Inferential Approaches for Network Analysis:
AMEN for Latent Factor Models[☆]

Abstract

We introduce a Bayesian approach to conduct inferential analyses on dyadic data while accounting for interdependencies between observations through a set of additive and multiplicative effects (AME). The AME model is built on a generalized linear model framework and is thus flexible enough to be applied to binary, ordinal, and continuous network data in both directed and undirected contexts. We contrast the AME model to two prominent approaches in the literature: the latent space model (LSM) and the exponential random graph model (ERGM). We show that the LSM has two notable limitations relative to AME. First, estimation of the effect of exogenous parameters of interest is biased due to the construction of the latent variable term in the LSM. Second, the LSM is limited in its applicability to studying networks as it can only account for a particular type of dependence pattern that emerges in relational data. ERGMs have also long been a powerful method through which to conduct inference on networks. Relative to the AME framework, ERGMs are best suited for cases in which scholars are interested in studying the role that particular types of node- and dyad-based network configurations play in generating the network. Though valuable this is often orthogonal to the interest of most researchers who are focused on studying the affect of a particular exogenous variable, such as democracy, on a dyadic variable like conflict while simply accounting for network dependencies. We show that the benefits of the AME approach are that it can be more easily interpreted, avoids computational issues such as degeneracy, and is easily implemented. Further, using an out-of-sample cross-validation procedure, we show that this approach out-performs alternative network based procedures in capturing the data generating process underlying networks.

Network analysis provides a way to represent and study "relational data", that is data with characteristics extending beyond those of the individual. Data structures that define relations between pairs of actors are ubiquitous in political science – examples include the study of events such as legislation cosponsorship, trade, interstate conflict, and the formation of international agreements. The dominant paradigm for dealing with such data, however, is not a network approach but rather a dyadic design, in which an interaction between a pair of actors is considered independent of interactions between any other pair in the system. To highlight the ubiquity of this approach the following represent just a sampling of the articles published from the 1980s to the present in the American Journal of Political Science (AJPS) and American Political Science Review (APSR) that assume dyadic independence: Dixon (1983); Mansfield, Milner and Rosendorff (2000); Lemke and Reed (2001); Mitchell (2002); Dafoe (2011); Fuhrmann and Sechser (2014); Carnegie (2014).

The implication of this assumption is that when, for example, Vietnam and the United States decide to form a trade agreement, they make this decision independently of what they have done with other countries and what other countries in the international system have done among themselves.[1] An even stronger assumption is that Japan declaring war against the United States is independent of the decision of the United States to go to war against Japan.[2] A common refrain from those that favor the dyadic approach is that many events are only bilateral (Diehl and Wright 2016), thus alleviating the need for an approach that incorporates interdependencies between observations. However, the network perspective asserts that even bilateral events and processes take place within a broader system. What takes place in one part of the system may be dependent upon events in another. At a minimum, we don't know whether independence of events and processes characterizes what we observe. We should at least examine this assertion.

---

[1]There has been significant work done on treaty formation that would challenge this claim, e.g., see Manger, Pickup and Snijders (2012); Kinne (2013).

[2]Maoz et al. (2006); Minhas, Hoff and Ward (2016) would each note the importance of taking into account network dynamics in the study of interstate conflict.

In this article, we introduce the additive and multiplicative effects (AME) model and compare it to two popular alternatives: the latent space model (LSM) and exponential random graph model (ERGM). The AME approach to network modeling is a flexible framework that can be used to estimate many different types of cross-sectional and longitudinal networks with binary, ordinal, or continuous edges within a generalized linear model framework. Our approach addresses ways in which observations can be interdependent while still allowing scholars to focus on examining theories that may only be relevant in the monadic or dyadic level. Further, at the network level it accounts for nodal and dyadic dependence patterns, and provides a descriptive visualization of higher-order dependencies such as homophily and stochastic equivalence.

The paper is organized as follows, we begin by briefly discussing the difficulties in studying dyadic data through approaches that assume observational independence; and how ERGMs provide one avenue to to address these issues. Then we introduce the AME framework in two steps. We first discuss nodal and dyadic dependencies that may lead to non-iid observations and show how the additive effects portion of AME can be used to model these dependencies. Similarly, in the second step, we discuss how the multiplicative effects portion of the AME framework can be used to effectively model third order effects, particularly when compared to alternatives such as the LSM. We conclude with

To highlight the differences between AME and alternative approaches for conducting statistical inference on networks, we examine a cross-sectional network measuring collaborations during the policy design of the Swiss CO_2 act. We show that AME provides a superior goodness of fit to the data in terms of ability to predict linkages and capture network dependencies.

Addressing Dependencies in Dyadic Data

Relational, or dyadic, data provide measurements of how pairs of actors relate to one another. The easiest way to organize such data is the directed dyadic design in which the unit of analysis is some set of n actors that have been paired together to form a dataset of z directed dyads. A tabular design such as this for a set of n actors, $\{i, j, k, l\}$ results in $n \times (n-1)$

observations, as shown in Table 1.

Sender	Receiver	Event
i	j	y_{ij}
	k	y_{ik}
$\vdots$	l	y_{il}
j	i	y_{ji}
	k	y_{jk}
$\vdots$	l	y_{jl}
k	i	y_{ki}
	j	y_{kj}
$\vdots$	l	y_{kl}
l	i	y_{li}
	j	y_{lj}
$\vdots$	k	y_{lk}

Table 1: Structure of datasets used in canonical design.

$\longrightarrow$

	i	j	k	l
i	NA	y_{ij}	y_{ik}	y_{il}
j	y_{ji}	NA	y_{jk}	y_{jl}
k	y_{ki}	y_{kj}	NA	y_{kl}
l	y_{li}	y_{lj}	y_{lk}	NA

Table 2: Adjacency matrix representation of data in Table 1. Senders are represented by the rows and receivers by the columns.

When modeling these types of data, scholars typically employ a generalized linear model (GLM) estimated via maximum-likelihood. The stochastic component of this model reflects our assumptions about the probability distribution from which the data are generated: $y_{ij} \sim P(Y|\theta_{ij})$, with a probability density or mass function such as the normal, binomial, or Poisson, and we assume that each dyad in the sample is independently drawn from that particular distribution, given θ_{ij}. The systematic component characterizes the model for the parameters of that distribution and describes how θ_{ij} varies as a function of a set of nodal and dyadic covariates, $\mathbf{X}_{ij}$: $\theta_{ij} = \boldsymbol{\beta}^T \mathbf{X}_{ij}$. The key assumption we make when applying this modeling technique is that given $\mathbf{X}_{ij}$ and the parameters of our distribution, each of the dyadic observations is conditionally independent. Specifically, we construct the joint density function over all dyads using the observations from Table 1 as an example.

$$P(y_{ij}, y_{ik}, \ldots, y_{lk} | \theta_{ij}, \theta_{ik}, \ldots, \theta_{lk}) = P(y_{ij}|\theta_{ij}) \times P(y_{ik}|\theta_{ik}) \times \ldots \times P(y_{lk}|\theta_{lk})$$

$$P(\mathbf{Y} \mid \boldsymbol{\theta}) = \prod_{\alpha=1}^{n \times (n-1)} P(y_\alpha | \theta_\alpha)$$

(1)

We next convert the joint probability into a likelihood: $\mathcal{L}(\boldsymbol{\theta}|\mathbf{Y}) = \prod\limits_{\alpha=1}^{n\times(n-1)} P(y_\alpha|\theta_\alpha)$.

The likelihood as defined above is only valid if we are able to make the assumption that, for example, y_{ij} is independent of y_{ji} and y_{ik} given the set of covariates we specified, or the values of θ_{ij}.[3] Assuming that the dyad y_{ij} is conditionally independent of the dyad y_{ji} asserts that there is no level of reciprocity in a dataset, an assumption that in many cases would seem quite untenable. A harder problem to handle is the assumption that y_{ij} is conditionally independent of y_{ik}, the difficulty here follows from the possibility that i's relationship with k is dependent on how i relates to j and how j relates to k, or more simply put the "enemy of my enemy [may be] my friend".

ERGMs

One way to account for dependence patterns in dyadic data are ERGMs.[4] The ERGM framework is particularly useful when researchers are interested in the role that a specific list of network statistics have in giving rise to a certain network. These network statistics could include the number of transitive triads in a network, balanced triads, reciprocal pairs and so on.[5] In the ERGM framework, a set of statistics, $S(\mathbf{Y})$, define a model. Given the chosen set of statistics, the probability of observing a particular network dataset $\mathbf{Y}$ can be expressed as:

---

[3]The difficulties of applying the GLM framework to data that have structural interdependencies between observations is a problem that has long been recognized. Beck and Katz (1995), for example, detail the issues with pooling observations in time-series cross-section datasets.

[4]This approach was first developed by Erdős and Rényi (1959), and has became more widely understood as it has been applied to particular problems. Frank (1971) undertook an early examination and Julian Besag developed interesting applications and methods promoting their examination (Besag, 1977). But obtaining estimates was computational challenging and it was not until Frank and Strauss (1986) and Wasserman and Pattison (1996) that these methods found widespread application.

[5]Morris, Handcock and Hunter (2008) and Snijders et al. (2006) provide a detailed list of network statistics that can be included in an ERGM model specification.

$$\Pr(Y = y) = \frac{\exp(\boldsymbol{\beta}^\top S(y))}{\sum_{z \in \mathcal{Y}} \exp(\boldsymbol{\beta}^\top S(z))} \ , y \in \mathcal{Y}. \tag{2}$$

$\boldsymbol{\beta}$ represents a vector of model coefficients for the specified network statistics, $\mathcal{Y}$ denotes the set of all obtainable networks, and the denominator is used as a normalizing factor (Hunter et al. 2008).[6] This approach provides a way to state that the probability of observing a given network depends on the patterns that it exhibits, which are operationalized in the list of network statistics specified by the researcher. Within this approach one can test the role that a variety of network statistics play in giving rise to a particular network.

Further if one has specified the "correct" set of network statistics, then our ability to conduct inference on exogenous parameters also improves relative to GLM.

A notable issue when estimating ERGMs is that the estimated model can become degenerate even if the observed graph is not degenerate. This means that the model is placing a large amount of probability on a small subset of networks that fall in the set of obtainable networks, $\mathcal{Y}$, but share little resemblance with the observed network (Schweinberger, 2011). For example, most of the probability may be placed on empty graphs, no edges between nodes, or nearly complete graphs, almost every node is connected by an edge. Some have argued that model degeneracy is simply a result of model misspecification (Goodreau et al., 2008; Handcock et al., 2008).

This points to an important

Failure to properly account for higher-order dependence structures through an appropriate specification can at best lead to model degeneracy, which provides an obvious indication that the specification needs to be altered, and at worst deliver a result that converges but does not

---

[6]One issue that arises when conducting statistical inference with this model is in the calculation of the normalizing factor, which is what ensures that the expression above corresponds to a legitimate probability distribution. For even a trivially sized directed network that has only 20 actors, calculating the denominator means summing over $2^{20 \times (20-1)} = 2^{380}$ possible networks, or, to put it another way, more than the total number of atoms in the universe.

appropriately capture the interdependencies in the network. The consequence of the latter case is a set of inferences that will continue to be biased as a result of unmeasured heterogeneity, thus defeating the major motivation for pursuing an inferential network model in the first place.

Nonetheless, as cranmer et al have shown this approach can be very useful in not only accounting for dependence patterns in networks but also for better understanding how networks shape .

two parts of the AME model

Social Relations Model: Additive Part of AME

The dependencies that tend to develop in relational data can be more easily understood when we move away from stacking dyads on top of one another and turn instead to a matrix design as illustrated in Table 2. Operationally, this type of data structure is represented as a $n \times n$ matrix, $\mathbf{Y}$, where the diagonals are typically undefined. The ij^{th} entry defines the relationship sent from i to j and can be continuous or discrete. Relations between actors in a network setting at times does not involve senders and receivers. Networks such as these are referred to as undirected and all the relations between actors are symmetric, meaning $y_{ij} = y_{ji}$.

The most common type of dependency that arises in networks are first-order, or nodal dependencies, and these point to the fact that we typically find significant heterogeneity in activity levels across nodes. The implication of this across-node heterogeneity is within-node homogeneity of ties, meaning that values across a row, say $\{y_{ij}, y_{ik}, y_{il}\}$, will be more similar to each other than other values in the adjacency matrix because each of these values has a common sender i. This type of dependency manifests in cases where sender i tends to be more active or less active in the network than other senders. Similarly, while some actors may be more active in sending ties to others in the network, we might also observe that others are more popular targets, this would manifest in observations down a column, $\{y_{ji}, y_{ki}, y_{li}\}$, being more similar. Last, we might also find that actors who are more likely to send ties in a network are also more likely to receive them, meaning that the row and column means of an adjacency matrix

may be correlated. Another ubiquitous type of structural interdependency is reciprocity. This is a second-order, or dyadic, dependency relevant only to directed datasets, and asserts that values of y_{ij} and y_{ji} may be statistically dependent. The prevalence of these types of potential interactions within directed dyadic data also complicates the basic assumption of observational independence.[7]

The relevance of modeling first- and second-order dependencies has long been recognized within fields such as psychology. In 1979, Warner et al. developed the social relations model (SRM), a type of ANOVA decomposition technique, that facilitates this undertaking Warner, Kenny and Stoto (1979). The SRM is of particular note as it provides the error structure for the additive effects component of the AME framework that we introduce here. The goal of the SRM is to decompose the variance of observations in an adjacency matrix in terms of heterogeneity across row means (out-degree), heterogeneity along column means (in-degree), correlation between row and column means, and correlations within dyads. Wong and Li & Loken and provide a random effects representation of the SRM Wong (1982); Li and Loken (2002):

$$y_{ij} = \mu + e_{ij}$$

$$e_{ij} = a_i + b_j + \epsilon_{ij}$$

$$\{(a_1, b_1), \ldots, (a_n, b_n)\} \overset{\text{iid}}{\sim} N(0, \Sigma_{ab})$$

$$\{(\epsilon_{ij}, \epsilon_{ji}) : \ i \neq j\} \overset{\text{iid}}{\sim} N(0, \Sigma_\epsilon), \text{ where}$$

$$\Sigma_{ab} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix} \quad \Sigma_\epsilon = \sigma_\epsilon^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

(3)

In the above, μ provides a baseline measure of the density mean of a network, and e_{ij} represents residual variation. The residual variation decomposes into parts: a row/sender effect (a_i), a column/receiver effect (b_j), and a within-dyad effect (ϵ_{ij}). The row and column effects

---

[7]In the Appendix, we include a lengthier discussion on the limitations of studying relational data via the typical GLM framework.

are modeled jointly to account for correlation in how active an actor is in sending and receiving ties. Heterogeneity in the row and column means is captured by σ_a^2 and σ_b^2, respectively, and σ_{ab} describes the linear relationship between these two effects (i.e., whether actors who send a lot of ties also receive a lot of ties). Beyond these first-order dependencies, second-order dependencies are described by σ_ϵ^2 and a within dyad correlation, or reciprocity, parameter ρ.

The SRM covariance structure described in Equation 3 can be incorporated into the systematic component of a GLM framework to produce the social relations regression model (SRRM): $\boldsymbol{\beta}^\top \mathbf{X}_{ij} + a_i + b_j + \epsilon_{ij}$, where $\boldsymbol{\beta}^\top \mathbf{X}_{ij}$ accommodates the inclusion of dyadic, sender, and receiver covariates. The SRRM approach incorporates row, column, and within-dyad dependence in way that is widely used and understood by applied researchers: a regression framework and additive random effects to accommodate variances and covariances often seen in relational data. Furthermore, this handles a diversity of outcome distributions. In the case of binary data, this can be done by utilizing a random effects logistic (or probit) latent variable representation.

**Latent Factor Model: Multiplicative Part of AME**

Missing from the framework provided by the SRM is an accounting of third-order dependence patterns that can arise in relational data. A third-order dependency is defined as the dependency between triads, not dyads. The ubiquity of third-order effects in relational datasets can arise from the presence of some set of shared attributes between nodes that affects their probability of interacting with one another. Another reason why we may see the emergence of third-order effects is the "sociology" explanation: that individuals want to close triads because this is putatively a more stable or preferable social situation (Wasserman and Faust 1994).

For example, one finding from the gravity model of trade is that neighboring countries are more likely to trade with one another; in this case, the shared attribute is simply geographic proximity. A finding common in the political economy literature is that democracies are more likely to form trade agreements with one another, and the shared attribute here is a country's political system. Both geographic proximity and a country's political system are examples of homophily, which captures the idea that the relationships between actors with similar character-

istics in a network are likely to be stronger than nodes with different characteristics. Homophily can be used to explain the emergence of patterns such as transitivity ("a friend of a friend is a friend") and balance ("an enemy of a friend is an enemy").

A binary network where actors tend to form ties with others based on some set of shared characteristics often leads to a network graph with a high number of "transitive triads" in which sets of actors $\{i, j, k\}$ are each linked to one another. The left-most plot in Figure 1 provides a representation of a network that exhibits this type of pattern. The relevant implication of this when it comes to conducting statistical inference is that–unless we are able to specify the list of exogenous variable that may explain this prevalence of triads–the probability of j and k forming a tie is not independent of the ties that already exist between those actors and i.

Figure 1: Graph on the left is a representation of an undirected network that exhibits a high degree of homophily, while on the right we show an undirected network that exhibits stochastic equivalence.
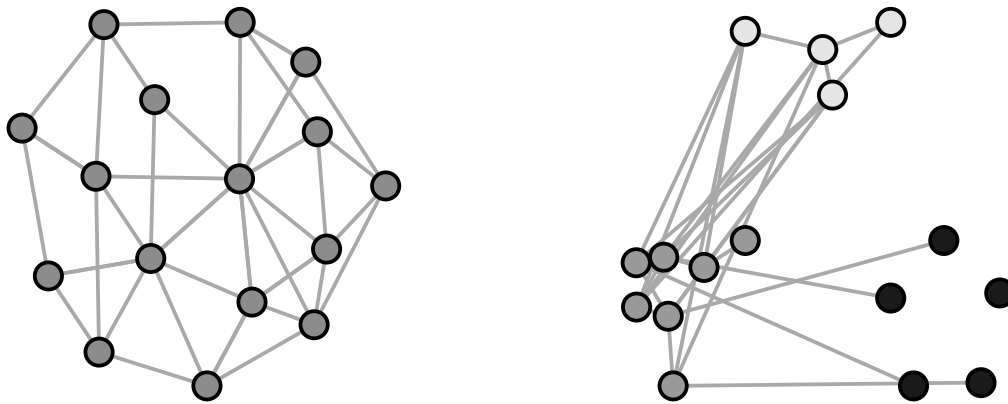
Another third-order dependence pattern that cannot be accounted for in the additive effects framework is stochastic equivalence. A pair of actors ij are stochastically equivalent if the probability of i relating to, and being related to, by every other actor is the same as the probability for j. This refers to the idea that there will be groups of nodes in a network with similar relational patterns. The occurrence of a dependence pattern such as this is not uncommon in the social science applications. Recent work estimates a stochastic equivalence structure to explain the formation of preferential trade agreements (PTAs) between countries Manger, Pickup and Snijders (2012). Specifically, they suggest that PTA formation is related to differences in per

capita income levels between countries. Countries falling into high, middle, and low income per capita levels will have patterns of PTA formation that are determined by the groups into which they fall. Such a structure is represented in the right-most panel of Figure 1, here the lightly shaded group of nodes at the top can represent high-income countries, nodes on the bottom-left middle-income, and the darkest shade of nodes low-income countries. The behavior of actors in a network can at times be governed by group level dynamics, and failing to account for such dynamics leaves potentially important parts of the data generating process ignored.

To account for third-order dependence patterns within the context of the SRRM we turn to latent variable models, which have become a popular approach for modeling relational data in fields as diverse as biology to computer science to the social sciences. These models assumes that relationships between nodes are mediated by a small number (K) of node-specific unobserved latent variables. One reason for their increased usage is that they enable researchers to capture and visualize third-order dependencies in a way that other approaches are not able to replicate.

A number of latent variable approaches have been developed to represent third-order dependencies in relational data, we focus on two here: the the latent space model (LSM) and the latent factor model (LFM).[8] For the sake of exposition, we consider the case where relations are symmetric to describe the differences between these approaches. These approaches can be incorporated into the framework that we have been constructing through the inclusion of an additional term, $\alpha(\mu_i, \mu_j)$, that captures latent third order characteristics of a network. General definitions for how $\alpha(u_i, u_j)$ are defined for these latent variable models are shown in Equations 4. One other point of note about these approaches is that researchers have to specify a value for K.[9] These two approaches are described below:

---

[8]An alternative approach with a similar latent variable formulation is known as the stochastic block model (Nowicki and Snijders, 2001).

[9]In the case of both the latent space and factor models, a value of K equal to two or three is typically large enough to account for third-order dependencies in relational data.

Latent space model

$$\alpha(\mathbf{u}_i, \mathbf{u}_j) = -|\mathbf{u}_i - \mathbf{u}_j|$$

$$\mathbf{u}_i \in \mathbb{R}^K, \ i \in \{1, \ldots, n\}$$

Latent factor model   (4)

$$\alpha(\mathbf{u}_i, \mathbf{u}_j) = \mathbf{u}_i^\top \Lambda \mathbf{u}_j$$

$$\mathbf{u}_i \in \mathbb{R}^K, \ i \in \{1, \ldots, n\}$$

Λ a $K \times K$ diagonal matrix

In the LSM approach, each node i has some unknown latent position in K dimensional space, $\mathbf{u}_i \in \mathbb{R}^K$, and the probability of a tie between a pair ij is a function of the negative Euclidean distance between them: $-|\mathbf{u}_i - \mathbf{u}_j|$. Because latent distances for a triple of actors obey the triangle inequality, this formulation models the tendencies toward homophily commonly found in social networks. This approach is implemented in the **latentnet** which is part of the **statnet** R package Krivitsky and Handcock (2015). However, this approach also comes with an important shortcoming: it confounds stochastic equivalence and homophily. Consider two nodes i and j that are proximate to one another in K dimensional Euclidean space, this suggests not only that $|\mathbf{u}_i - \mathbf{u}_j|$ is small but also that $|\mathbf{u}_i - \mathbf{u}_l| \approx |\mathbf{u}_j - \mathbf{u}_l|$, the result being that nodes i and j will by construction assumed to possess the same relational patterns with other actors such as l (i.e., that they are stochastically equivalent). Thus LSMs confound strong ties with stochastic equivalence. This approach cannot adequately model data with many ties between nodes that have different network roles.

This is problematic as real-world networks exhibit varying degrees of stochastic equivalence and homophily. In these situations, using the LSM would end up representing only a part of the network structure. In the latent factor model, each actor has an unobserved vector of characteristics, $\mathbf{u}_i = \{u_{i,1}, \ldots, u_{i,K}\}$, which describe their behavior as an actor in the network. The probability of a tie from i to j depends on the extent to which $\mathbf{u}_i$ and $\mathbf{u}_j$ are "similar" (i.e.,

point in the same direction) and on whether the entries of Λ are greater or less than zero.

More specifically, the similarity in the latent factors, $\mathbf{u}_i \approx \mathbf{u}_j$, corresponds to how stochastically equivalent a pair of actors are and the eigenvalue determines whether the network exhibits positive or negative homophily. For example, say that we estimate a rank-one latent factor model (i.e., $K = 1$), in this case $\mathbf{u}_i$ is represented by a scalar $u_{i,1}$, similarly, $\mathbf{u}_j = u_{j,1}$, and Λ will have just one diagonal element λ. The average effect this will have on y_{ij} is simply $\lambda \times u_i \times u_j$, where a positive value of $\lambda > 0$ indicates homophily and $\lambda < 0$ heterophily. This approach can represent both varying degrees of homophily and stochastic equivalence. In the directed version of this approach, we use the singular value decomposition, here actors in the network have a vector of latent characteristics to describe their behavior as a sender, denoted by $\mathbf{u}$, and as a receiver, $\mathbf{v}$: $\mathbf{u}_i, \mathbf{v}_j \in \mathbb{R}^K$. This can alter the probability of an interaction between ij additively: $\mathbf{u}_i^\top \mathbf{D} \mathbf{v}_j$, where $\mathbf{D}$ is a $K \times K$ diagonal matrix.

Both the latent space and factor models are "conditional independence models" in that they assume that ties are conditionally independent given all of the observed predictors and unknown node-specific parameters: $p(Y|X, U) = \prod_{i<j} p(y_{i,j}|x_{i,j}, u_i, u_j)$. Typical parametric models of this form relate $y_{i,j}$ to $(x_{i,j}, u_i, u_j)$ via some sort of link function:

$$p(y_{i,j}|x_{i,j}, u_i, u_j) = f(y_{i,j} : \eta_{i,j})$$
$$\eta_{i,j} = \beta^\top x_{i,j} + \alpha(\mathbf{u}_i, \mathbf{u}_j).$$

The structure of $\alpha(\mathbf{u}_i, \mathbf{u}_j)$ can result in very different interpretations for any estimates of the regression coefficients β. For example, suppose the latent effects $\{u_1, \ldots, u_n\}$ are near zero on average (if not, their mean can be absorbed into an intercept parameter and row and column additive effects). Under the LFM, the average value of $\alpha(\mathbf{u}_i, \mathbf{u}_j) = \mathbf{u}_i^\top \Lambda \mathbf{u}_j$ will be near zero and so we have

$$\eta_{i,j} = \beta^\top x_{i,j} + \mathbf{u}_i^\top \Lambda \mathbf{u}_j$$

$$\bar{\eta} \approx \beta^\top \bar{x}.$$

The implication of this is that the values of β can be interpreted as yielding the "average" value of $\eta_{i,j}$. On the other hand, under the LSM

$$\eta_{i,j} = \beta^\top x_{i,j} - |\mathbf{u}_i - \mathbf{u}_j|$$

$$\bar{\eta} \approx \beta^\top \bar{x} - \overline{|\mathbf{u}_i - \mathbf{u}_j|} < \beta^\top \bar{x}.$$

In this case, $\beta^\top \bar{x}$ does not represent an "average" value of the predictor $\eta_{i,j}$, it represents a maximal value as if all actors were zero distance from each other in the latent social space. For example, consider the simplest case of a normally distributed network outcome with an identity link. In this case,

$$y_{i,j} = \beta^\top x_{i,j} + \alpha(\mathbf{u}_i, \mathbf{u}_j) + \epsilon_{i,j}$$

$$\bar{y} \approx \beta^\top \bar{x} + \overline{\alpha(\mathbf{u}_i, \mathbf{u}_j)}.$$

Under the LSM, $\bar{y} \approx \beta^\top \bar{x} + \overline{|\mathbf{u}_i - \mathbf{u}_j|} < \beta^\top \bar{x}$, and so we no longer can interpret β as representing the linear relationship between y and x. Instead, it may be thought of as describing some sort of average hypothetical "maximal" relationship between $y_{i,j}$ and $x_{i,j}$.

Thus the LFM provides two important benefits. First, we are able to capture a wider assortment of dependence patterns that arise in relational data, and, second, parameter interpretation is more straightforward. The AME approach considers the regression model shown in Equation 5:

$$y_{ij} = g(\theta_{ij})$$

$$\theta_{ij} = \boldsymbol{\beta}^\top \mathbf{X}_{ij} + e_{ij}$$

$$e_{ij} = a_i + b_j + \epsilon_{ij} + \alpha(\mathbf{u}_i, \mathbf{v}_j) \text{, where} \tag{5}$$

$$\alpha(\mathbf{u}_i, \mathbf{v}_j) = \mathbf{u}_i^\top \mathbf{D} \mathbf{v}_j = \sum_{k \in K} d_k u_{ik} v_{jk}.$$

Using this framework, we are able to model the dyadic observations as conditionally independent given $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ depends on the the unobserved random effects, $\mathbf{e}$. $\mathbf{e}$ is then modeled to account for the potential first, second, and third-order dependencies that we have discussed. As described in Equation 3, $a_i + b_j + \epsilon_{ij}$, are the additive random effects in this framework and account for sender, receiver, and within-dyad dependence. The multiplicative effects, $\mathbf{u}_i^\top \mathbf{D} \mathbf{v}_j$, are used to capture higher-order dependence patterns that are left over in $\boldsymbol{\theta}$ after accounting for any known covariate information.

In the following section we undertake a comparison of the latent space model, ERGM, and the AME model. In doing so, we are able to compare and contrast these various approaches.

Empirical Comparison

To contrast AME with these alternative approaches, we utilize a cross-sectional network measuring whether an actor indicated that they collaborated with each other during the policy design of the Swiss CO_2 act (Ingold 2008). This is a directed relational matrix as an actor i can indicate that they collaborated with j but j may not have stated that they collaborated with i. The Swiss government proposed this act in 1995 with the goal of undertaking a 10% reduction in CO_2 emissions by 2012. The act was accepted in the Swiss Parliament in 2000 and implemented in 2008. Ingold Ingold (2008), and subsequent work by Ingold & Fischer Ingold and Fischer (2014), sought to determine what drives collaboration among actors trying to affect climate change policy. The set of actors included in this network are those that were identified by experts

as holding an important position in Swiss climate policy. In total, Ingold identifies 34 relevant actors: five state actors, eleven industry and business representatives, seven environmental NGOs and civil society organizations, five political parties, and six scientific institutions and consultants. We follow Ingold & Fischer and Cranmer et al. Cranmer et al. (2016) in developing a model specification to understand and predict link formation in this network.[10]

The LSM we fit on this network includes a two-dimensional Euclidean distance metric. The ERGM specification for this network includes the same exogenous variables as LSM, but also includes a number of endogenous characteristics of the network. The AME model we fit here includes the same exogenous covariates and accounts for nodal and dyadic heterogeneity using the SRM. Third-order effects are represented by the latent factor model with $K = 2$. Parameter estimates for these three approaches are shown in Table 3.

The first point to note is that, in general, the parameter estimates returned by the AME are similar to those of ERGM but quite different from the LSM. For example, while the LSM returns a result for the `Opposition/alliance` variable that diverges from ERGM, the AME returns a result that is similar to Ingold & Fischer. Similar discrepancies appear for other parameters such as `Influence attribution` and `Alter's influence degree`. Each of these discrepancies are eliminated when using AME. As described previously, this is because the LSM approach complicates the interpretation of the effects of exogenous variables.[11]

---

[10]We do not review the specification in detail here, instead we just provide a summary of the variables to be included and the theoretical expectations of their effects in the Appendix.

[11]In the SI Appendix, we show that these differences persist even when incorporating sender and receiver random effects.

Table 3: ERGM results are shown with standard errors in parentheses. LSM and AME are shown with 95% posterior credible intervals provided in brackets.

	LSM	ERGM	AME
Intercept/Edges	0.94	-12.17	-3.39
	[0.09; 1.82]	(1.40)	[-4.38; -2.50]
Conflicting policy preferences			
Business vs. NGO	-1.37	-1.11	-1.37
	[-2.42; -0.41]	(0.51)	[-2.44; -0.47]
Opposition/alliance	0.00	1.22	1.08
	[-0.40; 0.39]	(0.20)	[0.72; 1.47]
Preference dissimilarity	-1.76	-0.44	-0.79
	[-2.62; -0.90]	(0.39)	[-1.55; -0.08]
Transaction costs			
Joint forum participation	1.51	0.90	0.92
	[0.86; 2.17]	(0.28)	[0.40; 1.47]
Influence			
Influence attribution	0.08	1.00	1.09
	[-0.40; 0.55]	(0.21)	[0.69; 1.53]
Alter's influence indegree	0.01	0.21	0.11
	[-0.03; 0.04]	(0.04)	[0.07; 0.15]
Influence absolute diff.	0.04	-0.05	-0.07
	[-0.01; 0.09]	(0.01)	[-0.11; -0.03]
Alter = Government actor	-0.46	1.04	0.55
	[-1.08; 0.14]	(0.34)	[-0.07; 1.15]
Functional requirements			
Ego = Environmental NGO	-0.60	0.79	0.67
	[-1.32; 0.09]	(0.17)	[-0.38; 1.71]
Same actor type	1.17	0.99	1.04
	[0.63; 1.71]	(0.23)	[0.63; 1.50]
Endogenous dependencies			
Mutuality		0.81	0.39
		(0.25)	[-0.12; 0.96]
Outdegree popularity		0.95	
		(0.09)	
Twopaths		-0.04	
		(0.02)	
GWIdegree (2.0)		3.42	
		(1.47)	
GWESP (1.0)		0.58	
		(0.16)	
GWOdegree (0.5)		8.42	
		(2.11)	

There are also a few differences between the parameter estimates that result from the ERGM and AME. Using the AME we find evidence that `Preference dissimilarity` is associated with a reduced probability of collaboration between a pair of actors, which is in line with the theoretical expectations of Ingold & Fischer. Additionally, the AME results differ from ERGM for the nodal effects related to whether a receiver of a collaboration is a government actor, `Alter=Government actor`, and whether the sender is an environmental NGO, `Ego=Environmental NGO`.

Tie Formation Prediction

Next, we utilize a cross-validation procedure to assess the out-of-sample performance for each of the models presented in Table 3 as follows:

- Randomly divide the $n \times (n-1)$ data points into S sets of roughly equal size, letting s_{ij} be the set to which pair $\{ij\}$ is assigned.

- For each $s \in \{1, \ldots, S\}$:

    - Obtain estimates of the model parameters conditional on $\{y_{ij} : s_{ij} \neq s\}$, the data on pairs not in set s.

    - For pairs $\{kl\}$ in set s, let $\hat{y}_{kl} = E[y_{kl}|\{y_{ij} : s_{ij} \neq s\}]$, the predicted value of y_{kl} obtained using data not in set s.

The procedure summarized in the steps above generates a sociomatrix of out-of-sample predictions of the observed data. Each entry $\hat{y}_{ij}$ is a predicted value obtained from using a subset of the data that does not include y_{ij}. In this application we set S to 45 which corresponds to randomly excluding approximately 2% of the data from the estimation. Such a low number of observations were excluded in every sample (denoted a fold) because excluding any more observations would cause the ERGM specification to result in a degenerate model that empirically can not be fit. This highlights the computational difficulties associated with ERGMs in the presence of even small amounts of missing data.

Using the set of out-of-sample predictions we generate from the cross-validation procedure, we provide a series of tests to assess model fit. The left-most plot in Figure 2 compares the five approaches in terms of their ability to predict the out-of-sample occurrence of collaboration based on Receiver Operating Characteristic (ROC) curves. ROC curves provide a comparison of the trade-off between the True Positive Rate (TPR), sensitivity, False Positive Rate (FPR), 1-specificity, for each model. Models that have a better fit according to this test should have curves that follow the left-hand border and then the top border of the ROC space. On this diagnostic, the AME model performs best closely followed by ERGM. The LSM approach lags notably behind the other specifications.

A more intuitive visualization of the differences between these modeling approaches can be gleaned through examining the separation plots included on the right-bottom edge of the ROC plot. This visualization tool plots each of the observations, in this case actor pairs, in the dataset according to their predicted value from left (low values) to right (high values). Models with a good fit should have all network links, here these are colored by the modeling approach, towards the right of the plot. Using this type of visualization emphasizes that the AME and ERGM models performs better than the latent space approach.

The last diagnostic we highlight to assess predictive performance are precision-recall (PR) curves. In both ROC and PR space we utilize the TPR, also referred to as recall–though in the former it is plotted on the y-axis and the latter the x-axis. The difference, however, is that in ROC space we utilize the FPR, while in PR space we use precision. FPR measures the fraction of negative examples that are misclassified as positive, while precision measures the fraction of examples classified as positive that are truly positive. PR curves are useful in situations where correctly predicting events is more interesting than simply predicting non-events (Davis and Goadrich 2006). This is especially relevant in the context of studying many relational datasets in political science such as conflict, because events in such data are extremely sparse and it is relatively easy to correctly predict non-events. In the case of our application dataset, the vast majority of dyads, 80%, do not have a network linkage, which points to the relevance of assessing performance using the PR curves as we do in the right-most plot of Figure 2. We

Figure 2: Assessments of out-of-sample predictive performance using ROC curves, separation plots, and PR curves. AUC statistics are also provided.
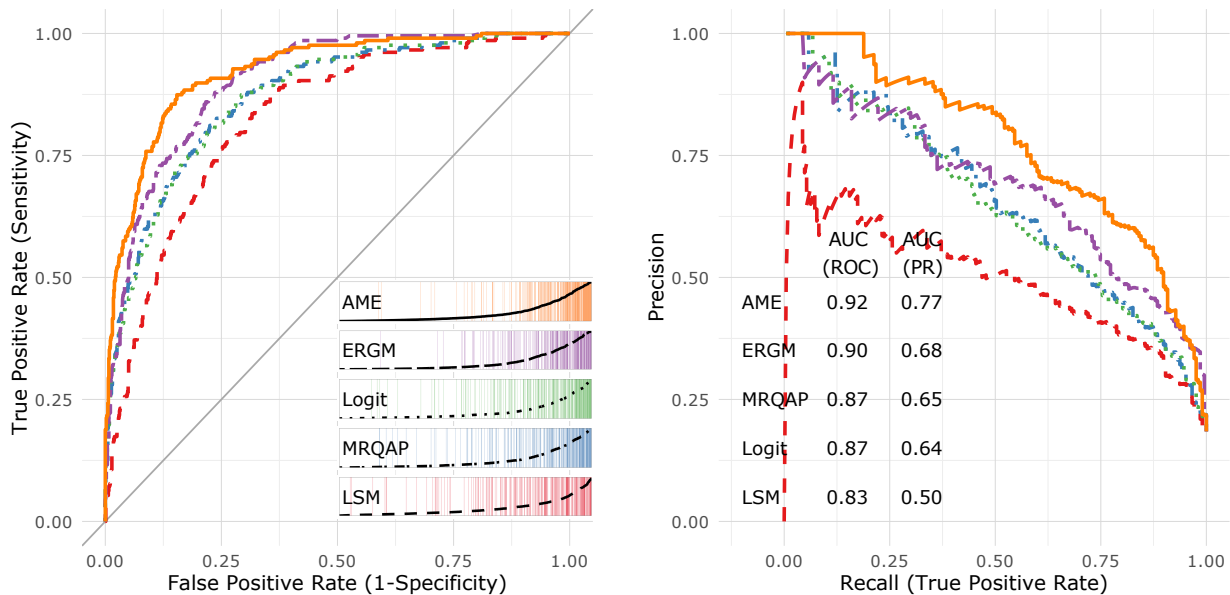

can see that the relative-ordering of the models remains similar but the differences in how well they perform become much more stark. Here we find that the AME approach performs notably better in actually predicting network linkages than each of the alternatives. Area under the curve (AUC) statistics are provided in Figure 2 and these also highlight AME's superior out-of-sample performance.

Capturing Network Attributes

We also assess which of these models best captures the network features of the dependent variable. To do this, we compare the observed network with a set of networks simulated from the estimated models. We simulate 1,000 networks from the three models and compare how well they align with the observed network in terms of four network statistics: (1) the empirical standard deviation of the row means (i.e., heterogeneity of nodes in terms of the ties they send); (2) the empirical standard deviation of the column means (i.e., heterogeneity of nodes in terms of the ties they receive); (3) the empirical within-dyad correlation (i.e., measure of reciprocity in the network); and (4) a normalized measure of triadic dependence. A comparison of the LSM,

ERGM, and AME models among these four statistics is shown in Figure 3.[12]

Figure 3: Network goodness of fit summary using **amen**.
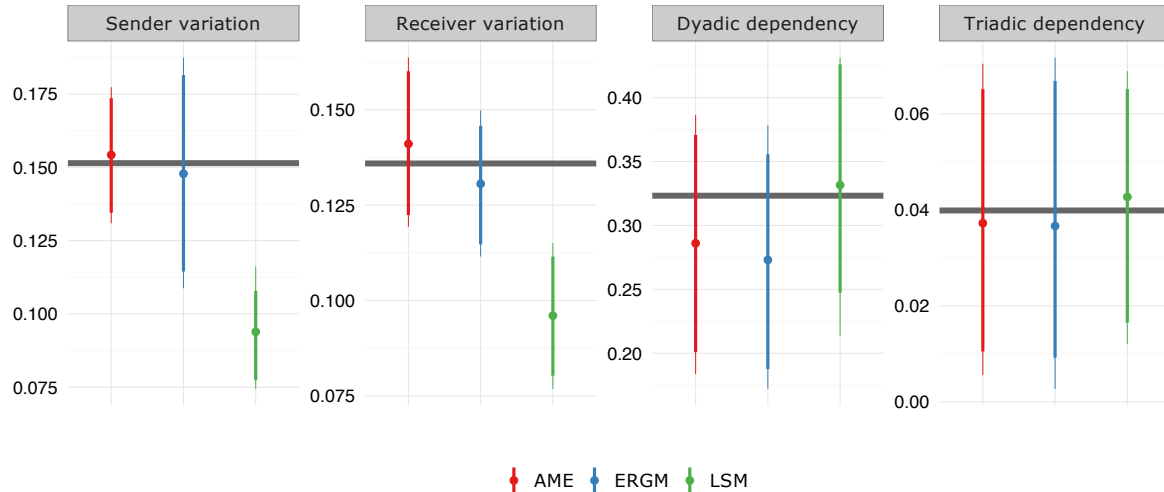
Here it becomes quickly apparent that the LSM model fails to capture how active and popular actors are in the Swiss climate change mitigation network.[13] The AME and ERGM specifications again both tend to do equally well. If when running this diagnostic, we found that the AME model did not adequately represent the observed network this would indicate that we might want to increase K to better account for network interdependencies. No changes to the model specification as described by the exogenous covariates a researcher has chosen would be necessary. If the ERGM results did not align with the diagnostic presented in Figure 3, then this would indicate that an incorrect set of endogenous dependencies have been specified. Failing to identify (or find) the right specification will leave the researcher with the problems we introduced earlier.

---

[12]In the SI Appendix, we compare the ability of these models to capture network attribute across a wider array of statistics, and the results are consistent with what we present below.

[13]Further even after incorporating random sender and receiver effects into the LSM framework this problem is not completely resolved, see the SI Appendix for details.

Conclusion

The AME approach to estimation and inference in network data provides a number of benefits over alternative approaches. Specifically, it provides a modeling framework for dyadic data that is based on familiar statistical tools such as linear regression, GLM, random effects, and factor models. We have an understanding of how each of these tools work, they are numerically more stable than ERGM approaches, and more general than alternative latent variable models. Further the estimation procedure utilized in AME avoids complicating interpretation of parameter estimates for exogenous variables. For researchers in the social sciences this is of primary interest, as many studies that employ relational data still have conceptualizations that are monadic or dyadic in nature. Additionally, through the application dataset utilized herein we show that the AME approach outperforms both ERGM and LSM in out-of-sample prediction, and also is better able to capture network dependencies than the LSM.

More broadly, relational data structures are composed of actors that are part of a system. It is unlikely that this system can be viewed simply as a collection of isolated actors or pairs of actors. The assumption that dependencies between observations occur can at the very least be examined. Failure to take into account interdependencies leads to biased parameter estimates and poor fitting models. By using standard diagnostics such as shown in Figure 3, one can easily assess whether an assumption of independence is reasonable. We stress this point because a common misunderstanding that seems to have emerged within the social science literature relying on dyadic data is that a network based approach is only necessary if one has theoretical explanations that extend beyond the dyadic. This is not at all the case and findings that continue to employ a dyadic design may misrepresent the effects of the very variables that they are interested in. The AME approach that we have detailed here provides a statistically familiar way for scholars to account for unobserved network structures in relational data. Additionally, through this approach we can visualize these dependencies in order to better learn about the network patterns that remain in the event of interest after having accounted for observed covariates.

When compared to other network based approaches, AME is easier to specify and utilize.

Beck, Nathaniel and Jonathan N. Katz. 1995. "What To Do (and Not To Do) with Pooled Time-Series Cross-Section Data." *American Political Science Review* 89(3):634–647.

Besag, Julian E. 1977. "Efficiency of Pseudolikelihood Estimation for Simple Gaussian Fields." *Biometrika* 64:616.

Carnegie, Allison. 2014. "States held hostage: Political hold-up problems and the effects of international institutions." *American Political Science Review* 108(01):54–70.

Cranmer, Skyler J., Philip Leifeld, Scott D. McClurg and Meredith Rolfe. 2016. "Navigating the Range of Statistical Tools for Inferential Network Analysis." *American Journal of Political Science* .
    URL: *http://onlinelibrary.wiley.com/doi/10.1111/ajps.12263/abstract*

Dafoe, Allan. 2011. "Statistical Critiques of the Democratic Peace: Caveat Emptor." *American Journal of Political Science* 55(2):247–262.

Davis, Jesse and Mark Goadrich. 2006. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning*. ACM pp. 233–240.

Diehl, Paul F. and Thorin M. Wright. 2016. "A Conditional Defense of the Dyadic Approach." *International Studies Quarterly* .
    URL: *http://isq.oxfordjournals.org/content/60/2/363*

Dixon, William. 1983. "Measuring interstate affect." *American Journal of Political Science* pp. 828–851.

Erdős, Paul and Alfréd Rényi. 1959. "On random graphs." *Publicationes Mathematicae* 6:290–297.

Frank, Ove. 1971. *Statistical Inference in Graphs*. Stockholm: FOA.

Frank, Ove and David Strauss. 1986. "Markov Graphs." *Journal of the American Statistical Association* 81(395):832–842.

Fuhrmann, Matthew and Todd S. Sechser. 2014. "Signaling Alliance Commitments: Hand-Tying and Sunk Costs in Extended Nuclear Deterrence." *American Journal of Political Science* 58(4):919–935.

Goodreau, Steven M., Mark S. Handcock, Carter T. Hunter, David R .and Butts and Martina Morris. 2008. "A statnet Tutorial." *Journal of Statistical Software* 24(9):1.

Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau and Martina Morris. 2008. "`statnet`: Software tools for the representation, visualization, analysis and simulation of network data." *Journal of Statistical Software* 24(1):1548.

Hunter, David, Mark Handcock, Carter Butts, Steven M Goodreau and Martina Morris. 2008. "`ergm`: A package to fit, simulate and diagnose exponential-family models for networks." *Journal of Statistical Software* 24(3):1–29.

Ingold, Karin. 2008. *Les mécanismes de décision: Le cas de la politique climatique Suisse. Politikanalysen*. Rüegger Verlag, Zürich.

Ingold, Karin and Manuel Fischer. 2014. "Drivers of collaboration to mitigate climate change: An illustration of Swiss climate policy over 15 years." *Global Environmental Change* 24:88–98.

Kinne, Brandon J. 2013. "Network dynamics and the evolution of international cooperation." *American Political Science Review* 107(04):766–785.

Krivitsky, Pavel N. and Mark S. Handcock. 2015. *latentnet: Latent Position and Cluster Models for Statistical Networks*. The Statnet Project (`http://www.statnet.org`). R package version 2.7.1.
URL: *http://CRAN.R-project.org/package=latentnet*

Lemke, Douglas and William Reed. 2001. "War and Rivalry among Great Powers." *American Journal of Political Science* 45(2):457–469.

Li, Heng and Eric Loken. 2002. "A Unified Theory of Statistical Analysis and Inference for Variance Component Models For Dyadic Data." *Statistica Sinica* 12(2):519–535.

Manger, Mark S., Mark A. Pickup and Tom A.B. Snijders. 2012. "A Hierarchy of Preferences: A Longitudinal Network Analysis Approach to PTA Formation." *Journal of Conflict Resolution* 56(5):852–877.

Mansfield, Edward, Helen V. Milner and B. Peter Rosendorff. 2000. "Free to Trade? Democracies, Autocracies, and International Trade Negotiations." *American Political Science Review* 94(2):305–321.

Maoz, Zeev, Ranan D. Kuperman, Lesley Terris and Ilan Talmud. 2006. "Structural Equivalence and International Conflict: A Social Networks Analysis." *Journal of Conflict Resolution* 50(5):664–689.

Minhas, Shahryar, Peter D. Hoff and Michael D. Ward. 2016. "A New Approach to Analyzing Coevolving Longitudinal Networks in International Relations." *Journal of Peace Research* 53(3):491–505.

Mitchell, Sara McLaughlin. 2002. "A Kantian system? Democracy and third-party conflict resolution." *American Journal of Political Science* pp. 749–759.

Morris, Martina, Mark S. Handcock and David R. Hunter. 2008. "Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects." *Journal of Statistical Software* 24(4):1554–7660.

Nowicki, Krzysztof and Tom A.B. Snijders. 2001. "Estimation and Prediction for Stochastic Blockstructures." *Journal of the American Statistical Association* 96(455):1077–1987.

Schweinberger, Michael. 2011. "Instability, sensitivity, and degeneracy of discrete exponential families." *Journal of the American Statistical Association* 106(496):1361–1370.

Snijders, Tom A.B., Philippa E. Pattison, Garry L. Robins and Mark S. Handcock. 2006. "New specifications for exponential random graph models." *Sociological Methodology* 36(1):99–153.

Warner, Rebecca, David Kenny and Michael Stoto. 1979. "A New Round Robin Analysis of Variance for Social Interaction Data." *Journal of Personality and Social Psychology* 37:1742–1757.

Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Wasserman, Stanley and Phillipa Pattison. 1996. "Logit, models and logistic regression for social networks: I. An introduction to Markov graphs and p^*." *Psychometrika* 61:401–425.

Wong, George Y. 1982. "Round robin analysis of variance via maximum likelihood." *Journal of the American Statistical Association* 77(380):714–724.