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

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

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

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 modeling framework and is thus flexible enough to be applied to a variety of 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). Relative to these approaches, we show that the AME approach is 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 ERGMs and LSMs on a variety of metrics and in an out-of-sample context. In summary, AME offers a straightforward way to undertake nuanced, principled inferential network analysis for a wide range of social science questions.

[*]Corresponding author
Email address: `minhassh@msu.edu` (Shahryar Minhas)

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, even bilateral events and processes take place within a broader system, and occurrences in one part of the system may be dependent upon events in another. At a minimum, we do not know whether independence of events and pro-

---

[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.

cesses characterizes what we observe.

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 article is organized as follows, we begin by briefly discussing the difficulties in studying dyadic data through approaches that assume observational independence. 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 while still enabling researchers to study exogenous covariates of interest. We then briefly contrast these latent variables models with ERGM and conclude with an application on 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: $\mathcal{L}(\boldsymbol{\theta}|\mathbf{Y}) = \prod_{\alpha=1}^{n \times (n-1)} P(y_\alpha|\theta_\alpha)$.

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.[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". 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.

---

[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.

Additive Part of AME

The dependencies that tend to develop in relational data can be more easily under-
stood 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 het-
erogeneity in activity levels across nodes. The implication of this across-node het-
erogeneity 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 ma-
trix 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 net-
work 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 net-
work 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 interde-
pendency 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.

We model first- and second-order dependencies in AME using a set of additive effects that are motivated by the social relations model (SRM) developed by (Warner, Kenny and Stoto, 1979; Li and Loken, 2002). Specifically, we 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:

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

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

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

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

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

(2)

μ here 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 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.

We incorporate the covariance structure described in Equation 2 into the systematic

component of a GLM framework: $\boldsymbol{\beta}^\top \mathbf{X}_{ij} + a_i + b_j + \epsilon_{ij}$, where $\boldsymbol{\beta}^\top \mathbf{X}_{ij}$ accommodates the inclusion of dyadic, sender, and receiver covariates. This 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.

Multiplicative Part of AME

Missing from the additive effects portion of the model 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.[4]

For example, 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. 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.

---

[4]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

Figure 1: Graph on the left is a representation of an undirected network that exhibits a high degree of homophily (linkages forming because of shared attributes), while on the right we show an undirected network that exhibits stochastic equivalence.
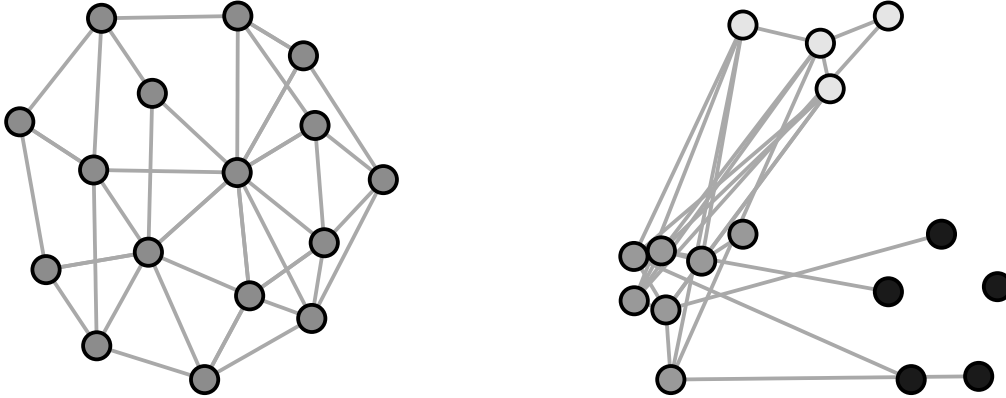
Another third-order dependence pattern that cannot be accounted for in the additive effects framework is stochastic equivalence. A pair of actors ij are stochastically equivalent if the probability of i relating to, and being related to, by every other actor is the same as the probability for j. This 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

(Wasserman and Faust 1994).

dynamics, and failing to account for such dynamics leaves potentially important parts of the data generating process ignored.

We account for third order dependence patterns using a latent variable framework, and our goal in doing so is twofold: 1) be able to adequately represent third order dependence patterns, 2) improve our ability to conduct inference on exogenous covariates. Latent variable models assume that relationships between nodes are mediated by a small number (K) of node-specific unobserved latent variables. We contrast the approach that we utilize within AME, the latent factor model (LFM), to the latent space model, which is among the most widely used in the networks literature.[5] 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 3:

[5]An alternative approach with a similar latent variable formulation is known as the stochastic block model (Nowicki and Snijders, 2001), however, this approach is typically only used to model community structure in networks and not used to conduct inference on exogenous covariates.

Latent space model

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

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

Latent factor model   (3)

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

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

$$\Lambda \text{ a } K \times K \text{ diagonal matrix}$$

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 struc-

ture.

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.[6]

In addition to summarizing dependence patterns in networks, scholars are often concerned with accounting for interdependencies so that they can better estimate the effects of exogenous covariates. Both the latent space and factor models attempt to do this as they 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 a link function:

---

[6]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.

$$p(y_{i,j}|x_{i,j}, u_i, u_j) = f(y_{i,j} : \eta_{i,j})$$

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

However, 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:

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

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

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 4:

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

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

$$e_{ij} = a_i + b_j + \epsilon_{ij} + \alpha(\mathbf{u}_i, \mathbf{v}_j) \text{ , where} \tag{4}$$

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

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 2, $a_i + b_j + \epsilon_{ij}$, 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.[7]

ERGMs

An alternative approach to accounting for third-order dependence patterns are ERGMs. Whereas AME seeks to estimate interdependencies in a network through a set of latent variables, ERGM approaches are useful when researchers are interested in the role that a specific network statistic(s) has in giving rise to an observed network. These network statistics could include the number of transitive triads in a network, balanced triads, reciprocal pairs and so on.[8] 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:

$$\Pr(Y = y) = \frac{\exp(\boldsymbol{\beta}^T S(y))}{\sum_{z \in \mathcal{Y}} \exp(\boldsymbol{\beta}^T S(z))} \text{ , } y \in \mathcal{Y} \tag{5}$$

$\boldsymbol{\beta}$ 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). 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. Additionally, researchers can easily accommodate nodal and

---

[7]The MCMC algorithm describing the estimation procedure is available in the Appendix.

[8]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.

dyadic covariates. Further because of the Hammersley-Clifford theorem any probability distribution over networks can be represented by the form shown in Equation 5.

A notable issue when estimating ERGMs, however, is that the estimated model can become degenerate. Degeneracy here means that the model places 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, $\mathbf{Y}$ (Schweinberger, 2011).[9] Some have argued that model degeneracy is simply a result of model misspecification (Handcock, 2003; Goodreau et al., 2008; Handcock et al., 2008). However, this points to an important caveat in interpreting the implications of 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 using an application presented in Cranmer et al. (2016).[10] In

---

[9]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.

[10]The reason we use the same dataset is because of the model specification issue that arises when using ERGMs. As Cranmer et al. (2016, p. 8) note, when using ERGMs scholars must model third-order effects and "must also specify them in a complete and correct manner" or the model will be misspecified. Thus to avoid providing an incorrect specification when comparing ERGM with the AME we use the

doing so, we are able to compare and contrast these various approaches.

Empirical Comparison

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 (2008), and subsequent work by 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. (2016) in developing a model specification to understand and predict link formation in this network.[11]

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 includes the same exogenous covariates and accounts for nodal and

specification that they constructed.

[11]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.

dyadic heterogeneity using the SRM.[12] Third-order effects are represented by the latent factor model with $K = 2$. Last, we also include a logistic model as that is still the standard in most of the field. 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 while similar to those of ERGM are 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 due to the construction of the latent variable term.[13]

---

[12]Convergence diagnostics for AME are provided in the Appendix.

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

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

	Logit	LSM	ERGM	AME
Intercept/Edges	-4.44	0.95	-12.17	-3.40
	(0.34)	[0.09; 1.85]	(1.40)	[-4.40; -2.51]
Conflicting policy preferences				
Business vs. NGO	-0.86	-1.37	-1.11	-1.38
	(0.46)	[-2.39; -0.40]	(0.51)	[-2.47; -0.49]
Opposition/alliance	1.21	0.00	1.22	1.08
	(0.20)	[-0.40; 0.40]	(0.20)	[0.72; 1.49]
Preference dissimilarity	-0.07	-1.77	-0.44	-0.79
	(0.37)	[-2.64; -0.91]	(0.39)	[-1.55; -0.07]
Transaction costs				
Joint forum participation	0.88	1.51	0.90	0.92
	(0.27)	[0.85; 2.17]	(0.28)	[0.40; 1.46]
Influence				
Influence attribution	1.20	0.08	1.00	1.10
	(0.22)	[-0.40; 0.54]	(0.21)	[0.70; 1.55]
Alter's influence indegree	0.10	0.01	0.21	0.11
	(0.02)	[-0.03; 0.04]	(0.04)	[0.07; 0.15]
Influence absolute diff.	-0.03	0.04	-0.05	-0.07
	(0.02)	[-0.01; 0.09]	(0.01)	[-0.11; -0.03]
Alter = Government actor	0.63	-0.46	1.04	0.56
	(0.25)	[-1.08; 0.14]	(0.34)	[-0.06; 1.16]
Functional requirements				
Ego = Environmental NGO	0.88	-0.60	0.79	0.68
	(0.26)	[-1.30; 0.08]	(0.17)	[-0.36; 1.73]
Same actor type	0.74	1.17	0.99	1.03
	(0.22)	[0.62; 1.72]	(0.23)	[0.62; 1.48]
Endogenous dependencies				
Mutuality	1.22		0.81	
	(0.21)		(0.25)	
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.[14] 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

To test which model more accurately captures the data generating process for this network, we utilize a cross-validation procedure to assess the out-of-sample performance for each of the models presented in Table 3. Our cross-validation approach proceeds 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

---

[14]See the Appendix for details.

to 45 which corresponds to randomly excluding approximately 2% of the data from the estimation.[15] 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 four 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 Logit and LSM approach lag 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 perform better than the alternatives.

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

---

[15]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 is an example of the computational difficulties associated with ERGMs.

Figure 2: Assessments of out-of-sample predictive performance using ROC curves, separation plots, and PR curves. AUC statistics are also provided.
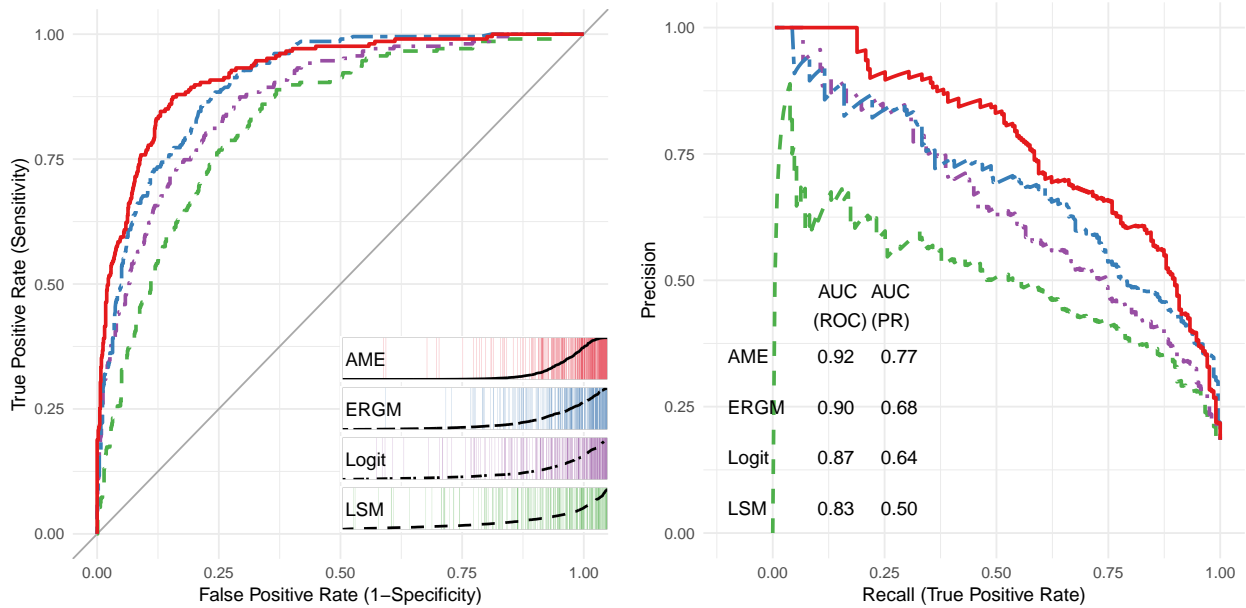
difference, however, is that in ROC space we utilize the FPR, while in PR space we use precision. FPR measures the fraction of negative examples that are misclassified as positive, while precision measures the fraction of examples classified as positive that are truly positive. PR curves are useful in 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 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.[16]

Capturing Network Attributes

We also assess which of these models best captures the network features of the dependent variable.[17] To do this, we compare the observed network with a set of networks simulated from the estimated models.[18] 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.

---

[16]We also test AME against the multiple regression quadratic assignment procedure (MRQAP). This approach also perform notably worse than AME in terms of predicting tie formation. Results are available upon request.

[17]We restrict our focus to the three approaches–LSM, ERGM, and AME–that explicitly seek to model network interdependencies.

[18]In the Appendix, we compare the ability of these models to capture network attribute across a wider array of statistics (e.g., dyad-wise shared partners, incoming k-star, etc.), and the results are consistent with what we present below.

Figure 3: Network goodness of fit summary using **amen**.
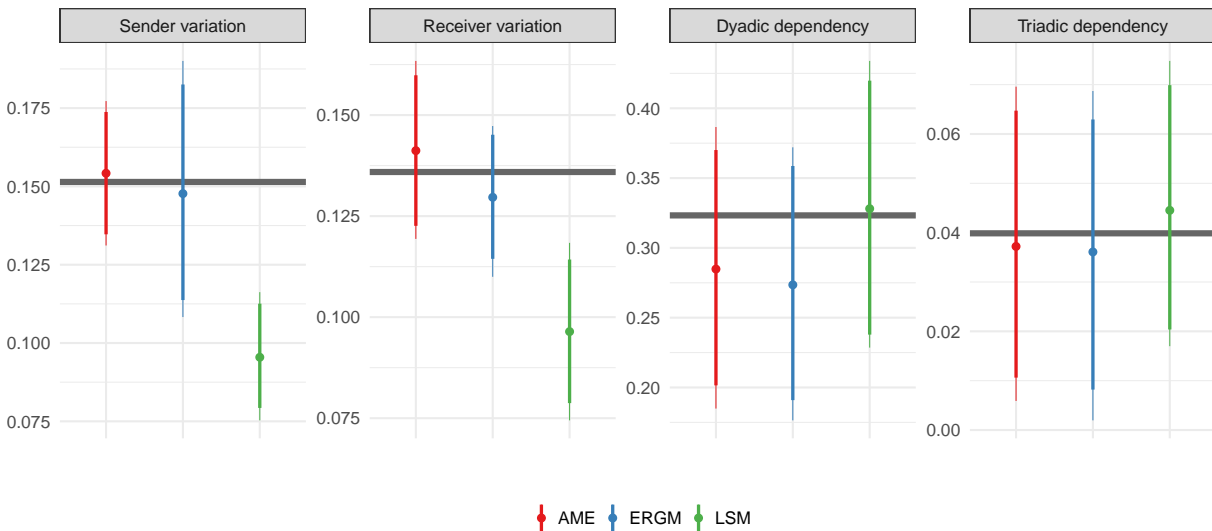
Here it becomes quickly apparent that the LSM model fails to capture how active and popular actors are in the Swiss climate change mitigation network.[19] The AME and ERGM specifications again both tend to do equally well. If when running this diagnostic, we found that the AME model did not adequately 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.

Conclusion

The AME approach to estimation and inference in network data provides a number of benefits over extant alternatives in political science. Specifically, it provides a

---

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

modeling framework for dyadic data that is based on familiar statistical tools such as linear regression, GLM, random effects, and factor models.[20] Further we have shown that alternatives such as the LSM complicate parameter interpretation due to the construction of the latent variable term. The benefit of AME is that its focus intersects with the interest of most IR scholars, which is primarily on the effects of exogenous covariates. 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.

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. 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.[21] It is unlikely that this system can be viewed simply as a collection of isolated

---

[20]A number of related approaches have been developed that also stem from latent variable models: Sewell and Chen (2015); Gollini and Murphy (2016); Durante, Dunson and Vogelstein (2017); Kao, Smith and Airoldi (2018). Each of these approaches differ in how they construct the latent variable term to account for third-order dependencies, but they each are based off of a similar framework as the model we present here. We hope that this paper motivates further interest in exploring the utility of latent variable models to studying networks in political science.

[21]Additionally, in most political science applications, we are interested in how actors behave towards each other over time. Accounting for repeated interaction within AME can be done by including time-dependent regression terms such as lags of the dependent variable or simply time-varying regression

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.

---

parameters.

Appendix

Additive and Multiplicative Effects Gibbs Sampler

To estimate, the effects of our exogenous variables and latent attributes we utilize a Bayesian probit model in which we sample from the posterior distribution of the full conditionals until convergence. Specifically, given observed data **Y** and **X** – where **X** is a design array that includes our sender, receiver, and dyadic covariates – we estimate our network of binary ties using a probit framework where: $y_{ij,t} = 1(\theta_{ij,t} > 0)$ and $\theta_{ij,t} = \boldsymbol{\beta}^\top \mathbf{X}_{ij,t} + a_i + b_j + \mathbf{u}_i^\top \mathbf{D} \mathbf{v}_j + \epsilon_{ij}$. The derivation of the full conditionals is described in detail in Hoff (2005) and Hoff (2008), thus here we only outline the Markov chain Monte Carlo (MCMC) algorithm for the AME model that we utilize in this paper.

- Given initial values of $\{\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2\}$, the algorithm proceeds as follows:

    - sample $\boldsymbol{\theta} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2$ (Normal)

    - sample $\boldsymbol{\beta} \mid \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2$ (Normal)

    - sample $\mathbf{a}, \mathbf{b} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2$ (Normal)

    - sample $\Sigma_{ab} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \rho, \text{ and } \sigma_\epsilon^2$ (Inverse-Wishart)

    - update ρ using a Metropolis-Hastings step with proposal $p^*|p \sim$ truncated normal$_{[-1,1]}(\rho, \sigma_\epsilon^2)$

    - sample $\sigma_\epsilon^2 \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \text{ and } \rho$ (Inverse-Gamma)

    - For each $k \in K$:

        * Sample $\mathbf{U}_{[,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}_{[,-k]}, \mathbf{V}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2$ (Normal)

        * Sample $\mathbf{V}_{[,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}_{[,-k]}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2$ (Normal)

        * Sample $\mathbf{D}_{[k,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho, \text{ and } \sigma_\epsilon^2$ (Normal)[22]

---

[22]Subsequent to estimation, **D** matrix is absorbed into the calculation for **V** as we iterate through K.

*Ingold & Fischer Model Specification and Expected Effects*

Variable	Description	Expected Effect
Conflicting policy preferences		
Business v. NGO	Binary, dyadic covariate that equals one when one actor is from the business sector and the other an NGO.	$-$
Opposition/alliance	Binary, dyadic covariate that equals one when i, sender, perceives j, receiver, as having similar policy objectives regarding climate change.	$+$
Preference dissimilarity	Transformation of four core beliefs into a Manhattan distance matrix, smaller the distance the closer the beliefs of i and j.	$-$
Transaction costs		
Joint forum participation	Binary, dyadic covariate that equals one when i and j belong to the same policy forum.	$+$
Influence		
Influence attribution	Binary, dyadic covariate that equals one when i considers j to be influential.	$+$
Alter's influence in-degree	Number of actors that mention i as being influential, this is a measure of reputational power.	$+$
Influence absolute diff.	Absolute difference in reputational power between i and j.	$-$
Alter = Government Actor	Binary, nodal covariate that equals one when j is a state actor.	$+$
Functional requirements		
Ego = Environment NGO	Binary, nodal covariate that equals one when i is an NGO.	$+$
Same actor type	Binary, dyadic covariate that equals when i and j are the same actor type.	$+$
Endogenous dependencies: ERGM Specific Parameters		
Mutuality	Captures concept of reciprocity, if i indicates they collaborated with j then j likely collaborates with i.	$+$
Outdegree popularity	Captures idea that actors sending more ties will be more popular targets themselves for collaboration.	$+$
Twopaths	Counts the number of two-paths in the network, two-path is an instance where i is connected to j, j to k, but i is not connected to k.	$-$
GWIdegree (2.0)	Takes into account how many ties a node sends in the network, used to capture network structures that result from some highly active nodes.	$+$
GWESP (1.0)	Counts the number of shared partners for each pair and sums across.	$+$
GWOdegree (0.5)	Takes into account how many ties a node receives in the network, used to capture networks structures that result from some highly popular nodes.	$+$

Table A.1: Summary of variables to be included in model specification.

AME Model Convergence

Trace plot for AME model presented in paper.
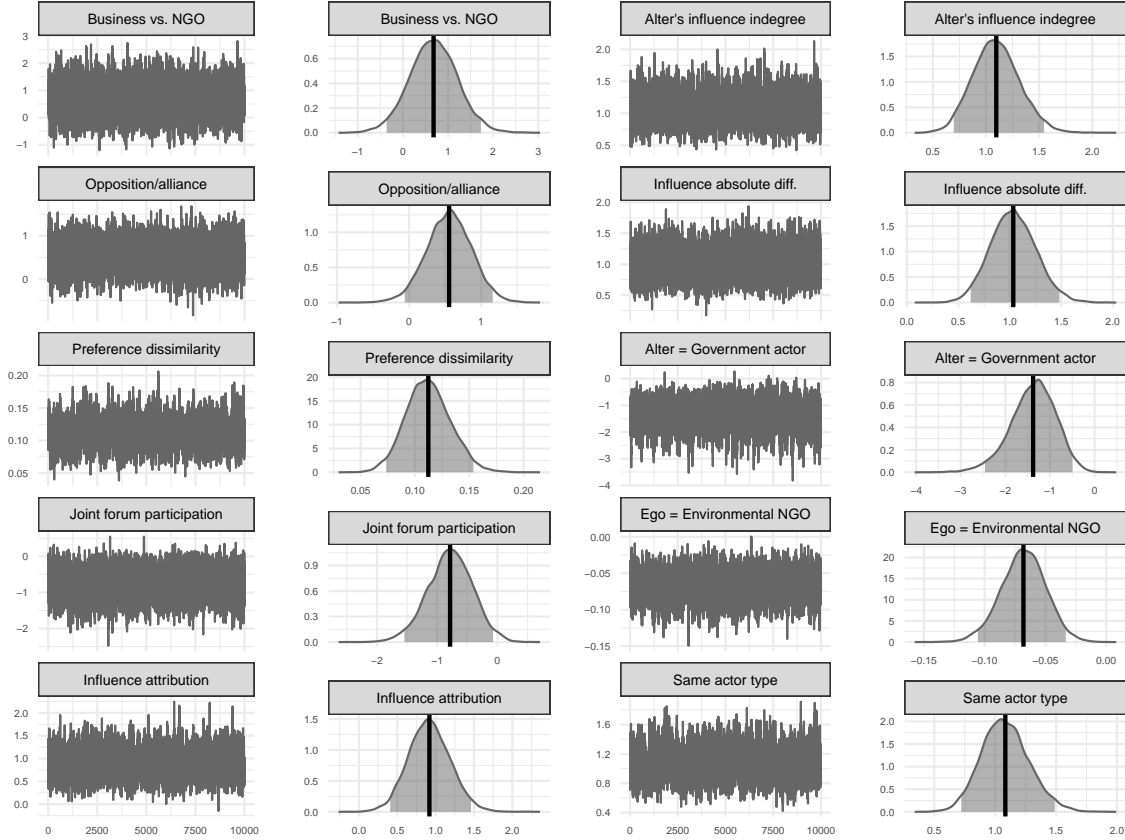
Figure A1: Trace plot for AME model presented in paper. In this model, we utilize the SRM to account for first and second-order dependence. To account for third order dependencies we use the latent factor approach with $K = 2$.

Multiplicative Effects Visualization

When it comes to estimating higher-order effects, ERGM is able to provide explicit estimates of a variety of higher-order parameters, however, this comes with the caveat that these are the "right" set of endogenous dependencies. The AME approach, as shown in Equation 4 of the manuscript, estimates network dependencies by examining patterns left over after taking into account the observed covariates. For the sake of space, we focus on examining the third-order dependencies left over after accounting for the observed covariates and network covariance structure modeled by the SRM. A visualization of remaining third-order dependencies is shown in Figure A2.
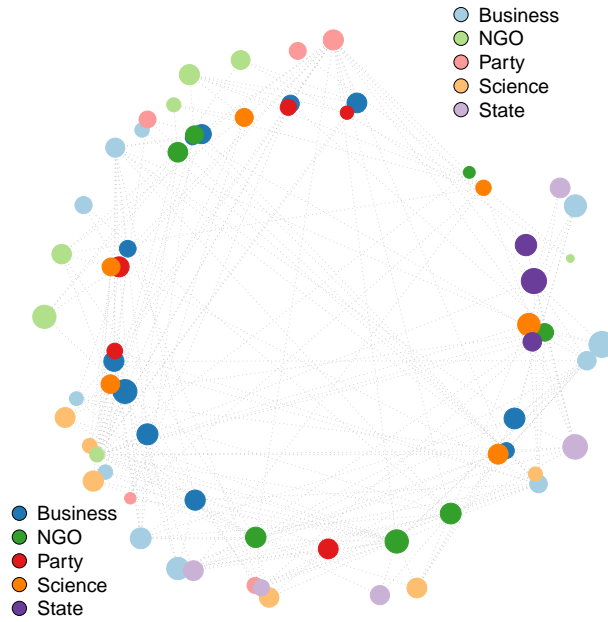
Figure A2: Circle plot of estimated latent factors.

In Figure A2, the directions of $\hat{u}_i$'s and $\hat{v}_i$'s are noted in lighter and darker shades, respectively, of an actor's type.[23] The size of actors is a function of the magnitude of the vectors, and dashed lines between actors indicate greater than expected levels of

---

[23]For example, actors from industry and business are assigned a color of blue and the direction of $\hat{u}_i$ for these actors is shown in light blue and $\hat{v}_i$ in dark blue

collaboration based on the regression term and additive effects. In the case of the application dataset that we are using here organization names have been anonymized and no additional covariate information is available. However, if we were to observe nodes sharing certain attributes clustering together in this circle plot that would mean such an attribute could be an important factor in helping us to understand collaborations among actors in this network. Given how actors of different types are distributed in almost a random fashion in this plot, we can at least be sure that it is unlikely other third-order patterns can be picked up by that factor.

Other Network Goodness of Fit Tests

Below we show a standard set of statistics upon which comparisons are usually conducted:[24]

Variable	Description
Dyad-wise shared partners	Number of dyads in the network with exactly i shared partners.
Edge-wise shared partners	Similar to above except this counts the number of dyads with the same number of edges.
Geodesic distances	The proportion of pairs of nodes whose shortest connecting path is of length k, for $k = 1, 2, \ldots$. Also, pairs of nodes that are not connected are classified as $k = \infty$.
Incoming k-star	Propensities for individuals to have connections with multiple network partners.
Indegree	Proportion of nodes with the same value of the attribute as the receiving node.
Outdegree	Proportion of nodes with the same value of the attribute as the sending node.

Table A.2: Description of a set of standard statistics used to assess whether a model captures network dependencies.

We simulate 1,000 networks from the LSM, ERGM, and AME model and compare how well they align with the observed network in terms of the statistics described in Table A.2. The results are shown in Figure A3. Values for the observed network are indicated by a gray bar and average values from the simulated networks for the AME, ERGM, and LSM are represented by a diamond, triangle, and square, respectively. The densely shaded interval around each point represents the 95% interval from the simulations and the taller, less dense the 90% interval.[25] Looking across the panels in Figure A3 it is clear that there is little difference between the ERGM and AME models in terms of how well they capture network dependencies. The LSM model, however, does perform somewhat worse in comparison here as well. Particularly, when it comes to assessing the number of edge-wise shared partners and in terms of capturing the indegree and outdegree distributions of the collaboration network.

---

[24]See Morris, Handcock and Hunter (2008) for details on each of these parameters. If one was to examine goodness of fit in the **ergm** package these parameters would be calculated by default.

[25]Calculation for the incoming k-star statistic is not currently supported by the **latentnet** package.
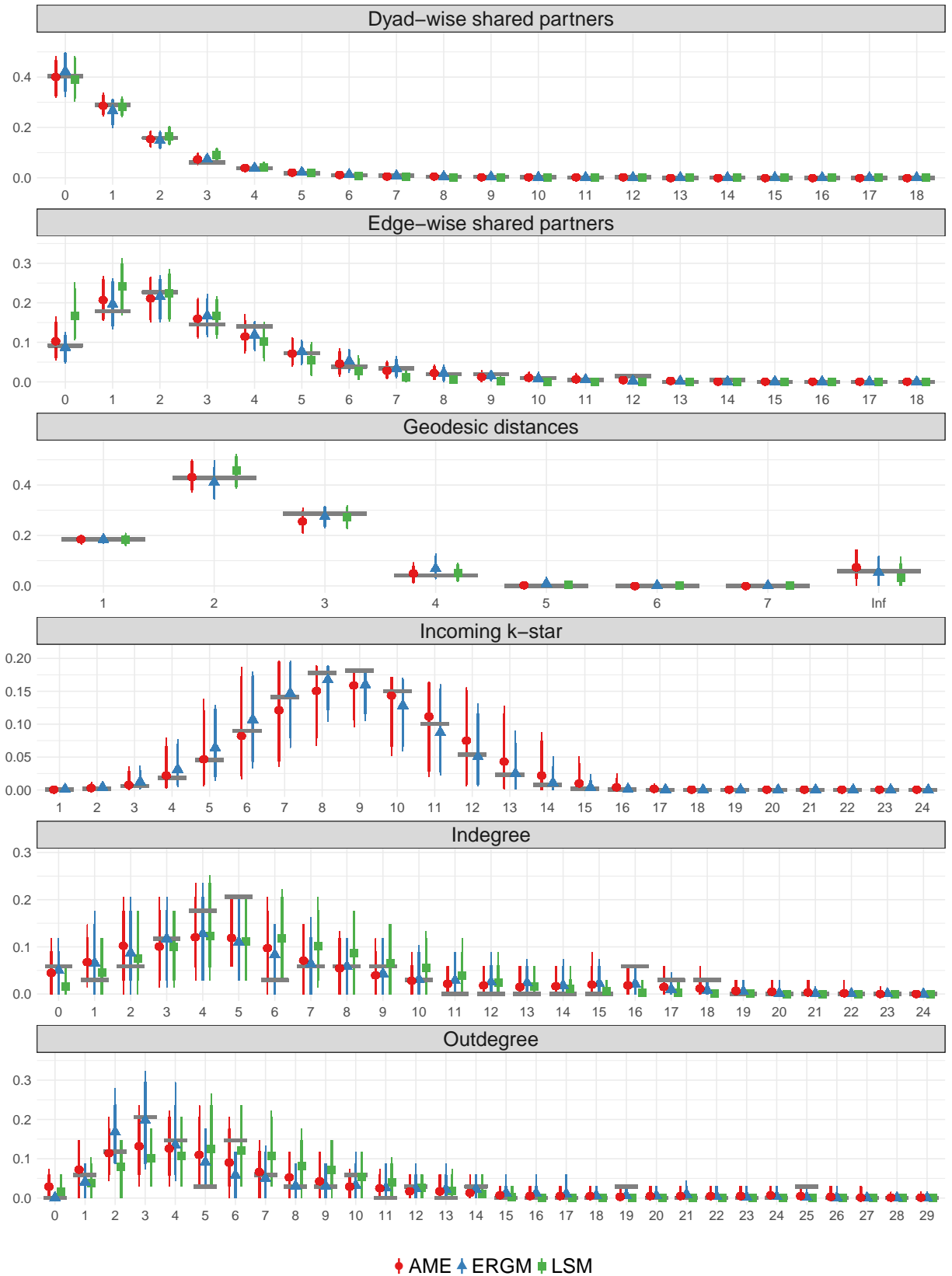
Figure A3: Goodness of fit statistics to assess how well the LSM, ERGM, and AME approaches account for network dependencies. Grey bars indicate true values.

33

Comparison with other AME Parameterizations

Here we provide a comparison of the AME model we present in the paper that uses $K = 2$ for multiplicative effects and show how results change when we use $K = \{1, 3, 4\}$. Trace plots for $K = \{1, 3, 4\}$ are available upon request.

	AME (k=1)	AME (k=2)	AME (k=3)	AME (k=4)
Intercept/Edges	-3.08	-3.40	-3.74	-3.92
	[-3.91; -2.30]	[-4.40; -2.51]	[-4.80; -2.75]	[-5.13; -2.87]
Conflicting policy preferences				
Business vs. NGO	-1.28	-1.38	-1.51	-1.50
	[-2.20; -0.46]	[-2.47; -0.49]	[-2.64; -0.53]	[-2.69; -0.52]
Opposition/alliance	0.95	1.08	1.19	1.27
	[0.65; 1.28]	[0.72; 1.49]	[0.80; 1.63]	[0.84; 1.76]
Preference dissimilarity	-0.65	-0.79	-0.92	-0.96
	[-1.29; -0.02]	[-1.55; -0.07]	[-1.76; -0.12]	[-1.80; -0.15]
Transaction costs				
Joint forum participation	0.84	0.92	1.02	1.05
	[0.37; 1.31]	[0.40; 1.46]	[0.46; 1.62]	[0.43; 1.73]
Influence				
Influence attribution	1.00	1.10	1.21	1.27
	[0.64; 1.39]	[0.70; 1.55]	[0.77; 1.70]	[0.80; 1.83]
Alter's influence indegree	0.10	0.11	0.13	0.13
	[0.07; 0.14]	[0.07; 0.15]	[0.08; 0.17]	[0.09; 0.18]
Influence absolute diff.	-0.06	-0.07	-0.08	-0.08
	[-0.10; -0.03]	[-0.11; -0.03]	[-0.13; -0.04]	[-0.12; -0.04]
Alter = Government actor	0.52	0.56	0.66	0.63
	[-0.04; 1.07]	[-0.06; 1.16]	[-0.04; 1.46]	[-0.08; 1.36]
Functional requirements				
Ego = Environmental NGO	0.61	0.68	0.77	0.79
	[-0.31; 1.56]	[-0.36; 1.73]	[-0.35; 1.93]	[-0.38; 2.04]
Same actor type	0.97	1.03	1.14	1.18
	[0.60; 1.36]	[0.62; 1.48]	[0.66; 1.66]	[0.70; 1.70]

Table A.3: 95% posterior credible intervals are provided in brackets.
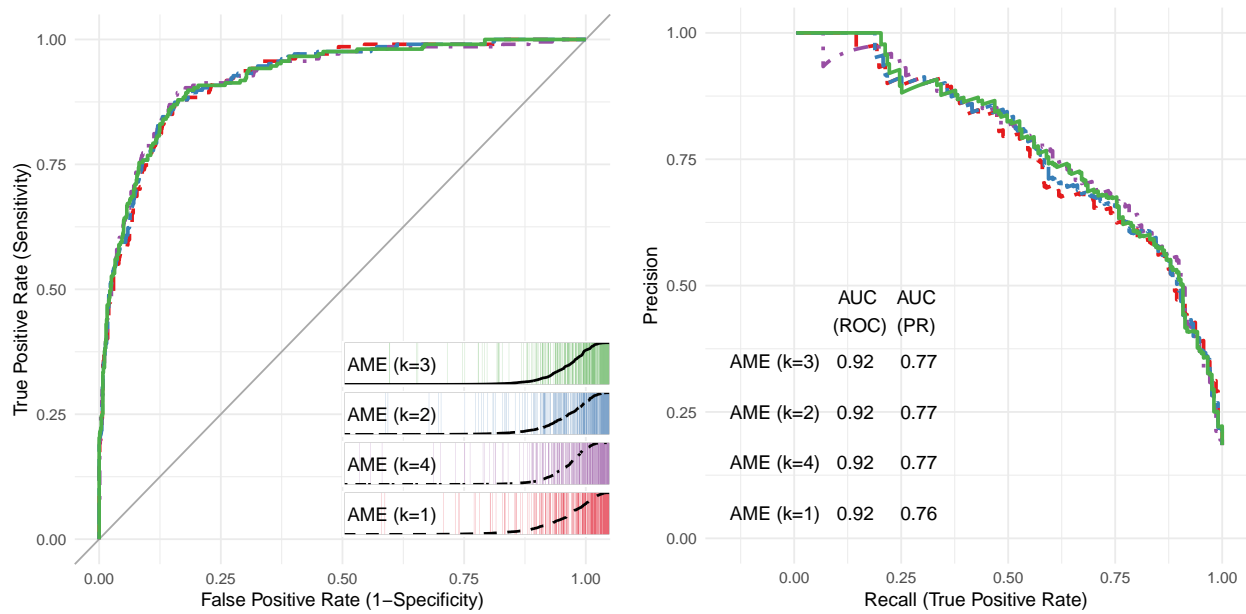
Figure A4: Assessments of out-of-sample predictive performance using ROC curves, separation plots, and PR curves. AUC statistics are provided as well for both curves.
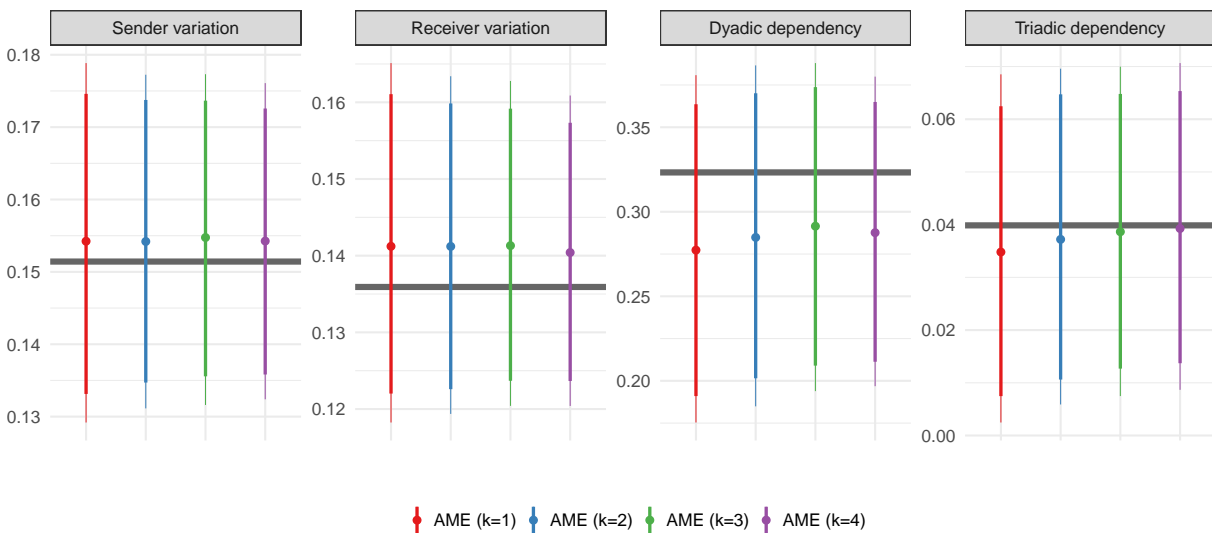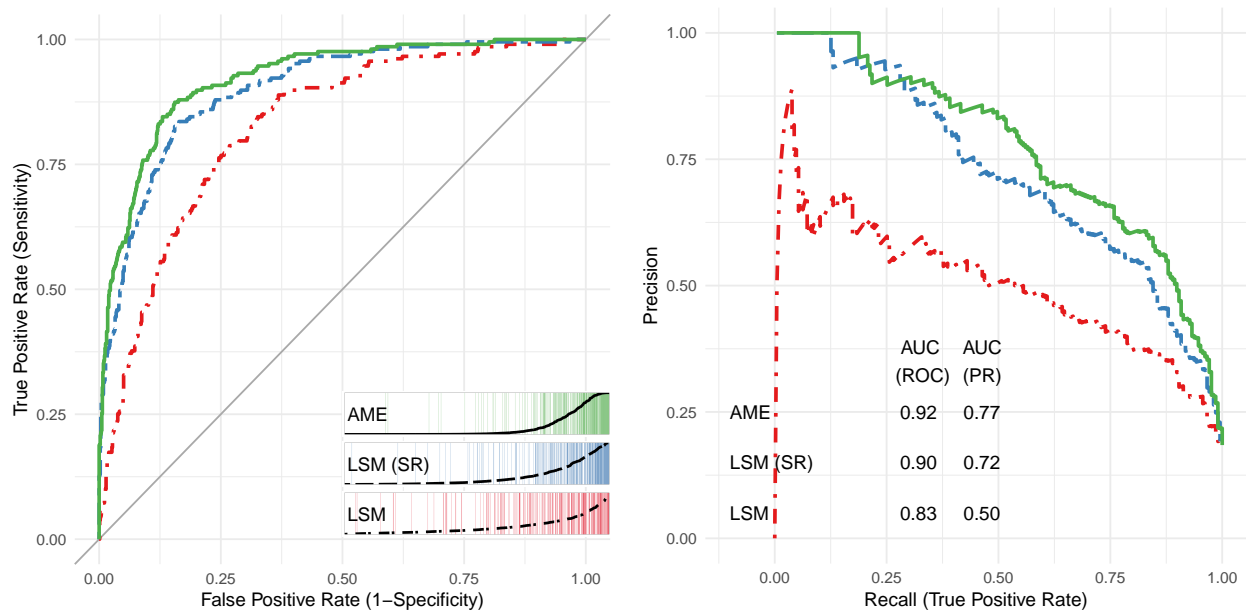

Figure A5: Network goodness of fit summary using **amen**.

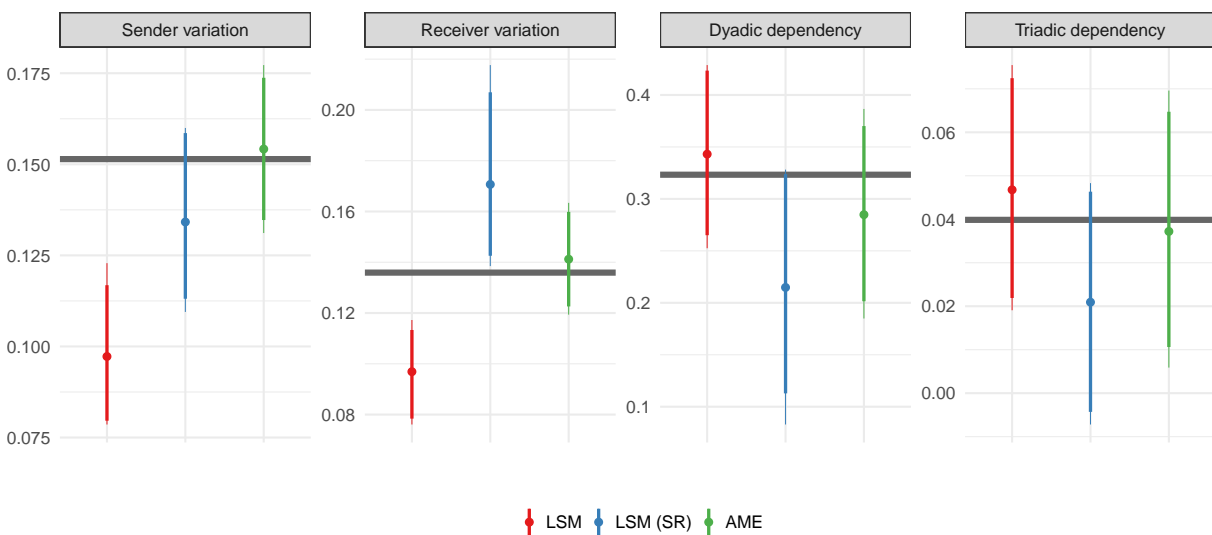*Comparison of **amen** & **latentnet** R Packages*

Here we provide a comparison of the AME model we present in the paper with a variety of parameterizations from the **latentnet** package. The number of dimensions in the latent space in each of these cases is set to 2. LSM (SR) represents a model in which random sender and receiver effects are included.

	LSM	LSM (SR)	AME
Intercept/Edges	0.95	0.61	-3.40
	[0.09; 1.85]	[-1.07; 2.34]	[-4.40; -2.51]
Conflicting policy preferences			
Business vs. NGO	-1.37	-3.06	-1.38
	[-2.39; -0.40]	[-4.72; -1.59]	[-2.47; -0.49]
Opposition/alliance	0.00	0.30	1.08
	[-0.40; 0.40]	[-0.26; 0.86]	[0.72; 1.49]
Preference dissimilarity	-1.77	-1.88	-0.79
	[-2.64; -0.91]	[-3.06; -0.69]	[-1.55; -0.07]
Transaction costs			
Joint forum participation	1.51	1.55	0.92
	[0.85; 2.17]	[0.66; 2.40]	[0.40; 1.46]
Influence			
Influence attribution	0.08	0.30	1.10
	[-0.40; 0.54]	[-0.38; 0.96]	[0.70; 1.55]
Alter's influence indegree	0.01	0.06	0.11
	[-0.03; 0.04]	[-0.03; 0.14]	[0.07; 0.15]
Influence absolute diff.	0.04	-0.08	-0.07
	[-0.01; 0.09]	[-0.14; -0.02]	[-0.11; -0.03]
Alter = Government actor	-0.46	-0.09	0.56
	[-1.08; 0.14]	[-1.85; 1.77]	[-0.06; 1.16]
Functional requirements			
Ego = Environmental NGO	-0.60	-1.72	0.68
	[-1.30; 0.08]	[-3.76; 0.25]	[-0.36; 1.73]
Same actor type	1.17	1.81	1.03
	[0.62; 1.72]	[1.09; 2.56]	[0.62; 1.48]

Table A.4: 95% posterior credible intervals are provided in brackets.

Figure A6: Assessments of out-of-sample predictive performance using ROC curves, separation plots, and PR curves. AUC statistics are provided as well for both curves.

**Figure A7:** Network goodness of fit summary using **amen**.

*Simulation Based Comparison of **amen** & **latentnet***

We construct a simulation study to examine differences in the ability of LSM and LFM to capture network dependencies under varying scenarios of "egalitarianism". By egalitarianism here we refer to how equally balanced the nodes are in terms of their number of ties. We construct six simulation scenarios representing varying degrees of egalitarianism. Note that to provide as fair a test as possible to the LSM we focus on comparing to just the LFM, the multiplicative effects portion of AME (see Equation 3 in the manuscript). This means that we exclude the additive effects described by the SRM portion of the model (see Equation 2 in the manuscript).

For each scenario, we simulate fifty binary, directed networks with 100 nodes each and then evaluate the performance of LFM and LSM to predict this network structure. The results are shown in Figure A8 below. Each panel here represents one scenario in which we vary the degree of egalitariansim. The left most panel represents the situation in which the structure of the network is most egalitarian. The numbers at the top of each panel indicate the standard deviation of the degree distribution averaged across fifty simulations. Across the diagnoal of the visualization, we also provide an example of the type of network that was simulated. The size of nodes in each example network corresponds to the number of ties that node has. As we go from left to right, we can see much greater variance in the size of nodes within the network, which indicates that the level of egalitarianism is changing.

We run a LFM and LSM on each of the simulated networks from each scenario, and compare the predictive performance based on AUC (ROC) and AUC (PR) statistics. We set $K = 2$ for both the LFM and LSM and estimate each model without any covariates. The results of this analysis indicate that under these varying scenarios of egalitarianism LFM consistently outperforms the LSM. However, the performance of both models tends to decline as the structure of the simulated networks become less egalitarian

(i.e., the extent of tie formation among just a few nodes becomes much higher than the typical node in the network). If covariate information was provided to the model about which nodes were more likely to form ties, then the predictive performance of both models would obviously improve. Additionally, if we were to estimate the full AME model (SRM + LFM) then the additive effects would be able to capture the degree heterogeneity. Typically, in most applied scenarios one would always to include both the additive and multiplicative effects portions when using AME.[26]
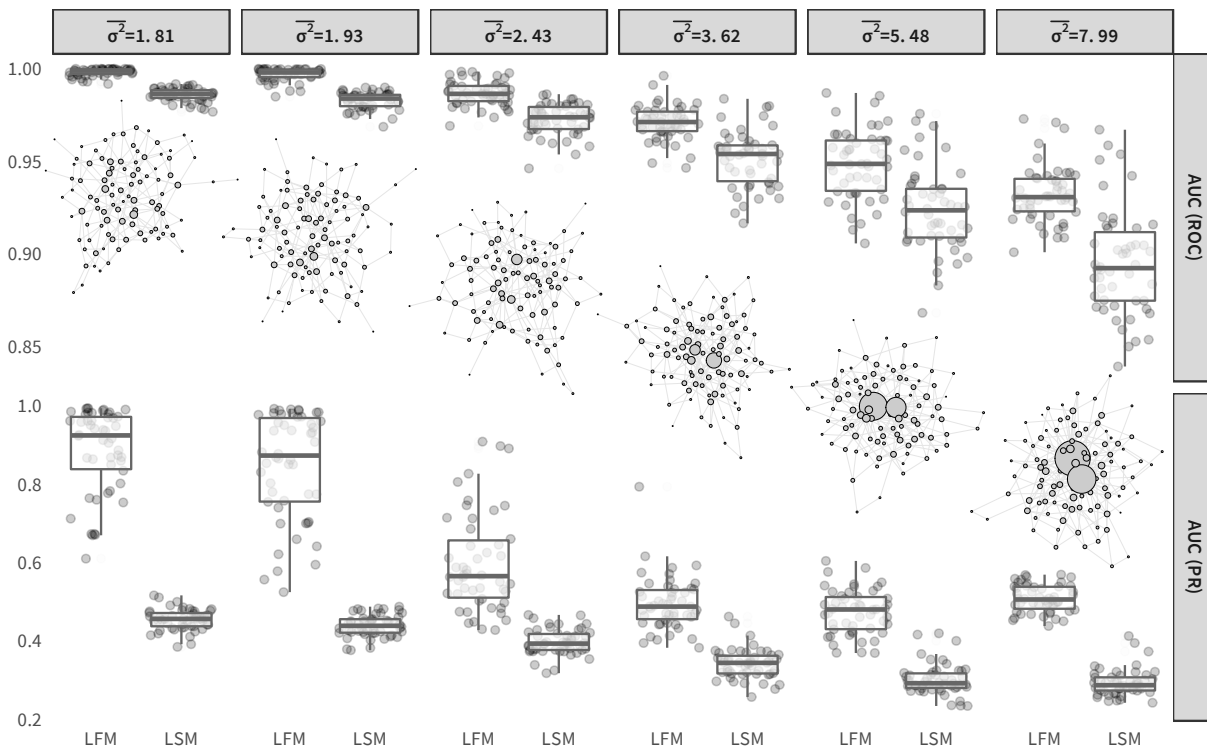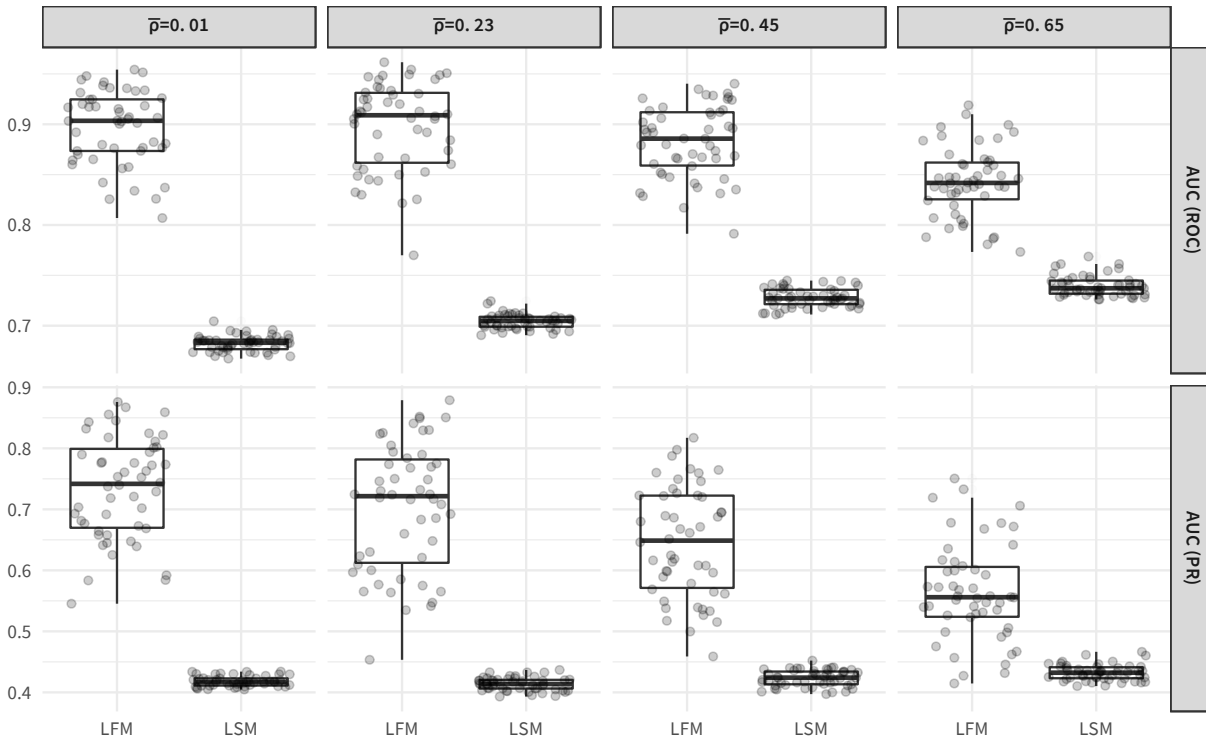

Figure A8: Predictive performance of LFM vs LSM for networks under five scenarios (the panels) that vary the extent to which the distribution of ties are egalitarian. We use a box plot to represent the performance of LFM and LSM across fifty simulations for each scenario. The set of network visualizations across the diagonal of the plot illustrate a representative network from one simulation under that scenario, and the size of nodes corresponds to their number of ties. The labels at the top of each panel indicate the standard deviation of the number of ties, which are averaged across the fifty simulations for that scenario.

---

[26]When using the **amen**, the SRM portion of the model will be included by default.

Next, we construct a second simulation study to compare the predictive perfor-
mance of LSM and LFM under varying levels of reciprocity. Here again we simulate a
set of scenarios, and for each scenario we simulate fifty binary, directed networks with
100 nodes. The results are shown in Figure A9. Each panel here represents represents
one scenario with a certain degree of reciprocity. The left-most panel highlights the
case where there is little to no reciprocity in the network and the right-most where
the level of reciprocity is quite high. The average level of reciprocity across the fifty
simulated networks is given at the top of each panel.

Figure A9: Predictive performance of LFM vs LSM for networks with varying levels of reciprocity.
We use a box plot to represent the performance of LFM and LSM across fifty simulations for
each scenario. The labels at the top of each panel indicate the average level of reciprocity across
the fifty simulated in that scenario.

To compare LFM and LSM, we again utilize AUC (ROC) and AUC (PR) statistics. K is
set to 2 for both models and no covariate information is provided. Here again we find
that the LFM consistently outpeforms the LSM, though at higher levels of reciprocity

the performance difference between the two approaches does shorten. If we were to estimate the full AME model, then we would be better able to capture reciprocity in the network, as dyadic reciprocity is estimated within the additive effects portion of AME.

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.

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. `https://dl.acm.org/citation.cfm?id=1143874` pp. 233–240.
  URL: *https://dl.acm.org/citation.cfm?id=1143874*

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.

Durante, Daniele, David B. Dunson and Joshua T. Vogelstein. 2017. "Nonparametric Bayes Modeling of Populations of Networks." *Journal of the American Statistical Association* 112(520):1516–1530.

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.

Gollini, Isabella and Thomas B. Murphy. 2016. "Joint Modeling of Multiple Network Views." *Journal of Computational and Graphical Statistics* 25(1):246–265.

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. 2003. Statistical Models for Social Networks: Inference and Degeneracy. In *Dynamic Social Network Modeling and Analysis*, ed. Breiger Ronald, Carley Kathlene and Pattison Pip. Vol. 126 of *Committee on Human Factors, Board on Behavioral, Cognitive, and Sensory Sciences* Washington D.C.: National Academy Press pp. 229–252.

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.

Hoff, Peter D. 2005. "Bilinear Mixed-Effects Models for Dyadic Data." *Journal of the American Statistical Association* 100(4690):286–295.

Hoff, Peter D. 2008. Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems 20*, ed. John C. Platt, Daphne Koller, Yoram Singer and Sam T. Roweis. Processing Systems 21 Cambridge, MA, USA: MIT Press pp. 657–664.

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.

Kao, E. K., S. T. Smith and E. M. Airoldi. 2018. "Hybrid Mixed-Membership Blockmodel for Inference on Realistic Network Interactions." *IEEE Transactions on Network Science and Engineering* .

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.

Sewell, Daniel K. and Yuguo Chen. 2015. "Latent space models for dynamic networks." *Journal of the American Statistical Association* 110(512):1646–1657.

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.