Inferential Approaches for Network Analysis: AMEN for Latent Factor Models

Shahryar Minhas[a,c,1], **Peter D. Hoff**[b,1,2], **and Michael D. Ward**[a]

[a]Department of Political Science, Michigan State University, East Lansing, MI 48824, USA; [b]Departments of Statistics, Duke University, Durham, NC 27701, USA; [c]Department of Political Science, Duke University, Durham, NC 27701, USA

This manuscript was compiled on June 21, 2017

Many network approaches have been developed in descriptive fashion, but attention to inferential approaches to network analysis has been growing. We introduce a new approach that models interdependencies among observations using additive and multiplicative effects (AME). This approach can be applied to binary, ordinal, and continuous network data, and provides a set of tools for inference from longitudinal networks as well. The AME approach is shown a) to be easy to implement; b) interpretable in a general linear model framework; c) computationally straightforward; d) not prone to degeneracy; e) captures 1st, 2nd, and 3rd order network dependencies; and f) notably outperforms multiple regression quadratic assignment procedures, exponential random graph models, and alternative latent variable approaches on a variety of metrics both in- and out-of-sample.

networks | latent variable models

Network analysis provides a way to represent and study "relational data", that is data which defines characteristics between pairs of actors. Data structures that extend beyond the actor level are common across many fields in the social sciences. In the study of international relations, for instance, the focus often rests on how countries conflict or cooperate with one another. Yet, the dominant paradigm in international relations for dealing with such data structures is not a network approach but rather a dyadic design, in which an interaction between a pair of countries is considered independent of interactions between any other pair in the system. 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.

A common defense of the dyad-only approach is that many events are only bilateral (1), thus alleviating the need for an approach that incorporates interdependencies between observations. The network perspective asserts that even bilateral events and processes take place within a broader system. At a minimum, we do not know whether independence of events and processes characterizes what we observe, thus we should at least examine this assertion.

The potential for interdependence among observations poses a challenge to theoretical as well as statistical modeling since the assumption made by standard approaches used across the social sciences is that observations are, at least, conditionally independent (2). The consequence of ignoring this assumption has been frequently noted already (3–5). Just as relevant is the fact that a wealth of research from other disciplines suggests that carrying the independence assumption into a study with relational data is misguided and leads to biased inferences.

A widely used approach is the exponential random graph model (ERGM), which allows researchers to estimate the role that particular dependencies play in giving rise to an observed network (**? ?**).

A variety of empirical frameworks have been developed to deal with the interdependencies inherent in relational data. A prominent class of approaches involves the use of latent variable models. The most cited latent variable model is the framework presented by Hoff et al. (13), here each actor is assigned a position in a lower-dimensional social space and the Euclidean distance between actors corresponds to their probability of a tie. This approach has received much attention but has two important problems. First, this approach is only able to capture a particular set of dependence patterns that arise in relational data, which substantially limits the class of networks that it can be used to study. Second, due to the construction of the random variables used to characterize the latent space, using this approach as a regression tool complicates parameter interpretation.

A variety of empirical frameworks have been developed to deal with the interdependencies inherent in relational data. In this article, we introduce the additive and multiplicative effects model (AME). To illustrate the contrasts between AME, earlier latent variable models, and other approaches such as ERGM we apply each to studying a cross-sectional network measuring collaborations during the policy design of the Swiss CO_2 act. By doing so we are able to show that AME provides a superior goodness of fit to the data than alternative approaches, both in terms of ability to predict linkages and capture network dependencies.

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.

When modeling relational data, scholars typically employ

Significance Statement

We introduce a statistical model for the evolution of dynamic networks, focusing on linkages that are observed as well as those that are highly likely, but unobserved. The approach we develop is distinctive in that it applies to data that include binary links as well links that may be counts of events, or amounts of trade flows, for example. Unlike the alternative approaches this approach is a) interpreted as a regression, b) captures 1st, 2nd, and 3rd order linkages, c) is easy to compute, and d) out performs alternative statistical approaches.

S.M. designed research and analyzed data; P.H. developed methodological approach; S.M., M.W., and P.H. wrote the paper.

The authors declare no conflict of interests.

[2]To whom correspondence should be addressed. E-mail: s7.minhasgmail.com

Table 1. Structure of datasets used in canonical design.

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

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

	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

a generalized linear model (GLM). An assumption we make when applying this modeling technique is that each of the dyadic observations is conditionally independent. However, this is a strong assumption to make given that events between actors in a network are often interdependent. 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 shown 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.

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.

The presence of these types of interdependencies in relational data complicates the *a priori* assumption of observational independence. Accordingly, inferences drawn from misspecified models that ignore potential interdependencies between dyadic observations are likely to have a number of issues including biased estimates of the effect of independent variables, uncalibrated confidence intervals, and poor predictive performance. By ignoring these interdependencies, we ignore a potentially important part of the data generating process behind relational data.

Social Relations Model: Additive Part of AME

The relevance of modeling first- and second-order dependencies has long been recognized within some social sciences particularly in psychology. Warner et al. developed the social relations model (SRM), a type of ANOVA decomposition technique, that facilitates this undertaking (14). 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 (15, 16):

$$
\begin{aligned}
y_{ij} &= \mu + e_{ij} \\
e_{ij} &= a_i + b_j + \epsilon_{ij} \\
\{(a_1, b_1), \ldots, (a_n, b_n)\} &\overset{\text{iid}}{\sim} N(0, \Sigma_{ab}) \\
\{(\epsilon_{ij}, \epsilon_{ji}) :\ i \neq j\} &\overset{\text{iid}}{\sim} N(0, \Sigma_{\epsilon}),\ \text{where} \\
\Sigma_{ab} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix} &\qquad \Sigma_{\epsilon} = \sigma_{\epsilon}^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}
\end{aligned}
\tag{1}
$$

In the above, μ provides a baseline measure of the density or sparsity 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 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 [receive] a lot of ties also receive [send] 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 1 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 (17). 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.

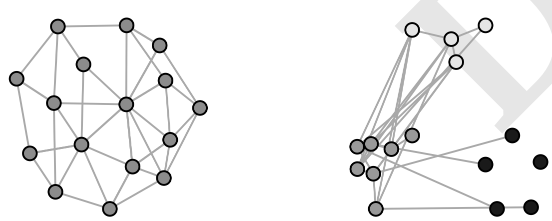
Latent Factor Model: Multiplicative Part of AME

Missing from the framework provided by the SRM is an accounting of third-order dependence patterns that can arise in relational data. The ubiquity of third-order effects in relational datasets arises 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 (18).

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 characteristics 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.

Fig. 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 (20). 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. Additionally, the conditional independence assumption facilitates the testing of a variety of nodal and dyadic level theories, and provides a range of computational advantages.

A number of latent variable approaches have been developed to represent third-order dependencies in relational data, we focus on two here: the latent space model – also known as the latent distance model – and the latent factor model. For the sake of exposition, we consider the case where relations are symmetric to describe the differences between these approaches. Both of these approaches can be incorporated into an undirected version of the framework that we have been constructing through the inclusion of an additional term to the model for y_{ij}, $\alpha(u_i, u_j)$, that captures latent third-order characteristics of a network, where u_i and u_j are node-specific latent variables. General definitions for how $\alpha(u_i, u_j)$ is defined for these latent variable models are shown in Equations 2. One other point of note about these approaches is that researchers have to specify a value for K.

$$
\begin{aligned}
&\text{Latent space model} \\
&\qquad \alpha(\mathbf{u}_i, \mathbf{u}_j) = -|\mathbf{u}_i - \mathbf{u}_j| \\
&\qquad \mathbf{u}_i \in \mathbb{R}^K,\ i \in \{1, \ldots, n\} \\
&\text{Latent factor model} \\
&\qquad \alpha(\mathbf{u}_i, \mathbf{u}_j) = \mathbf{u}_i^\top \Lambda \mathbf{u}_j \\
&\qquad \mathbf{u}_i \in \mathbb{R}^K,\ i \in \{1, \ldots, n\} \\
&\qquad \Lambda \text{ a } K \times K \text{ diagonal matrix}
\end{aligned}
\tag{2}
$$

The latent space model was developed by (13) to capture homophily. In this 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: $-|\mathbf{u}_i - \mathbf{u}_j|$. Hoff et al. show that 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 has been operationalized in the **latentnet** package developed by Krivitsky & Handcock (21). 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 latent space models confound strong ties with stochastic equivalence. This approach cannot adequately model data with many ties between nodes that have different network roles.

An alternative framework is the latent factor model. An early iteration of the latent factor approach was presented in (17). The revised approach is motivated by an eigenvalue decomposition of a network. An important difference in the earlier approaches compared to the model that we present here is that Λ was taken to be the identity matrix thus stochastic equivalence could not be characterized. The motivation for this alternative framework stems from the fact that many real networks exhibit varying degrees of stochastic equivalence and homophily. In these situations, using the latent space model 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 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 $\lambda < 0$ anti-homophily. This approach can represent both homophily and stochastic equivalence, and that the alternative latent variable approaches can be represented as a latent factor model but not vice versa (22). 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, $\mathbf{v}$: $\mathbf{u}_i, \mathbf{v}_j \in \mathbb{R}^K$. These again can alter the probability, or in the continuous case value, of an interaction between ij additively: $\mathbf{u}_i^\top \mathbf{D} \mathbf{v}_j$, where $\mathbf{D}$ is an $K \times K$ 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:

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

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 latent factor model, 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

$$\begin{aligned} \eta_{i,j} &= \beta^\top x_{i,j} + \mathbf{u}_i^\top \Lambda \mathbf{u}_j \\ \bar{\eta} &\approx \beta^\top \bar{x}, \end{aligned}$$

and so the values of β can be interpreted as yielding the "average" value of $\eta_{i,j}$. On the other hand, under the space model

$$\begin{aligned} \eta_{i,j} &= \beta^\top x_{i,j} - |\mathbf{u}_i - \mathbf{u}_j| \\ \bar{\eta} &\approx \beta^\top \bar{x} - \overline{|\mathbf{u}_i - \mathbf{u}_j|} < \beta^\top \bar{x}. \end{aligned}$$

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,

$$\begin{aligned} y_{i,j} &= \beta^\top x_{i,j} + \alpha(\mathbf{u}_i, \mathbf{u}_j) + \epsilon_{i,j} \\ \bar{y} &\approx \beta^\top \bar{x} + \overline{\alpha(\mathbf{u}_i, \mathbf{u}_j)} \\ &= \approx \beta^\top \bar{x}. \end{aligned}$$

Under the space model, $\bar{y} \approx \beta^\top \bar{x} + \overline{|\mathbf{u}_i - \mathbf{u}_j|} < \beta^\top \bar{x}$, 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 latent factor model 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 7:

$$y_{ij} = g(\theta_{ij}) \quad [3]$$

$$\theta_{ij} = \boldsymbol{\beta}^\top \mathbf{X}_{ij} + e_{ij} \quad [4]$$

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

$$\alpha(\mathbf{u}_i, \mathbf{v}_j) = \mathbf{u}_i^\top \mathbf{D} \mathbf{v}_j = \sum_{k \in K} d_k u_{ik} v_{jk} \quad [6]$$

[7]

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 1, $a_i + b_j + \epsilon_{ij}$, 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. A Bayesian procedure in which parameters are iteratively updated using a Gibbs sampler is available in the **amen** package to estimate this type of generalized linear mixed effects model from continuous, binary, ordinal, and other relational data types.

1. ERGMs

An alternative approach to accounting for third-order dependence patterns are ERGMs. ERGM approaches are 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. 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:

$$\Pr(Y = y) = \frac{\exp(\boldsymbol{\beta}^\top S(y))}{\sum_{z\in\mathcal{Y}} \exp(\boldsymbol{\beta}^\top S(z))} \text{ , } y \in \mathcal{Y} \qquad [8]$$

$\boldsymbol{\beta}$ 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 (?). 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.

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. One of the first approaches to deal with this issue was a computationally fast pseudo-likelihood approach developed by Strauss & Ikeda (?). However, this approach ignores the interdependent nature of observations in relational data, as a result, many have argued that the standard errors remain unreliable (?). Additionally, there is no asymptotic theory underlying this approach on which to base the construction of confidence intervals and hypothesis tests (?). The pseudo-likelihood approach has became increasingly unpopular in recent years among those in the network analysis community, particularly, as simulation based techniques have developed–though it has not disappeared. One favored approach in the literature is to approximate the MLE using Markov Chain Monte Carlo techniques, also referred to as MCMC-MLE (? ? ?).

The MCMC-MLE approach is an advancement but notable problems remain. Chatterjee & Diaconis (?) have shown that MCMC procedures can take an exponential time to converge for broad classes of ERGMs unless the dyadic observations are independent. This is a result of the fact that MCMC procedures visit an infinitesimally small portion of the set of possible graphs. A related 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 (?). 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 (? ?). This points to an important caveat in interpreting the implications of an often cited basis for ERGM, the Hammersley-Clifford theorem. Though this theorem ensures that any network can be represented through an ERGM, it says nothing about the complexity of the sufficient statistics ($S(y)$) required to do so. 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 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.

In the following section we undertake a comparison of the latent distance 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 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 (23). 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 (23), and subsequent work by Ingold & Fischer (24), 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 in developing a model specification. 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 SI appendix.

Parameter Estimates. Using the specification described in Table ?? we compare three different modeling approaches. We use an exponential random graph model (ERGM). Next, we use a latent space model (LSM) with a two-dimensional Euclidean distance metric. Last, we use the AME, in which we account for nodal and dyadic heterogeneity using the SRM and third-order effects represented by a latent factor approach with $K = 2$. In the SI Appendix, we show that the parameter estimates presented here for the AME model remain very similar no matter the K chosen.

Most relevant for us are how parameter estimates from AME relate to other approaches. 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 not only similar to those approaches but in line with the theoretical expectations of Ingold & Fischer. Similar discrepancies between LSM and other approaches appear for parameters such as `Influence attribution` and `Alter's influence degree`. Each of these discrepancies are resolved when using AME. In part, this is a result of how the LSM approach complicates the interpretation of the effect of exogenous variables. In the SI Appendix, we show that these differences persist even when incorporating sender and receiver random effects or when switching to a bilinear approach to handle third-order dependencies.

Table 3. * p $<$ 0.05. 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 notable 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 stated earlier. 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. How do these approaches fit the data out-of-sample? We utilize a cross-validation procedure to assess the out-of-sample performance for each of the models presented in Table 3 as follows:

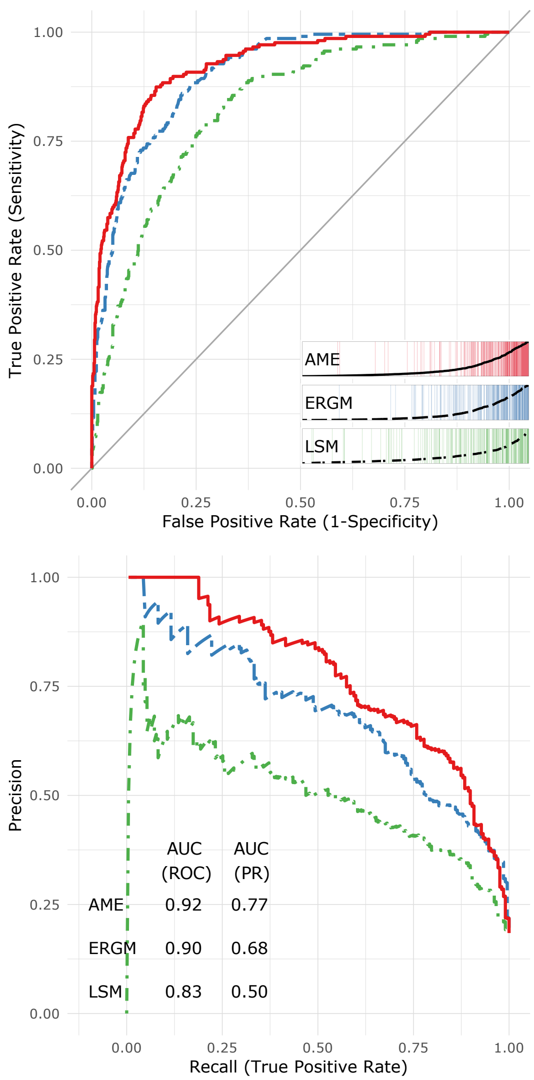
- Randomly divide the $n \times (n-1)$ data points into S sets of roughly equal size, letting s_{ij} be the set to which pair $\{ij\}$ is assigned.
- For each $s \in \{1, \ldots, S\}$:
  - Obtain estimates of the model parameters conditional on $\{y_{ij} : s_{ij} \neq s\}$, the data on pairs not in set 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 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 levels of missingness.

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. First, is a diagnostic that is common in the social sciences. 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. In the SI appendix, we provide additional comparisons between our AME approach and various parameterizations of the LSM, in each case we find that the AME approach provides far superior results in terms of out-of-sample predictive performance.

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 we can again see that the AME and ERGM models performs better than the alternatives.

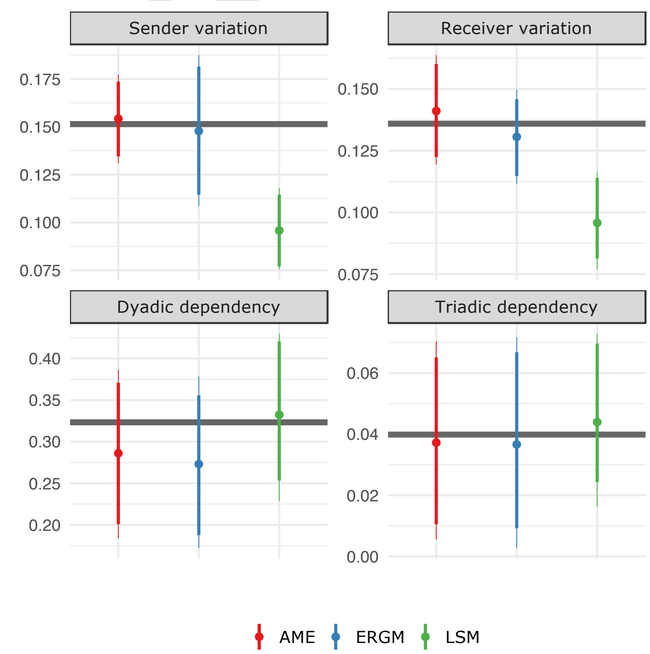
Fig. 2. Assessments of out-of-sample predictive performance using ROC curves, separation plots, and PR curves. AUC statistics are provided as well for both curves.



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 (?). 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 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 one can compare the observed network with a set of networks simulated from the estimated models. We restrict our focus to the three approaches–LSM, ERGM, and AME–that explicitly seek to model network interdependencies. 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.

Fig. 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. 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. 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.

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 latent space models in out-of-sample prediction, and also is better able to capture network dependencies than the latent space model.

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. It is also more straightforward to interpret since it does not require interpretation of unusual features such as *three-stars* which fall outside of the normal language for discussing social science. Further, the **amen** package facilitates the modeling of longitudinal network data. In sum, excuses for continuing to treat relational data as conditionally independent are no longer valid.

ACKNOWLEDGMENTS. This research was partially supported by the National Science Foundation Award 1259266.

1. Diehl PF, Wright TM (2016) A conditional defense of the dyadic approach. *International Studies Quarterly*.
2. Snijders TA (2011) Statistical models for social networks. *Annual Review of Sociology* 37:131–53.
3. Beck N, Katz JN, Tucker R (1998) Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science* 42(2):1260–1288.
4. Signorino C (1999) Strategic interaction and the statistical analysis of international conflict. *American Political Science Review* 92(2):279–298.
5. Aronow PM, Samii C, Assenova VA (2015) Cluster–robust variance estimation for dyadic data. *Political Analysis* 23(4):564–577.
6. Bonabeau E (2002) Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences* 99(suppl 3):7280–7287.
7. Brandes U, Erlebach T (2005) *Network Analysis: Methodological Foundations*. (Springer Science & Business Media) Vol. 3418.
8. Goyal S (2012) *Connections: an introduction to the economics of networks*. (Princeton University Press).
9. Jackson M (2014) Networks in the understanding of economic behaviors. *The Journal of Economic Perspectives* 28(4):3–22.
10. Pattison P, Wasserman S (1999) Logit models and logistic regressions for social networks. ii. multivariate relations. *British Journal of Mathematical and Statistical Psychology* 52:169–194.
11. Kenny DA, Kashy DA, Cook WL (2006) *Dyadic Data Analysis*. (Guilford Press, New York).
12. Snijders TA (1996) Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology* 21(1-2):149–172.
13. Hoff PD, Raftery AE, Handcock MS (2002) Latent space approaches to social network analysis. *Journal of the American Statistical Association* 97(460):1090–1098.
14. Warner R, Kenny D, Stoto M (1979) A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology* 37:1742–1757.
15. Wong GY (1982) Round robin analysis of variance via maximum likelihood. *Journal of the American Statistical Association* 77(380):714–724.
16. Li H, Loken E (2002) A unified theory of statistical analysis and inference for variance component models for dyadic data. *Statistica Sinica* 12(2):519–535.
17. Hoff PD (2005) Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association* 100(4690):286–295.
18. Wasserman S, Faust K (1994) *Social Network Analysis: Methods and Applications*. (Cambridge University Press, Cambridge).
19. Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research* 40(2):211–239.
20. Manger MS, Pickup MA, Snijders TA (2012) A hierarchy of preferences: A longitudinal network analysis approach to PTA formation. *Journal of Conflict Resolution* 56(5):852–877.
21. Krivitsky PN, Handcock MS (2015) *latentnet: Latent Position and Cluster Models for Statistical Networks* (The Statnet Project (http://www.statnet.org)). R package version 2.7.1.
22. Hoff PD (2008) Modeling homophily and stochastic equivalence in symmetric relational data in *Advances in Neural Information Processing Systems 20*, Processing Systems 21, eds. Platt JC, Koller D, Singer Y, Roweis ST. (MIT Press, Cambridge, MA, USA), pp. 657–664.
23. Ingold K (2008) *Les mécanismes de décision: Le cas de la politique climatique Suisse. Politikanalysen.* (Rüegger Verlag, Zürich).
24. Ingold K, Fischer M (2014) Drivers of collaboration to mitigate climate change: An illustration of swiss climate policy over 15 years. *Global Environmental Change* 24:88–98.