Supporting Information for "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*

[☆]S.M. and M.W. acknowledge support from National Science Foundation (NSF) Award 1259266 and P.H. acknowledges support from NSF Award 1505136.

[*]Corresponding author

Email address: s7.minhas@gmail.com (Shahryar Minhas)

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 $\{\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and $\sigma_\epsilon^2\}$, the algorithm proceeds as follows:
  - sample $\boldsymbol{\theta} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
  - sample $\boldsymbol{\beta} \mid \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
  - sample $\mathbf{a}, \mathbf{b} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
  - sample $\Sigma_{ab} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \rho,$ and σ_ϵ^2 (Inverse-Wishart)
  - update ρ using a Metropolis-Hastings step with proposal $p^*|p \sim$ truncated normal$_{[-1,1]}(\rho, \sigma_\epsilon^2)$
  - sample $\sigma_\epsilon^2 \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab},$ and ρ (Inverse-Gamma)
  - For each $k \in K$:
    * Sample $\mathbf{U}_{[,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}_{[,-k]}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
    * Sample $\mathbf{V}_{[,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}_{[,-k]}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)
    * Sample $\mathbf{D}_{[k,k]} \mid \boldsymbol{\beta}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{U}, \mathbf{V}, \Sigma_{ab}, \rho,$ and σ_ϵ^2 (Normal)[1]

[1] 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.

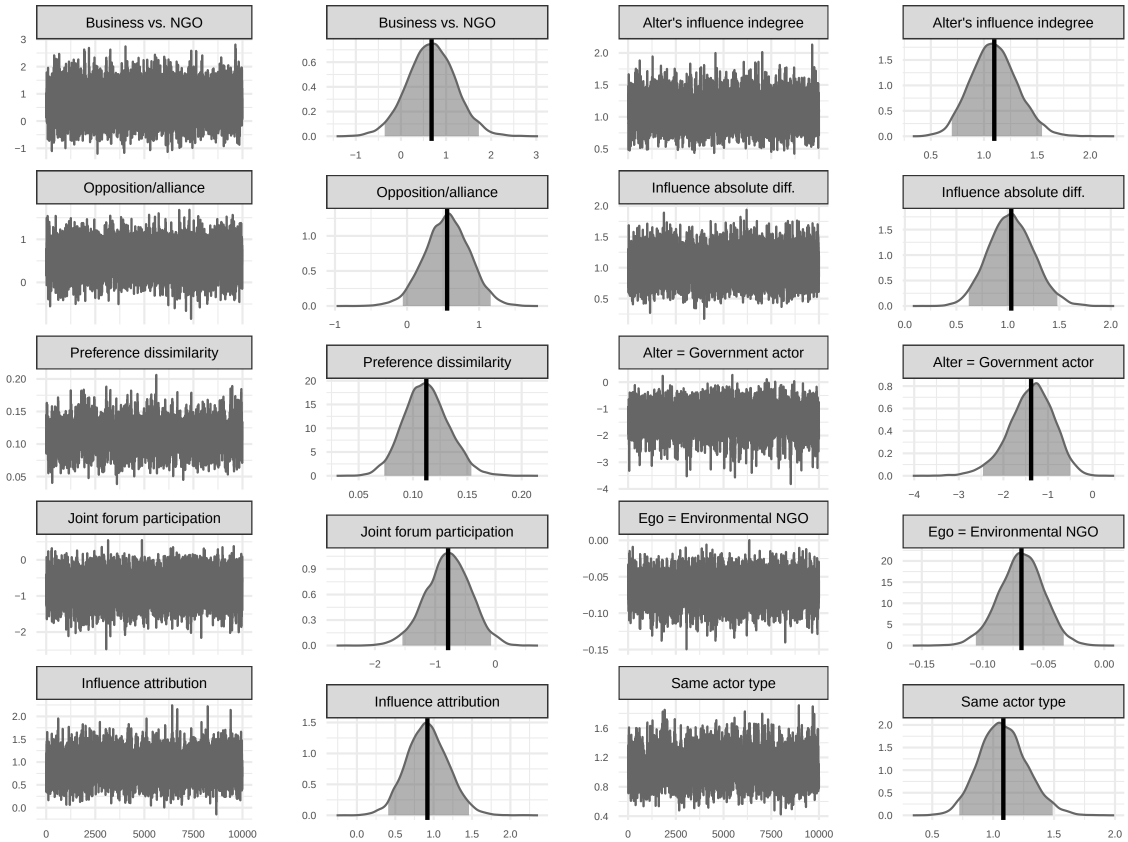

Figure A1: Trace plot for AME model presented in paper. In this model, we utilize the SRM to account for first and second-order dependence. To account for third order dependencies we use the latent factor approach with $K = 2$.

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.

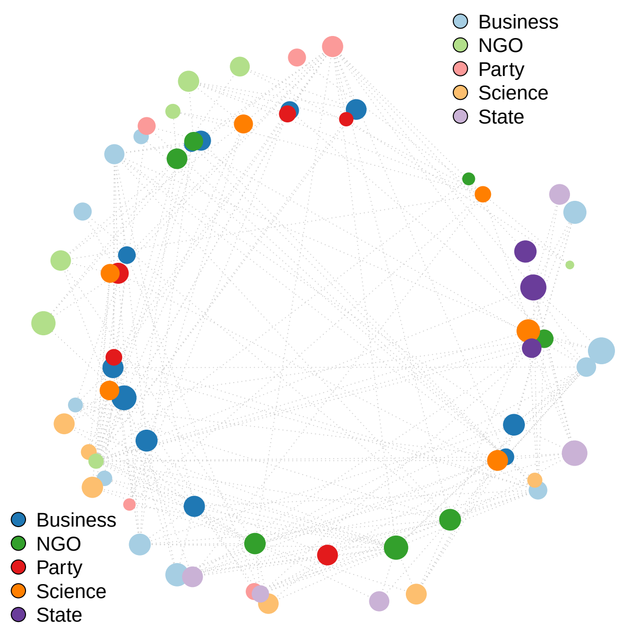

Figure A2: Circle plot of estimated latent factors.

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.[2] The size of actors is a function of the magnitude of the vectors, and dashed lines between actors indicate greater than expected levels of

[2]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:[3]

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.[4] 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.

[3] See Morris et al. (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.

[4] Calculation for the incoming k-star statistic is not currently supported by the **latentnet** package.

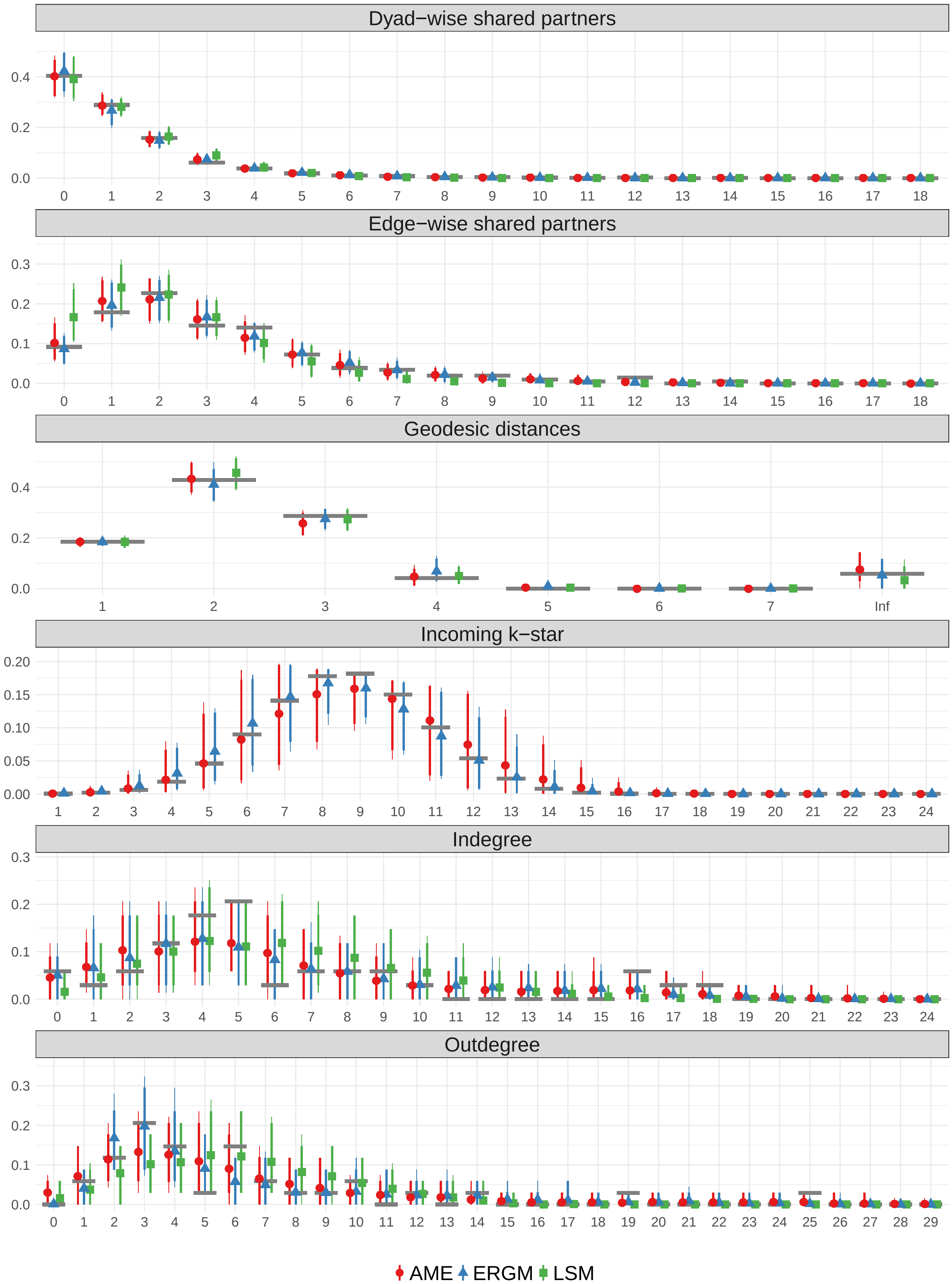

Figure A3: Goodness of fit statistics to assess how well the LSM, ERGM, and AME approaches account for network dependencies. Grey bars indicate true values.

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.

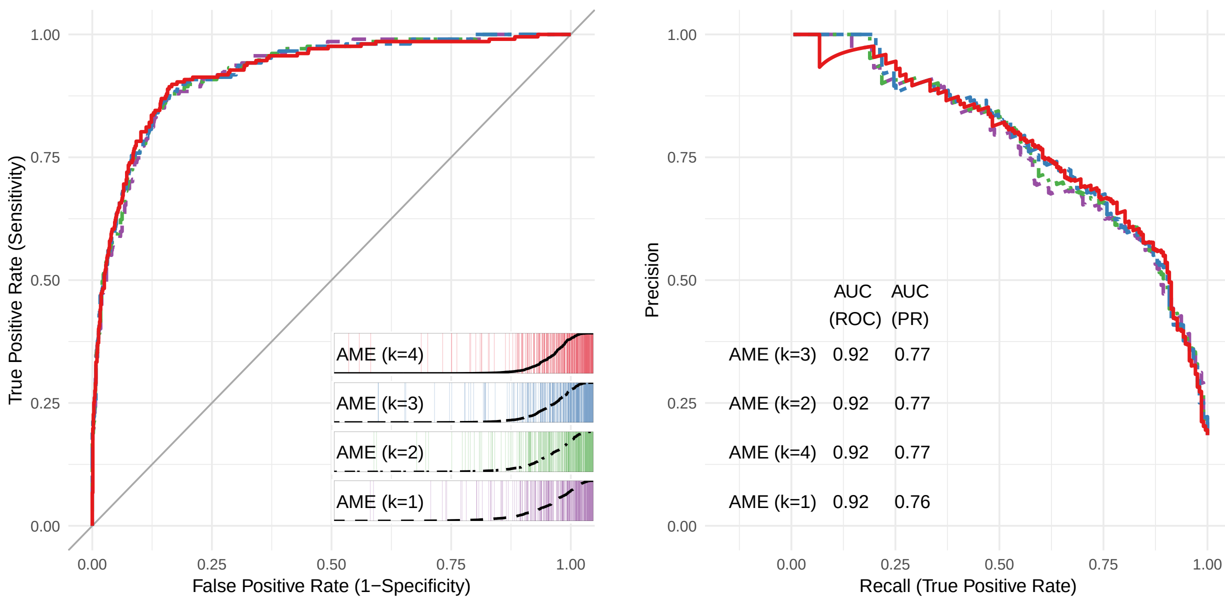

Figure A4: Assessments of out-of-sample predictive performance using ROC curves, separation plots, and PR curves. AUC statistics are provided as well for both curves.

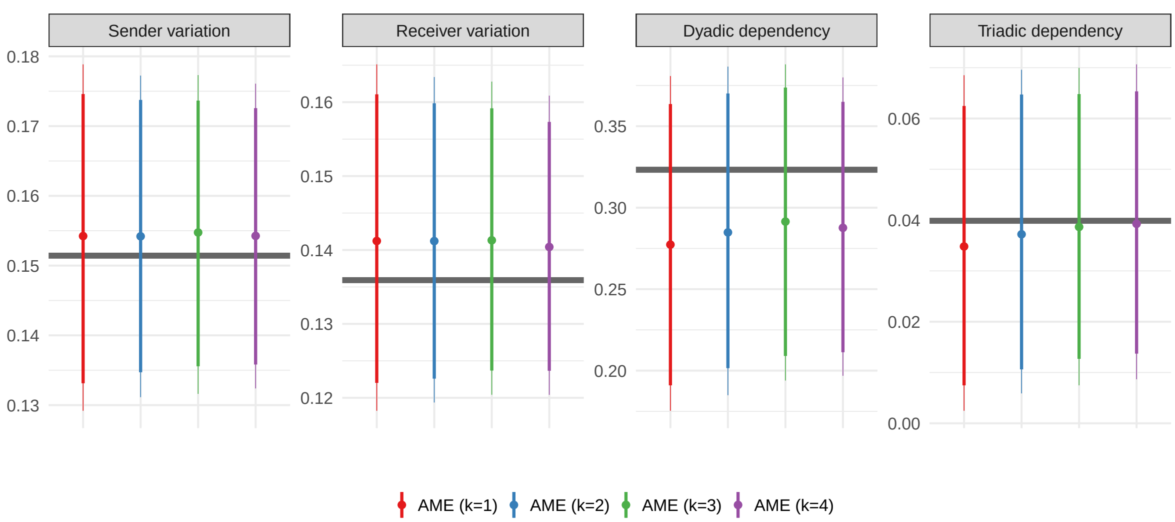

Figure A5: Network goodness of fit summary using **amen**.

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.

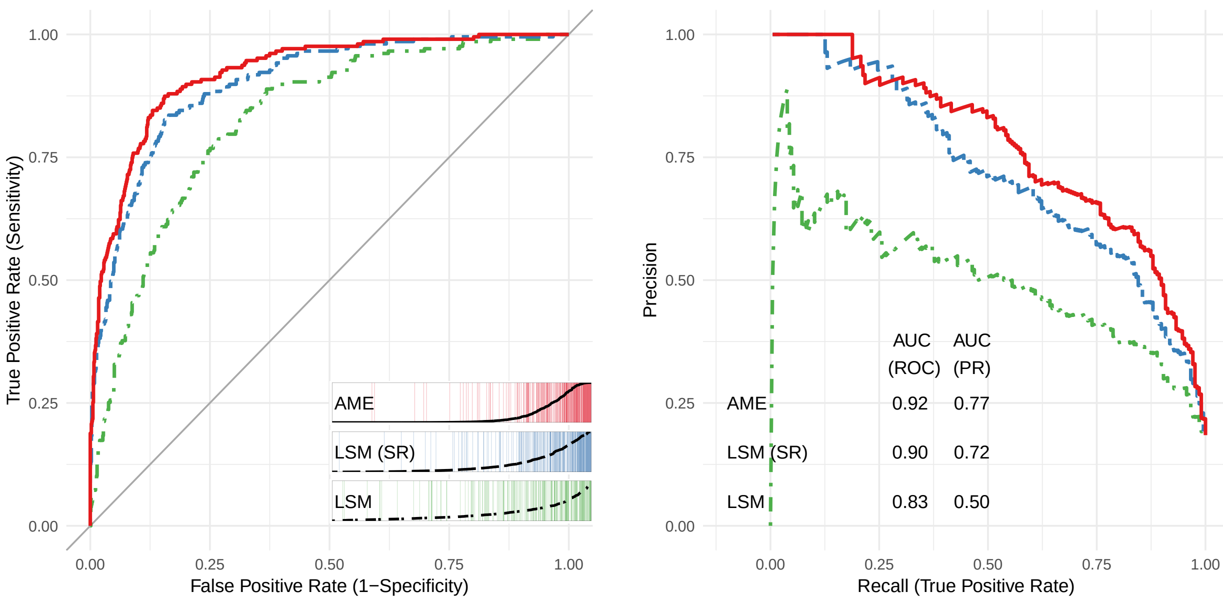

Figure A6: Assessments of out-of-sample predictive performance using ROC curves, separation plots, and PR curves. AUC statistics are provided as well for both curves.

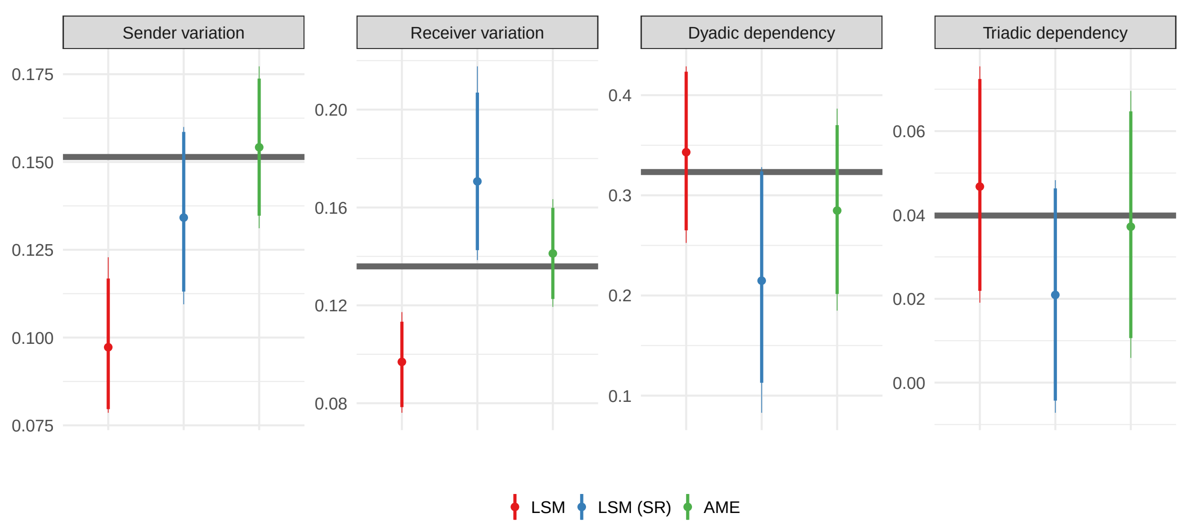

Figure A7: Network goodness of fit summary using **amen**.

*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.[5]

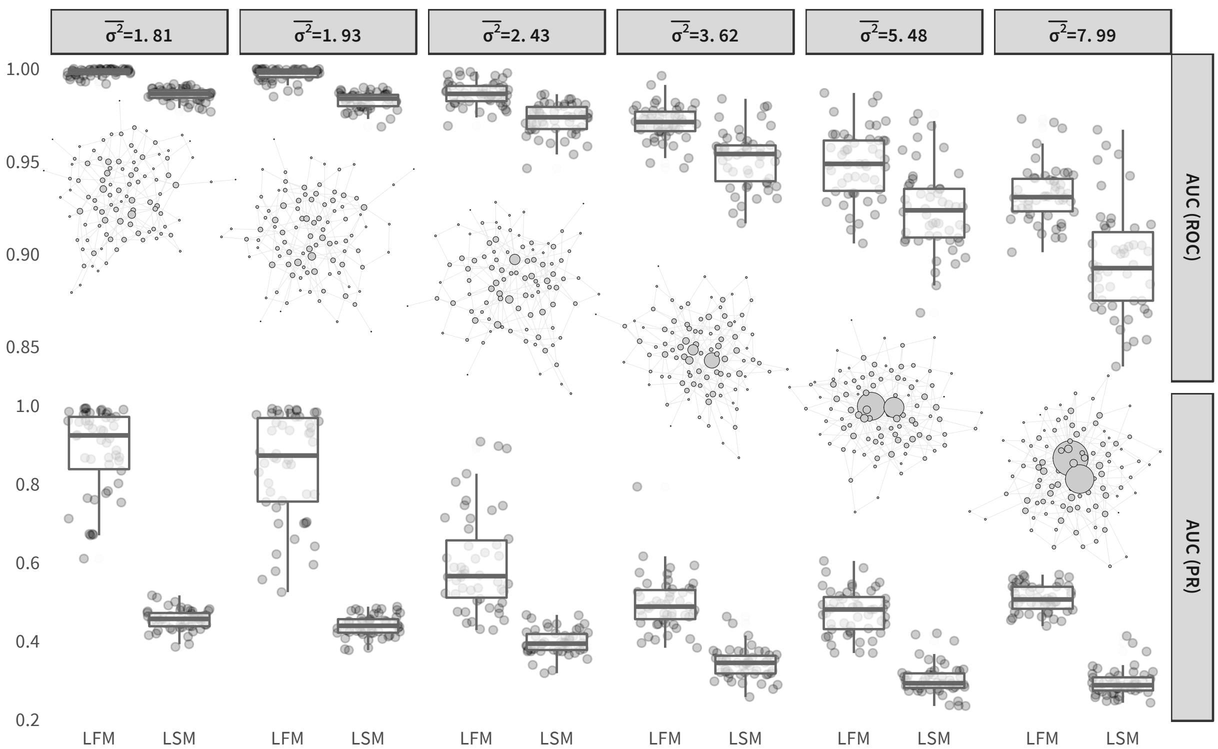

Figure A8: Predictive performance of LFM vs LSM for networks under five scenarios (the panels) that vary the extent to which the distribution of ties are egalitarian. We use a box plot to represent the performance of LFM and LSM across fifty simulations for each scenario. The set of network visualizations 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.

[5]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 performance 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.

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.

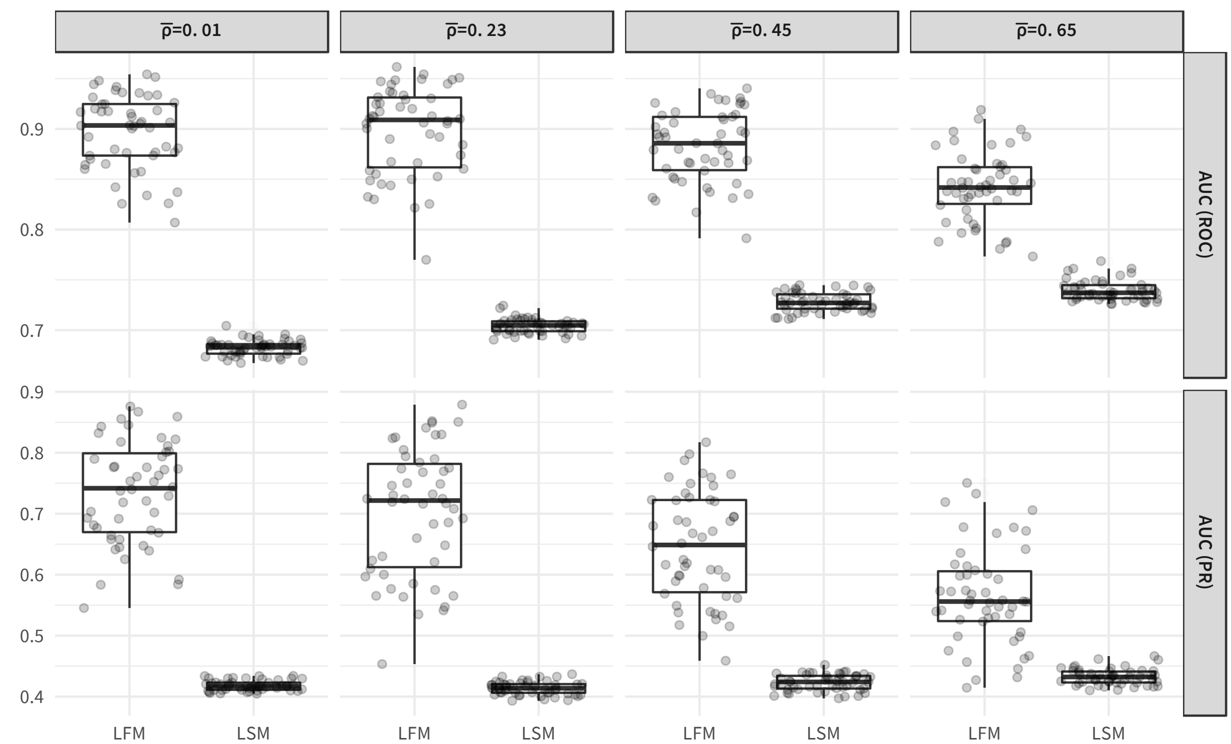

Figure A9: Predictive performance of LFM vs LSM for networks with varying levels of reciprocity. We use a box plot to represent the performance of LFM and LSM across fifty simulations for each scenario. The labels at the top of each panel indicate the average level of reciprocity across the fifty simulated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

Hoff, P. D., 2005. Bilinear mixed-effects models for dyadic data. Journal of the American Statistical Association 100 (4690), 286–295.

Hoff, P. D., 2008. Modeling homophily and stochastic equivalence in symmetric relational data. In: Platt, J. C., Koller, D., Singer, Y., Roweis, S. T. (Eds.), Advances in Neural Information Processing Systems 20. Processing Systems 21. MIT Press, Cambridge, MA, USA, pp. 657–664.

Morris, M., Handcock, M. S., Hunter, D. R., 2008. Specification of exponential-family random graph models: Terms and computational aspects. Journal of Statistical Software 24 (4), 1554–7660.