, allowing for the successful estimation of other parameters that would be inhibited by the large number of absent ties. Second, arc can be a powerful parameter when combined with actor attribute information, described in the next subsection.

In directed networks, *reciprocity*, or mutuality, is the next parameter of interest. A positive, significant reciprocity parameter estimate indicates there are a number of ties that are mutual among actors. Reciprocity can be a measure of the social exchange (Monge & Contractor, 2003), or equal legitimacy (Rogers & Marres, 2000).

Star parameters are useful substitutes for network centralization measures, which have been more commonly reported in communication research. Star parameters can be directed or non-directed. Directed star-parameters include in-stars, out-stars, and mixed stars. In-stars are nodes in which k other nodes have connections to the central node but not to each other. Out-stars are structures in which a central node has connections to k other nodes, but those nodes do not have connections with one another. Mixed-stars are structures in which a central node has ties from k_1 other nodes and ties to k_2 other nodes but the noncentral nodes do not have ties with one another. Star parameters are scalable, meaning the value of k can take on multiple values, such that 2-in-stars, 4-in-stars, and 8-in-stars are all possible (see Figure 1). Further, in Markov Chain Monte Carlo (MCMC) maximum likelihood estimation models, k does not have to be specified but can be estimated through simulation.

Finally, *triangle* parameters are an important network measure in ERGMs as they are often important features of communication networks. Triangles are network structures in which three nodes are all connected to one another (Robins et al., 2007). Triangles are further classified as transitive triangles, cyclical triangles (Robins, Pattison, & Wasserman, 1999), k-triangles, directed k-triangles, directed k-2 paths (Snijders et al., 2006), and alternating k-triangles of various types (Robins, Pattison, & Wang, in press). Some of these various types of triangles are highlighted in Figure 1. One application, for example, of triangles for communication research might be the test of the cognitive dissonance hypothesis in friendships (Monge & Contractor, 2003).

ERGMs can be used to estimate other structural parameters beyond reciprocity, varieties of stars, and varieties triangles. However, each of those other structures is a combination of these structures. For example, the k-L-star is a combination of a k-in-star and a k-out-star where the star node both receives and sends multiple ties to unconnected other actors. Likewise, a *T1* triangle combines transitivity and reciprocity.

Attributes

While network structures are an important part of ERGM analysis, most communication researchers, given the types of hypotheses tested, are likely to be interested in studying attributes (i.e., properties of the node). Attribute variables can be binary (e.g., male or female), categorical (e.g., quarterback, tight end, running back, or wide receiver), or continuous (e.g., age). Network attributes can

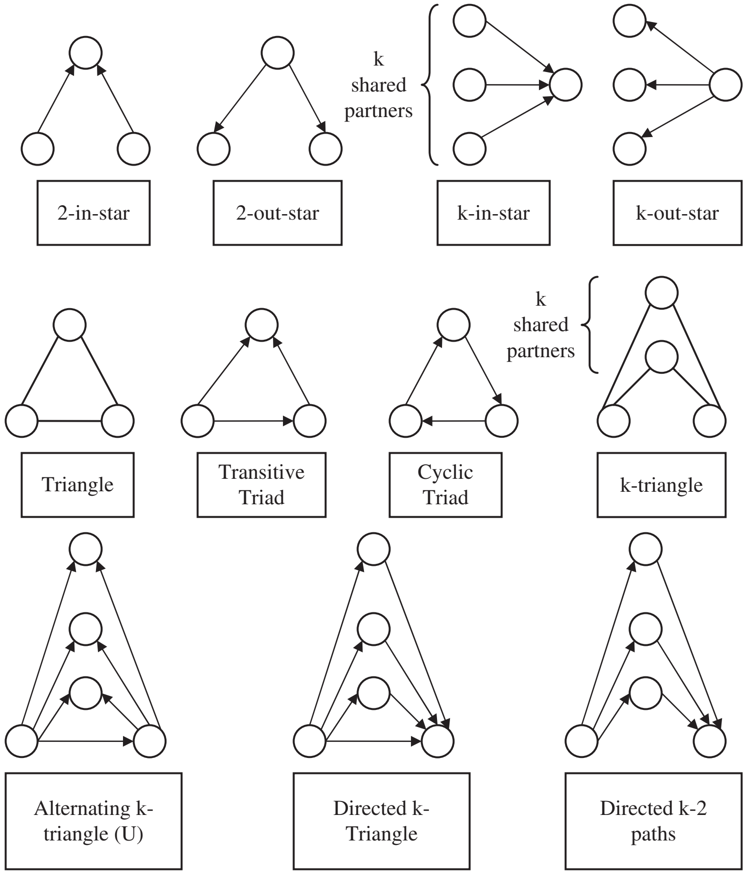

FIGURE 1 Stars and triangles able to be estimated with ERGMs.

be simultaneously included with network structures in exponential random graph models (Palazzolo, 2005; Shumate & Dewitt, 2008; Shumate & Lipp, 2008) to form social selection models (Robins, Elliott, & Pattison, 2001).

At the most basic level, researchers can test homophily hypotheses; that is, if actors who are similar (e.g., same gender) are more likely to have connections than by chance alone (Bryne, 1971; McPherson & Smith-Lovin, 1987; Monge & Contractor, 2003). Further, binary and categorical variables can be combined with any of the previously described network structures (Snijders et al., 2006). So, one might hypothesize that two nodes with the same attribute are more likely to have reciprocal ties (e.g., boys are more likely to mutually report being friends with each other in elementary school than boys and girls in the same class). Or one might hypothesize that nodes with a particular attribute are more likely to be the central node in out-stars (e.g., managers give instructions to subordinates, but subordinates do not give instructions to each other or to managers).

In addition, ERGMs allow researchers to specify relationships between nodes based on various contribution of an attribute, such as when the sum of the attribute values is highest, when the differences among the values is highest, and when the product of the values is highest. Using this feature, communication researchers could test hypotheses such as those with the least expertise would be more likely to have relationships with those with the greatest expertise (i.e., difference) or that two individuals with strong organizational identification would be more likely to have reciprocal communication ties (i.e., product).

Moreover, ERGMs can incorporate multiple attributes of various types simultaneously in the model estimation. This feature allows communication researchers an alternative to independently accounting for differences in people based on attributes. This option also allows for more parsimonious models to be developed which demonstrate when lower level structures (e.g., choice, reciprocity) account for the variance in the network that might otherwise be attributed to higher order structures (e.g., k-triangles).

Blockmodels

While attributes can be utilized to account for relationships, these measures do not allow for the specificity or configurations some researchers may require. Blockmodels are a way to specify higher order relationships among subnetworks within a larger network (e.g., exploring gender differences in the network) and allow researchers to examine tendencies that occur within and across groups. For researchers who wish to hypothesize relationships from a particular group in the network to a second group in the network, but not the reciprocated relationship, blockmodeling may be a good choice. Researchers for whom interactions among categorical variables play an important role, such as communication between Iraqi and Iranian sects, may wish to use blockmodeling. Further, the network parameters discussed earlier, such as reciprocity and triangles, can be estimated within blocks to build and estimate models that are more complex.

Additional relations

Finally, researchers may wish to estimate the probability of a relation using a second, third, or fourth network relation. For example, if a researcher hypothesized that people in minority communities are more likely to seek health information from family members than from nonfamily members, she would use the family ties relation to predict the second relation, the pattern of health information seeking within a minority community. ERGMs estimate the similarity of two relations and assesses if they are more or less similar than would be suggested by chance alone. The number of external relations that can be included in a model depends on the software used, but many software packages allow for multiple external relations to be estimated.

Overall, although each possible parameter to be estimated is presented on its own, ERGM analysis is elegant in that all parameters of interest can be *simultaneously analyzed*. Next, we present the ways to estimate the parameters in these models.

Model Estimation

ERGM parameter estimates can be calculated in two ways: (a) maximum pseudolikelihood estimation based upon the logit model and (b) MCMC maximum likelihood estimation based upon computer simulation. The distinction between maximum pseudolikelihood and MCMC maximum likelihood estimation may be familiar to communication researchers from logistic regression.

Maximum pseudolikelihood estimation

Maximum pseudolikelihood estimates are computed by fitting a logistic regression model, in this case computing the values of $z(x)$ and then comparing the fit to the values of the observed network (Pattison & Wasserman, 1999) rather than estimating the κ parameter. Thus, the form of the ERGM is transformed to:

$$PL(\theta) = \prod_{i \neq j} \prod_{m=1}^{r} P\left(X_{ijm} = 1 | X_{ijm}^{c}\right)^{X_{ijm}} P\left(X_{ijm} = 0 | X_{ijm}^{c}\right)^{1-X_{ijm}} \qquad (2)$$

In this estimation, the maximum pseudolikelihood estimate for a given network parameter is computed as a product of the log-odds ratio, or logit, of each probability for each tie being observed or not observed within a random network. Pseudolikelihood ratio statistics can then be used to compare the fitness of models. Like in logistic regression, pseudolikelihood estimation assumes independence of baseline logits (Robins et al., 1999; Strauss & Ikeda, 1990). Such an assumption enhances analytic simplicity and flexibility and produces consistent estimates (Robins et al., 1999). However, one must interpret the relationships between coefficient estimates and their associated standard errors with caution, especially when the probability of the baseline category, in this case zeros in the social network, is low or less frequent than other values in the network. In addition, small sample sizes are particularly problematic with this approach (Pattison & Wasserman, 1999). Lastly, pseudolikelihood estimation is at best approximate and, at times, may be seriously misleading; therefore, MCMC estimation techniques, when available, are preferred for estimation.

Markov chain monte carlo maximum likelihood estimation

In contrast, MCMC maximum likelihood estimates rely on computer simulations to produce a distribution of random networks to be compared with the observed network. The parameter estimates are revised through the simulation process until they stabilize (i.e., converge). MCMC maximum likelihood techniques do not assume the independence of logits, creating parameter estimates that can be interpreted in cases where the probability of the baseline category is low or in the cases of small sample sizes (Robins et al., 2009). However, MCMC maximum likelihood estimates can be obtained only when the model contains nondegenerate parameter estimates. Parameter estimates are considered degenerate or near degeneracy when the model implies only a few of the simulated networks, usually the full network or an empty network, have other than a very low probability of occurring (Robins et al., 2009). In degenerate cases, the parameter estimates will not converge, which suggests the model is poorly specified and different parameters are required.

Goodness-of-fit

Goodness-of-fit measures for ERGMs present another advantage over other social network analysis methods. Goodness-of-fit measures in ERGMs using MCMC maximum likelihood estimation compare the observed network with the distribution of simulated networks along many network properties including choice, reciprocity, and other network structures.

There are two types of goodness-of-fit estimates that result from this estimation process. First, goodness-of-fit measures in ERGMs using MCMC maximum likelihood estimation are indicators of the degree to which the estimated model represents a good explanation for the observed network (Goodreau, 2007; Wang, Robins, & Pattison, 2006). For the parameters estimated in the model, a good fit (i.e., convergence) is indicated by a convergence statistic of less than 0.1 (Snijders et al., 2006). In a model with a good fit, the number of particular network structures (e.g., reciprocal relations, 2-in-stars) in the observed network falls within the distribution of the number of network structures (e.g., reciprocal relations, 2-in-stars) in the simulated networks created using the reported parameter values.

Second, the goodness-of-fit measures in ERGMs using MCMC maximum likelihood estimation can be used to indicate whether the estimated model represents a plausible or good explanation for additional network dimensions not included in the estimated model (Goodreau, 2007; Wang et al., 2006). Network dimensions are properties of the network. Network dimensions, like variance in variables, allow researchers to determine how much of the network a particular model is explaining. Similar to accounting for variance in a regression analysis, some models explain more of the network than others do. A network is composed of various network structures and dimensions. For example, an observed network has a certain number of arcs, triangles, relations between actors with particular attributes, and a unique degree distribution. The greater the number of these structures and dimensions for which an estimated model can account, the closer it comes to explaining the entire network. Researchers should be guided by theory and research objectives in determining which network structures and dimensions they believe a particular model should explain (Goodreau, 2007).

For network structures and dimensions chosen not to be estimated in the model but which the model might explain, a good fit is indicated by a convergence statistic of less than 1.0, and a plausible fit is indicated by a convergence statistic between 1.0 and 2.0 (Robins, 2007). If the number of a particular network structure in an observed network is within the distribution of the number of those structures in the networks simulated by the estimated model, even though the simulated network distribution was not based upon an estimation of a parameter related to that structure, the model explains that structure or dimension of the network. For example, if a model including only an attribute, choice, arc, and reciprocity also explained the number of triangle structures in the observed network, the three estimated parameters explain another dimension. Thus, lower level parameters (i.e., choice, reciprocity, stars) may explain higher level network structures and dimensions of the model including the skew and standard deviation of indegree and outdegree centrality scores for nodes and global clustering in the network (Robins, 2007).

Developing such models will allow researchers to test theories and determine if there are parsimonious explanations for network dimensions. For example, Flanigan, Monge, and Fulk (2001), in their study of formative investments in interorganizational federations, hypothesized that formative investment would be positively related to CEO centrality in a federation. Controlling for several other variables and using ANCOVA, they found support for this hypothesis. In addition to the other advantages over ANCOVA mentioned above, ERGMs could have provided a measure of the degree to which formative investments not only correlated with indegree centrality, but the degree to which formative investments by CEOs accounted for the skew (i.e., extreme values) and standard deviation (i.e., variation) in the indegree centrality distribution. Further, ERGMs could have measured the degree to which formative investments explained dense clusters of CEOs who sought advice from each other.

Advantages of ERGMs

Exponential random graph models have several advantages over other social network analysis techniques. First, the parameter estimation method is suitable for the interdependency of network observations (Snijders, Pattison, Robins, & Handcock, 2006). Estimation techniques for ERGMs do not assume that observations are independent, like many of the parametric statistics that have been used in social network communication research.

Second, ERGMs can easily accommodate other relations, attributes, and structural estimates as predictors of a given network (Snijders et al., 2006). Each attribute, network structure, or other relation in the model is estimated given the other parameters in the model, thus giving researchers an estimate of the parameter's effect above and beyond the effects of the other parameters in the model. This may be of particular importance to communication researchers who study human communication networks. Reciprocity, for example, is a human communication behavior that appears to occur in many interactions, above the effect of individual attributes (Guerrero & Burgoon, 1996). Researchers, therefore, may wish to account for reciprocity when estimating interpersonal communication networks, even when their hypothesized models relate attributes to communication relations. Thus, structural predictors may act as control variables in the examination of human communication networks.

Although MRQAP and QAP can be adapted to account for attribute data, through the creation of matrices so that actors with the same attribute have ties to one another (e.g., see Doerfel & Barnett, 1999), ERGMs allow for greater flexibility in the use of attribute variables. For example, researchers can specify that actors with a particular attribute are more likely to be the senders of relations, the receiver of relations, or that actors with the greatest differences in valued attributes are likely to have relations.

Third, unlike currently implemented estimators in MRQAP (Dekker, Krackhardt, & Snijders, 2007), ERGMs better handle skewness in the distribution of network observations and collinearity between the multiple network observations. While the semi-partialing method addresses some of these concerns, UCINET 3.0 to version 6.1 did not implement this estimation method. Thus, ERGMs, especially using MCMC maximum likelihood estimators, provides better estimates than the most commonly implemented version of MRQAP.

ERGMs represent a promising class of models for communication researchers who are interested in social network analysis. The model allows researchers to predict a binary network from multiple attributes, network structures, and other relations. Further, researchers can assess the fit of their model and the degree to which it explains the network of interest. The example in the next section illustrates each aspect of ERGMs described above.

EXAMPLE USING PNET FOR ERGM (p^*) ANALYSIS

Shumate, Fulk, and Monge (2005) examined the HIV/AIDS international nonprofit, nongovernmental organization network over time. The purpose of this section is to illustrate the analysis potential of ERGMs for communication research by using the 2001 data from this previous research. While Shumate and colleagues (2005) utilized the limited MRQAP procedure, we were able to perform a more detailed analysis by developing four models including attributes, structural parameters, and other relations while not violating independence assumptions. We used PNET, a software package described in the following section, to estimate each model reported.

Tables 2 and 3 include the individual parameter estimates and goodness-of-fit indices for these models. While the values included in Table 2 would be useful in any evaluation of a model, the goodness-of-fit indices shown in Table 3 are just 6 of the 20 or more network parameters with which one can estimate. For instance, one could examine the distribution of transitive ties or the distribution of higher order *k*-triangles. As noted by Goodreau (2007), many observed

TABLE 2
Four ERGM Models with Maximum Likelihood Estimates

Parameters	*ML Estimate*	*Standard Error*	*Magnitude*	*Convergence Statistic*[a]
Model 1: Arc Baseline				
Arc	−4.42	0.13	34.00*	0.07
Model 2: Structural Parameters				
Arc	−8.77	1.02	8.60*	0.03
Reciprocity	6.26	0.54	11.59*	0.01
2-In-Star	0.04	0.15	0.27	0.01
k-In-Star (Lambda 2)	2.18	0.91	2.40*	0.02
Model 3: Attributes				
Arc	−5.30	0.29	18.28*	0.002
Funded by UNAIDS as receiver of ties	1.70	0.41	4.15*	0.04
Age as receiver of ties	0.00	0.03	0.00	0.03
Same region	0.31	0.40	0.78	0.06
Previous relationship in 1999	6.17	0.48	12.79*	0.003
Model 4: Combined Model				
Arc	−7.62	1.15	6.63*	0.03
Reciprocity	5.92	0.63	9.40*	0.03
2-In-Star	0.09	0.24	0.38	0.04
k-In-Star (Lambda 2)	0.97	1.10	0.88	0.03
Funded by UNAIDS as receiver of ties	0.05	0.54	0.09	0.01
Age as receiver of ties	0.02	0.03	0.67	0.02
Same region	0.13	0.37	0.35	0.10[b]
Previous relationship in 1999	5.50	0.58	9.48*	0.07

*Indicates a magnitude value of over 1.96, or a significant parameter. [a]Convergence statistics for each parameter estimated are included as part of the estimation routine for ERGMs. Convergence statistics are also computed as part of the goodness-of-fit routine. This second set of convergence statistics should be given more weight than the convergence statistics calculated during the estimation routine, since these values are computed based upon a larger number of iterations. [b]Indicates a convergence value has been rounded up, but converged at less than 0.1.

TABLE 3
Goodness-of-fit Indices for Four Models

	Arc	*Reciprocity*	*SD Indegree Distribution*	*Skew Indegree Distribution*	*SD Outdegree Distribution*	*Skew Outdegree Distribution*
Baseline	0.08	32.19	5.41	3.31	7.23	6.00
Structural	0.89	0.93	1.39	0.22	3.30	3.21
Attribute	0.02	16.41	1.17	0.37	5.36	4.50
Combined	0.08	0.07	0.27	0.12	2.63	2.80

A convergence statistic value of less than 0.1 for parameters estimated and 1 are indicators of a good fit. Convergence statistic values of between 1 and 2 for parameters not estimated in a model are indicators of plausible fit of the model within the distribution of the specified parameter.

network parameters may be compared with the estimated model. Researchers should be guided by theory and research objectives in determining which network parameters to compare. In this instance, we have chosen standard deviation of the nodes' indegree centrality, skew of the indegree distribution, the standard deviation of the nodes' outdegree centrality, and the skew of the outdegree distribution. The reason for this selection was both pragmatic, to aid in illustrating the uses of ERGMs, and theoretical, since Shumate and colleagues (2005) were primarily concerned with alliance partner choice and not larger clustering or path lengths in the network. The standard deviation of both indegree and outdegree distribution represent an estimate of the variance in the number of alliance partners chosen in the network (Goodreau, 2007). The skew of both the indegree and outdegree distribution represents an estimate of extremeness of the values in the degree distribution (Goodreau, 2007), often of interest in scale free networks or networks with a power law distribution. Below we present four models that build in complexity, to exemplify the possibilities of ERGM social network analysis.

Baseline Model

The first model, the baseline model, includes only a parameter estimate for arc, which indicates the level of choice (i.e., number of ties) compared with the number of possible ties, given the size of the network. The goodness-of-fit indices reported in Table 3 indicate the fit of the parameters to the specified model. A parameter estimate in the model can be assumed to have converged if the goodness-of-fit index is below 0.10. Here, the goodness-of-fit index is 0.08, showing a good fit of the arc parameter in the baseline model.

In Table 2, we report the maximum likelihood (ML) estimate, standard error, magnitude, and convergence statistic values. The ML estimate indicates the direction of the propensity for the parameter to exist in the network. The negative estimate of −4.42 for the arc parameter indicates ties are less likely to exist in this network than by chance unless, perhaps, the tie is included in a higher order structure such as a star or triangle, where the negative arc effect might be counterbalanced by positive higher order effects. A parameter estimated in the model is considered significant when the absolute value of the ML estimate is greater than twice the magnitude of the standard error. Here, the ML estimate of −4.42 divided by the standard error of 0.13 yields an absolute value of 34.00, which is clearly greater than 1.96, showing the arc parameter to be significant. In this simple model, arc is shown to be an important component in the network. Specifically, international nonprofit, nongovernmental organizations are less likely to form alliances with one another than would occur by chance alone. In other words, organizations form alliances with only a few of the potential other organizations in the network. Next, we build on this model by adding structural configurations.

Structural Model

In the second model, the following four parameters are included: arc, reciprocity, 2-in-star, and *k*-in-star. This model increases in complexity from the baseline model by adding three parameters to measure the structural configurations of the ties in the network. Unfortunately, this model reveals a poor fit to the observed network.

As can be seen from Table 3, the estimated goodness-of-fit values are not below 0.10 for any of the modeled parameters (e.g., arc $t = 0.89$ and reciprocity $t = 0.93$), which suggests the estimates have not converged properly and the model may be degenerate. The goodness-of-fit indices for the standard deviation of the outdegree distribution ($t = 3.30$) and the skew of the outdegree distribution ($t = 3.21$) are not plausibly modeled by these four structural parameters. Given the poor fit of the overall model, the parameter estimates reported in Table 2 for the structural model should not be interpreted despite their initial convergence statistic values (see Table 2 note). Thus, a better model that explains, at minimum, the parameters estimated is needed.

Attribute Model

In the third model, network attributes are tested for their ability to explain properties of the network. The attribute parameters are considered independent of the structural parameters. Specifically, this model includes a binary variable, receiving funds from UNAIDS or not; a continuous variable, age of the organization; a categorical variable, being in the same geographic region (e.g., North America, Africa, or Asia); and a second network relation, previous alliance relationship.

The model fit the observed network's arc parameters well ($t = 0.02$), indicated by a value of less than 0.10, because this parameter was estimated. In addition, indicated by a goodness-of-fit index of less than 1 for parameters not estimated by the model, this model explains the skew of the indegree distribution ($t = 0.37$). Further, its standard deviation of the indegree distribution ($t = 1.17$) is a plausible fit to the observed network, as indicated by a goodness-of-fit index between 1 and 2 for parameters not estimated by the model. However, the model fails to explain the standard deviation of the outdegree distribution and the skew of the outdegree distribution.

From Table 2, the arc remains a negative parameter (*ML estimate* = −5.30, *SE* = 0.29), whereas previous alliance relationship (*ML estimate* = 6.17, *SE* = 0.48) and receiving funding from UNAIDS (*ML estimate* = 1.70, *SE* = 0.41) are both positive and significant parameters. Thus, this model shows network ties are not random. Previous alliances and UNAIDS funding partially account for the alliances reported. However, the model still can be better fit with the appropriate parameters.

Combined Model

Finally, a combined model, including both structural variables and attribute variables, was estimated. The parameters included in the model converged and the estimated model fits these parameters well, as indicated by a convergence statistic value of less than 0.1. In Table 3, we report both arc and reciprocity goodness-of-fit, as parameters included in the combined model. In addition, we have included the goodness-of-fit for four parameters not estimated in the final model. The combined model provides a plausible explanation, indicated by a convergence statistic value of less than 1.96, for the standard deviation and skew of the indegree distribution, or the relative popularity of HIV/AIDS international nonprofit, nongovernmental organizations as alliance partners. However, the combined model fails to provide a plausible explanation of the standard deviation or skew of the outdegree distribution, or the number of alliance partners sought by HIV/AIDS international nonprofit, nongovernmental organizations. Future work should introduce additional or alternative explanations to account for these dimensions of the network.

In the final model and as reported in Table 2, only arc, reciprocity, and previous alliance relationship remained significant. The maximum likelihood estimate for arc remained negative and significant (*ML estimate* = −7.62, *SE* = 1.15). The reciprocity parameter was positive and significant (*ML estimate* = 5.92, *SE* = 0.63). The previous alliance relationship also was positive and significant (*ML estimate* = 5.50, *SE* = 0.58). None of the other parameters was significant.

Shumate, Fulk, and Monge (2005), in their original study of HIV/AIDS international nonprofit, nongovernmental organizations using MRQAP, found that common cohort, geographic proximity, and past cooperative relationships were related significantly and positively to organizational alliances. This study, using the same data, yielded some similar and some different results. Past cooperative relationships remain significantly and positively related to current alliances. However, geographic proximity, or being located in the same geographic region, did not remain significantly related to alliances among organizations. We did not examine common cohort here, but age was not related to preference for alliances in the network. Finally, we added structural indicators of the network, a feature unavailable to be examined in MRQAP, and found reciprocity was more prevalent than would occur by chance alone. What accounts for these different findings? First, unlike MRQAP, ERGM simultaneously estimates each parameter, so each parameter's influence is in addition to that of the other parameters. Second, we were able to include an arc parameter, which controls for the sparseness of the network. The high number of zeros in the alliance network may have inflated the significance of some of the parameters in the prior study. Finally, ERGM allowed us to include structural parameters in the final model. These parameters effects were unaccounted for in the 2001 MRQAP analysis. In sum, ERGM analyses allowed us to include controls, examine the additive effect of each parameter, include structural effects in the final model, and identify previously unaccounted for effects.

LIMITATIONS OF ERGMS

Although ERGMs (p^*) have a number of advantages over previous social network analysis methods, the class of models has several limitations. First, ERGMs using MCMC estimation require a fair amount of computing power and time to run an MCMC simulation, especially as the size of the network increases. This limitation will decline over time as computing power increases and cost decreases. Second, the process of identifying a model that will converge or produce an excellent goodness-of-fit can be challenging. Choosing parameters to estimate can take some time, as there are many possibilities. Further, as many theories commonly used by communication researchers are not expressed in network terminology, linking the theoretical concepts and mechanisms of interest with network parameters will require further work (see Monge & Contractor, 2003, for more on this line of needed scholarly work). Third, once the best model is identified and parameter estimates are properly calculated, the interpretation of the results may be challenging. Parameter estimates are calculated given the effects of the other parameters in the model. As such, the estimates of any given parameter should be interpreted as controlling for or in addition to the other effects in any given model. Fourth, at the present time, only binary networks can be estimated. Researchers with valued networks will have to choose between new MRQAP techniques (Dekker et al., 2007) and dichotomizing valued network data in order to use ERGMs.

ERGMs have been suitably developed to the point where they are useful for communication researchers. They are a new class of social network models; as such, statistics and techniques are currently being developed and are continuously evolving. Ongoing work is examining the parameterization, or further parameters which may be estimated, and extending ERGMs to valued networks. ERGMs show considerable promise as an analytical method with advantages that outweigh, in most cases, the disadvantages of earlier developed social network analysis. In addition, many software packages are now available which communication researchers may use to estimate ERGMs.

SOCIAL NETWORK SOFTWARE

There are at least five software packages that will allow for the estimation of ERGM*s*: (a) PSPAR, (b) MultiNet, (c) Statnet, (d) SIENA in Stocnet, and (e) PNET. Each of these packages can currently handle networks of at least 500 nodes, and some can handle up to 5,000 nodes. Depending upon one's computer processor and patience, even larger datasets may be manageable. The first, PSPAR, is a DOS-based program, created by Seary (1999). A review of this package and instructions for examples are available in Monge and Contractor's (2003) *Theories of Communication Networks*. PSPAR program uses maximum pseudolikelihood estimation and is ideal for working with sparse networks with large numbers of nodes. PSPAR allows one external relation and one categorical attribute to be included in a model at a time and does not estimate any *k*-parameters.

The second, MultiNet, is a Windows-based shareware program created by Richards and Seary (2007). Like PSPAR, MultiNet also uses maximum pseudolikelihood estimation. MultiNet has a graphical interface that may be more appealing to users than PSPAR's command-driven interface. Again, only one external relation and categorical attribute can be included in the model and *k*-parameters cannot be included.

The third program, Statnet, is a suite of software packages designed for social network analysis of random graph models (Handcock, Hunter, Butts, Goodreau, & Morris, 2003; Morris, Handcock, & Hunter, 2008). This program can make both maximum pseudolikelihood and approximate MCMC maximum likelihood estimates. It is designed to work within the R statistical package giving it the versatility to work on most computer platforms. Further, R's flexibility allows others to write their own packages for future inclusion in Statnet. Statnet allows for multiple types of attributes, multiple external relations, and can estimate *k*-parameters.

The fourth program, SIENA using Stocnet, was developed by Snijders and colleagues (2007). Like PNET, SIENA uses MCMC maximum likelihood estimation.

In addition, SIENA is capable of doing over time network analysis (see Snijders, 2001, for an overview) making it well suited for communication researchers interested in longitudinal projects. SIENA is now only supported in the R environment and does not support ERGM analysis within this environment. SIENA has incorporated *k*-stars and *k*-triangles and can include multiple external relations and attributes SIENA 3.17e allows for the estimation of ERGM models across multiple networks simultaneously, such as examining multiple project teams.

The fifth program, PNET, was developed by Wang, Robins, and Pattison (2006). It uses MCMC maximum likelihood estimation. PNet allows for the external network to be valued and can include multiple external networks simultaneously into the estimated model. In addition, the various forms of *k*-triangles have been incorporated into PNET. Further, a complimentary package, XPNET, allows for the estimation of ERGM models across multiple networks, examining both individual network effects and effects that are prevalent across all networks (see Su, 2008).

ILLUSTRATIVE ANALYSIS USING PNET

In this section, we use data from Palazzolo (2005) to provide step-by-step directions to illustrate how communication researchers who are familiar with UCINET or similar network analysis programs might utilize software for ERGM analysis. We use PNET (Wang et al., 2006) in this illustration and refer the reader to the software authors' Web site for up-to-date releases and manuals (http://www.sna.unimelb.edu.au/pnet/pnet.html).

In Figure 2, we present the data from Team A, Knowledge Area 1 from the Palazzolo (2005) article. The two matrices and the vector are each stored in individual text documents, which can be created in a variety of ways. Researchers can directly enter the matrix into a text editor such as Microsoft Notepad. In addition, researchers can enter data into UCINET, Microsoft Excel, or other network analysis programs. These programs allow the researcher to export the network files and attribute files as text documents. The interested reader is encouraged to enter the data from Figure 2 into three text documents and follow along to replicate the illustration.

A screen shot of the PNET program is displayed in Figure 3. There are four tabs—Simulation, Estimation, Goodness-of-fit, and Bayes Goodness-of-fit. We have selected the *Estimation* tab to begin. PNET is similar to many network analysis programs in that output is saved as text documents external to the program. First, the researcher should enter a *Session Name*. In this case, we have named the session "PNET example". All documents created by this analysis will have this session name as part of its filename. Second, we have specified the *Session Folder*, or the folder in which all the output files will be saved. Next, the researcher must

Information Retrieval	Perceived Expertise	Self-report expertise
0 0 0 0 0 0 0 1 0 0 0 1 1 1 0 0 0	0 0 0 0 0 0 1 1 0 0 1 1 1 0 0 1 0	0
0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0
0 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0	0 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0	0
0 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0	0 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0	0
0 0 0 1 0 0 1 0 0 0 1 1 1 0 0 1 0	0 0 0 0 1 0 1 0 0 0 1 1 1 0 0 0 0	1
0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0	0
0 0 0 1 1 0 0 0 0 1 0 1 1 0 1 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0
0 0 0 0 0 0 0 0 0 0 1 0 1 1 0 1 1	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0	0
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0	0
0 0 1 0 0 0 1 0 0 0 0 0 1 0 1 0 0	0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0	0
0 0 0 0 1 0 1 1 0 1 0 0 1 0 0 1 1	0 0 0 0 0 0 1 0 0 0 1 0 1 0 0 1 0	1
0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0	0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0
0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0	0 0 0 0 0 0 1 0 0 0 1 0 1 0 0 0 0	0
0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0	0 0 0 0 0 0 0 1 0 0 0 1 1 0 1 0 0	1
1 1 1 0 1 1 0 0 1 0 1 0 1 1 1 0 0	0 0 1 0 0 1 1 0 0 1 1 1 0 1 1 1 0	1
0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0	0 0 0 0 0 0 1 1 0 0 1 0 0 1 0 1 0	0

FIGURE 2 Three text files of data from Palazzolo (2005) Team A, Knowledge Area 1. Each column represents a single text document which can be inputted into PNET. Text documents can be exported from UCINET and other network analysis programs.

specify the *Number of Actors* in the network. In this case, there are 17 actors in the network. The researcher must also indicate the text file containing the main network as the *Network File*; in this case, the information retrieval network pictured in the left column in Figure 2. Finally, the researcher must indicate if the network is *nondirected* or *directed*; the information retrieval network is directed.

Once the network to be estimated is identified in the program, the researcher must describe the model to be estimated. First, the researcher should specify the *Structural Parameters* she wishes to estimate. In the top callout box in Figure 3, a screen-shot shows the structural parameter interface that appears once the structural parameters button is clicked. In each of the callout boxes, PNET uses abbreviations for the various structures. Each of these is identified and visualized in the PNET manual, and we refer the reader to that manual for more detail. In this structural parameter selection box, we have selected arc, reciprocity, and in-2-star. Arc is the baseline parameter against which others are estimated and should always be included in the estimation. Palazzolo (2005) included the remaining two parameters as hypotheses one and two respectively. Click OK to return to the main PNET window.

Next, the researcher may identify the number of *Dyadic Attributes*, or external relations, to be estimated in the model by checking the dyadic attributes checkbox and clicking Select Parameters. The number of external relations practical to estimate depends on both the size of the network and the number of other parameters to be estimated in the model. In Palazzolo (2005), only one external relation

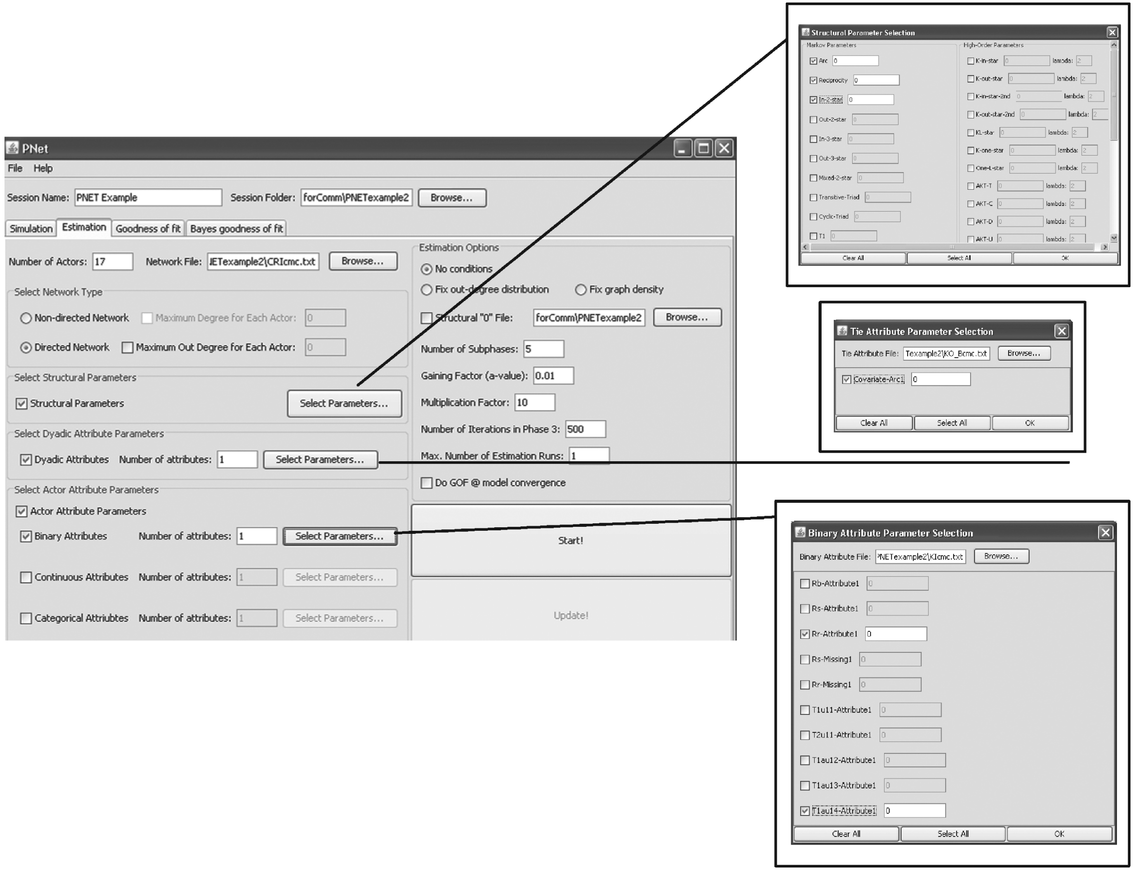

FIGURE 3 Screen shot of the PNET program (Wang, Robins & Pattison, 2006), showing the various elements of the estimation tab. Special thanks to Garry Robins for permission to use screen shots from the PNET program.

was examined, perceived expertise. Next, as displayed in the middle callout box, the researcher must both check the *covariate arc* to be estimated and indicate the location of the text file of the external relation data in the *tie attribute file* field. If more than one external attribute is requested to be estimated, multiple networks should be entered into the same text file. PNET will allow the user to estimate or not estimate one or more of the multiple networks by including a toggle box for each covariate arc. Click OK to return to the main PNET window.

The researcher may identify the number of binary, continuous, and categorical attributes by checking the actor attribute parameters checkbox and appropriate checkboxes for the types of attributes. Each type of attribute must be recorded in a separate text file, meaning that there should be one text file for binary attributes, one text file for continuous attributes, and one text file for categorical attributes. Again, the number of attributes practical to estimate depends both on the size of the network and the number of other parameters, both structural and external relations, that are also included in the model. The bottom callout box is displayed for binary attributes which can be opened by clicking the corresponding Select Parameters button. In Palazzolo (2005), one binary attribute was examined. Hypotheses four and five in the original article specified that actors with self-reported expertise would be more likely to be the receivers of information requests and were more likely to be the star of in-2-star structures. After specifying the binary attribute file, we have specified these same parameters in the screen shot indicated. Once again, click OK to return to the main PNET window.

Now that the parameters are indicated, the user clicks on "Start!" Once the estimation is complete, a dialogue notification box appears. The researcher should then locate a text file entitled "estimation-<Session Name>" in the specified session folder. The first output from this estimation, "estimation-PNET Example" is included as Figure 4. The researcher should examine the section of the text document entitled "Estimation results for network SUMMARY." In this section, the network parameter estimate, standard error, and t-value are indicated. In the top box in Figure 4, the absolute value of the t-values for the last three parameters are greater than 0.1. Thus, the model has not yet converged. In this case, the researcher should return to the PNET program and click "Update!" This will update the model to include the last estimate as the starting values. Then the researcher should again click "Start!" Achieving convergence is often time consuming, especially in larger networks. The PNET manual should be consulted for more advanced instruction if, after considerable time, convergence fails to be achieved. After updating the parameters, the illustrative model converged. In this model, the three out of the five hypothesized parameters were significant: reciprocity, in-2-star, and perceived expertise (see bottom box in Figure 4).

To examine how well a converged model explains additional elements of the data, a goodness-of-fit estimate should be created. The goodness-of-fit tab in PNET resembles the estimation tab. Again, the researcher will have to specify

```
estimation-PNET Example - Notepad
File Edit Format View Help
* Number of estimation runs = 1

*STOCHASTIC APPROXIMATION RUN 1

original statistics:61.000000 10.000000 142.000000 15.000000 22.000000 32.000000
starting parameters:-1.240984 0.000000 0.000000 0.000000 0.000000 0.000000
Phase1 Started with the following setup:
* a = 0.010000
* num of steps = 25
* num of iterations in each step = 502.773438
*************************************
mean statistics in phase1:62.280000 7.320000 106.160000 14.720000 24.080000 10.440000
END PHASE1 parameter:-1.240984 0.000000 0.000000 0.000000 0.000000 0.000000
Phase 2 started


Subphase 0 started with a valued 0.010000

Subphase 0 has gone up to 213 steps
Parameter after Subphase 0:-2.65410  0.91524  0.17788  3.99284 -1.12969  2.36124


Subphase 1 started with a valued 0.010000

Subphase 1 has gone up to 233 steps
Parameter after Subphase 1:-2.66047  0.95995  0.15320  4.33198 -1.24853  2.55235


Subphase 2 started with a valued 0.005000

Subphase 2 has gone up to 283 steps
Parameter after Subphase 2:-2.67749  0.97883  0.15303  4.49246 -1.31146  2.61248


Subphase 3 started with a valued 0.002500

Subphase 3 has gone up to 408 steps
Parameter after Subphase 3:-2.69093  0.99167  0.15296  4.63939 -1.36524  2.65897


Subphase 4 started with a valued 0.001250

Subphase 4 has gone up to 725 steps
Parameter after Subphase 4:-2.70693  1.00292  0.15281  4.74641 -1.40293  2.68946
END PHASE2 parameter:-2.706930 1.002921 0.152814 4.746411 -1.402926 2.689457
Phase3 Started with the following setup:
* num of steps = 500
* num of iterations in each step = 502.773438
*****************************************
mean statistics in phase3:61.406000 10.232000 144.260000 15.236000 22.900000 31.520000

*Estimation Result for Network SUMMARY (parameter, standard error, t-statistics)
NOTE: t-statistics = (observation - sample mean)/standard error
        Arc: -2.706930, 0.34875, -0.06543          *
        Reciprocity: 1.002921, 0.47947, -0.08088           *
        2-In-Star: 0.152814, 0.06129, -0.07263  *
        Rr for Attribute1:  4.746411, 3.18845, -0.15121
        Tlau14 for Attribute1: -1.402926, 0.95213, -0.17212
        Covariat Arc1: 2.689457, 0.43540, 0.16512


Estimate
0.121626
-0.08786
-0.01474
-0.09258
0.014807
-0.02003
```

```
*Estimation Result for Network SUMMARY (parameter, standard error, t-statistics)
NOTE: t-statistics = (observation - sample mean)/standard error
        Arc: -2.697703, 0.34923, -0.08420          *
        Reciprocity: 1.018126, 0.46792, -0.02394          *
        2-In-Star: 0.145029, 0.06482, -0.05319  *
        Rr for Attribute1:  5.198782, 2.94946, 0.03058
        Tlau14 for Attribute1: -1.595518, 0.90926, 0.05003
        Covariat Arc1: 2.752549, 0.43348, -0.02850          *
```

FIGURE 4 Top box includes the first set of estimates produced by PNET. The estimates have not yet converged. The bottom off-set box includes the converged set of estimates.

the model and the data files. Once all of the appropriate boxes for the parameters estimated in the model are checked, the researcher can click "Update!" and the estimates from the last estimation run will be entered as the values into the appropriate fields. Again, the user clicks "Start!" to begin the estimation and a dialogue box appears when the estimation is complete. A file with the

```
gof-PNET Example - Notepad
File Edit Format View Help

* GOODNESS OF FIT

* Parameter Values:
    # Arc:              -2.69770
    # Reciprocity:       1.01813
    # 2-In-Star:         0.14503
    # Rr for Attribute1:      5.19878
    # T1au14 for Attribute1:      -1.59552
    # Covariant Arc1:      2.75255
* There are 1000000 proposed digraphs.
* Statistic samples are picked up at 1 per 1000 digraphs.
* Accepted 292770 proposed digraphs.

observation, sample mean (standard error), t-statistic
t-statistics = (observation - sample mean)/standard deviation

# Arc: 61.0000 Mean= 60.9290 ( 6.5019 )  t = 0.0109
# Reciprocity: 10.0000 Mean= 9.8980 ( 2.8006 )  t = 0.0364
# 2-In-Star: 142.0000 Mean= 140.2710 ( 31.9329 )  t = 0.0541
# Rr for Attribute1: 15.0000 Mean= 14.9950 ( 1.4502 )  t = 0.0034
# T1au14 for Attribute1: 22.0000 Mean= 21.9470 ( 4.7897 )  t = 0.0111
# Covariant Arc1: 32.0000 Mean= 31.8950 ( 3.0271 )  t = 0.0347
# Std Dev In-degree dist: 2.7237 Mean= 2.6327 ( 0.4051 )  t = 0.2248
# Skew In-degree dist: 1.6942 Mean= 0.9736 ( 0.4066 )  t = 1.7721
# Std Dev Out-degree dist: 2.4748 Mean= 1.7140 ( 0.2830 )  t = 2.6885
# Skew Out-degree dist: 0.9190 Mean= 0.4277 ( 0.5036 )  t = 0.9756
# CorrCoef In-Out-degree dists: -0.1037 Mean= -0.0359 ( 0.2186 )  t = -0.3101
# Global Clustering Cto: 0.2863 Mean= 0.2793 ( 0.0529 )  t = 0.1321
# Global Clustering Cti: 0.2641 Mean= 0.2101 ( 0.0362 )  t = 1.4921
# Global Clustering Ctm: 0.3623 Mean= 0.2983 ( 0.0546 )  t = 1.1728
# Global Clustering Ccm: 0.2174 Mean= 0.1979 ( 0.0564 )  t = 0.3450
# Global Clustering AKC-T: 0.3604 Mean= 0.2814 ( 0.0445 )  t = 1.7771
# Global Clustering AKC-D: 0.2713 Mean= 0.2711 ( 0.0438 )  t = 0.0044
# Global Clustering AKC-U: 0.2450 Mean= 0.2130 ( 0.0337 )  t = 0.9490
# Global Clustering AKC-C: 0.2518 Mean= 0.2034 ( 0.0513 )  t = 0.9450
```

FIGURE 5 Goodness-of-fit estimates produced by PNET.

goodness-of-fit estimates entitled "GOF-<session name>" in the specified session folder should be examined for the results. The results of this illustrative test appear in Figure 5. In this file, the estimates used for the goodness-of-fit test are listed first. Then both estimated and additional parameters goodness-of-fit is indicated. The parameters are listed and then followed by the observed number, the mean count, and the standard error across the 1,000 simulated samples randomly selected from the 1,000,000 simulated networks, and the *t*-value comparing the observed count to the simulated mean count. In this illustration, all of the estimated parameters had *t*-values less than 0.1. Several additional parameters were explained well by the model, indicated by *t*-values less than 2.

CONCLUSION

The purpose of this article was to introduce communication researchers to ERGMs as a technique that will allow them to test better their research hypotheses. Communication researchers using social network analysis have consistently sought out new methods to test their hypotheses, and with the improvements in methodologies, their hypotheses have become more advanced. However, the current social analysis methods used by many communication researchers have three limitations. First, some researchers have sacrificed the richness of network observations by reducing their data to descriptive measures and then subjecting them to other statistical analyses. Second, some researchers have ignored the interdependence of observations and treated network measures as independent. Third, and perhaps most important, the analysis methods available to communication researchers limited the number and type of variables that could be analyzed.

ERGMs (p^*) provide a superior method to analyze the influence of both attributes and internal network structures on an observed network. In the example, we demonstrate how attributes and structural variables can be estimated simultaneously and then the fit of the estimated model can be compared to the observed network. ERGMs provide a method that better suits the nature of network data and is flexible enough to allow communication researchers to adapt the method for various research projects.

REFERENCES

Barnett, G. A., & Danowski, J. (1992). The structure of communication: A network analysis of the international communication association. *Human Communication Research, 19*, 264–285.

Beinstein, J. (1977). Friends, the media, and opinion formation. *Journal of Communication, 27*, 30–39.

Bryne, D. E. (1971). *The attraction paradigm*. New York, NY: Academic Press.

Carley, K. M., & Kaufer, D. S. (1993). Semantic connectivity: An approach for analyzing symbols in semantic networks. *Communication Theory, 3*, 183–213.

Choi, J. H., & Danowski, J. (2002). Making a global community on the net — global village or global metropolis? A network analysis of USENET newsgroups. *Journal of Computer Mediated Communication, 7*(3), article 7. Retrieved from http://jcmc.indiana.edu/vol7/issue3/choi.html

Corman, S. R. (1990). A model of perceived communication in collective networks. *Human Communication Research, 16*, 582–602.

Danowski, J. A. (1980). Group attitude uniformity and connectivity of organizational communication networks for production, innovation, and maintenance content. *Human Communication Research, 6*, 299–308.

Danowski, J. A., & Edison-Swift, P. (1985). Crisis effects on intraorganizational computer-based communication. *Communication Research, 12*, 251–270.

Dekker, D., Krackhardt, D., & Snijders, T. (2007). Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika*, *72*, 563–581.

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

Doerfel, M. L., & Taylor, M. (2004). Network dynamics of interorganizational cooperation: The Croatian civil society movement. *Communication Monographs, 71*, 373–394.

Feeley, T. H. (2000). Testing a communication network model of employee turnover based on centrality. *Journal of Applied Communication Research, 28*, 262–278.

Fine, M. G. (1981). Soap opera conversations: The talk that binds. *Journal of Communication, 31*, 97–107.

Flanagin, A. J., Monge, P. R., & Fulk, J. (2001). The value of formative investment in organizational federations. *Human Communication Research, 27*, 69–93.

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

Goodreau, S. M. (2007). Advance in exponential random graph (p^*) models applied to large social networks. *Social Networks, 29*, 231–248.

Guerrero, L. K., & Burgoon, J. K. (1996). Attachment styles and reactions to nonverbal involvement change in romantic dyads patterns of reciprocity and compensation. *Human Communication Research, 22*, 335–370.

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. Retrieved from http://csde.washington.edu/statnet/

Kang, N., & Choi, J. H. (1999). Structural implications of the crossposting network of international news in cyberspace. *Communication Research, 26*, 454–481.

Livingstone, S. M. (1987). The implicit representation of characters in Dallas: A multidimensional scaling approach. *Human Communication Research, 13*, 399.

MacDonald, D. (1976). Communication roles and communication networks in a formal organization. *Human Communication Research, 2*, 365–375.

McPherson, M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations: Status distance and the composition of face to face groups. *American Sociological Review, 52*, 370–379.

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

Monge, P. R., & Matei, S. A. (2004). The role of the global telecommunications network in bridging economic and political divides, 1989 to 1999. *Journal of Communication, 54*, 511–531.

Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of exponential-family random graph models: Terms and computational aspects. *Journal of Statistical Software, 24*(4). Retrieved from http://www.jstatsoft.org/v24/i04/

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

Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology, 52*, 169–193.

Prell, C. (2003). Community networking and social capital: Early investigations. *Journal of Computer Mediated Communication, 8*(3), article 4. Retrieved from http://jcmc.indiana.edu/vol8/issue3/prell.html

Reese, S. D., Grant, A., & Danielian, L. H. (1994). The structure of news sources on television: A network analysis of "CBS news," "Nightline," "MacNeil/Lehrer," and "This week with David Brinkley." *Journal of Communication, 44*, 84–107.

Reeves, B., & Borgman, C. L. (1983). A bibliometric evaluation of core journals in communication research. *Human Communication Research, 10*, 119–136.

Rice, R. E., Borgman, C. L., & Reeves, B. (1988). Citation networks of communication journals, 1977-1985: Cliques and positions, citations made and citations received. *Human Communication Research, 15*, 256–283.

Rice, R. E., & Love, G. (1987). Electronic emotion: Socioemotional content in a computer-mediated communication network. *Communication Research, 14*(1), 85–108.

Richards, W., & Seary, A. (2007). MULTINET for Windows (Version 4.75). Available from http://www.sfu.ca/~richards/ Multinet/Pages/multinet.htm

Robins, G. (2007). *Workshop on exponential random graph models for social networks*. Paper presented at the Age of Networks, University of Illinois, Urbana-Champaign.

Robins, G., Elliott, P., & Pattison, P. (2001). Network models for social selection processes. *Social Networks, 23*, 1–30.

Robins, G., & Morris, M. (Eds.). (2007a). Advances in exponential random graph (p*) models [Special section] *Social Networks*, *29*, 169–248.

Robins, G., & Morris, M. (2007b). Advances in exponential random graph (*p**) models. *Social Networks, 29*, 169–172.

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

Robins, G., Pattison, P., & Wang, P. (2009). Closure, connectivity and degrees: New specifications for exponential random graph (*p**) models for directed social networks. *Social Networks, 31*, 105–117.

Robins, G., Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social networks: III. Valued relations. *Psychometrika, 64*, 371–394.

Robins, G., Snijders, T., Wang, P., Handcock, M., & Pattison, P. (2007). Recent developments in exponential random graph (*p**) models for social networks. *Social Networks, 29*, 192–215.

Rogers, E. M., & Antola, L. (1985). Telenovelas: A Latin American success story. *Journal of Communication, 35*, 24–35.

Rogers, R., & Marres, N. (2000). Landscaping climate change: A mapping technique for understanding science and technology debates on the World Wide Web. *Public Understanding Science, 9*, 141–163.

Scott, J. (2000). *Social network analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Seary, A. (1999). PSPAR: Sparse matrix version of PSPAR. Available from http://www.sfu.ca/~richards/Pages/ pspar.html

Shumate, M., & Dewitt, L. (2008). The North/South divide in NGO hyperlink networks. *Journal of Computer Mediated Communication, 13*, 405–428.

Shumate, M., & Lipp, J. (2008). The role of generalists in connective collective action online. *Journal of Computer Mediated Communication, 14*, 178–201.

Shumate, M., Fulk, J., & Monge, P. R. (2005). Predictors of the international HIV/AIDS INGO network over time. *Human Communication Research, 31*, 482–510.

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology, 31*, 361–395.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology, 36*, 99–153.

Snijders, T. A. B., Steglich, C. E. G., Schweinberg, M., & Huisman, M. (2007). Siena version 3.1. Available from http://stat.gamma.rug.nl/siena.html

So, C. Y. K. (1988). Citation patterns of core communication journals: An assessment of the developmental status of communication. *Human Communication Research, 15*, 236–255.

Stohl, C. (1993). European managers' interpretations of participation: A semantic network analysis. *Human Communication Research, 20*, 97–117.

Strauss, D., & Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association, 85*, 204–212.

Sykes, R. E. (1983). Initial interaction between strangers and acquaintances: A multivariate analysis of factors affecting choice of communication partners. *Human Communication Research, 10*, 27–53.

Su, C. (2008). Where to get information in the workplace? A multi-theoretical network perspective on information retrieval from team members and digital knowledge repositories. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.

Tateo, L. (2005). The Italian extreme right on-line network: An exploratory study using an integrated social network analysis and content analysis approach. *Journal of Computer-Mediated Communication, 10*(2), article 10. Retrieved from http://jcmc.indiana.edu/vol10/issue2/tateo.html

Taylor, M., & Doerfel, M. L. (2003). Building interorganizational relationships that build nations. *Human Communication Research, 29*, 153–181.

Wang, P., Robins, G., & Pattison, P. (2006). *PNET: Program for the simulation and estimation of p* exponential random graph models*. Melbourne, Australia: University of Melbourne. Retrieved from http://www.sna.unimelb.edu.au/pnet/ pnet.html

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge, England: Cambridge University Press.

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

Yuan, Y. C., & Gay, G. (2006). Homophily of network ties and bonding and bridging social capital in computer-mediated distributed teams. *Journal of Computer-Mediated Communication, 11*, 1062–1084.

Yum, J. O. (1982). Communication diversity and information acquisition among Korean immigrants in Hawaii. *Human Communication Research, 8*, 154.