POLNET2017: AMEN for Latent Factor Models

Shahryar Minhas[†], Peter D. Hoff[‡], & Michael D. Ward[†]
Duke University
[†] Department of Political Science
[‡] Department of Statistical Science

June 16, 2017

Motivation

Relational data consists of

- a set of units or nodes
- a set of measurements, y_{ij}, specific to pairs of nodes (i, j)

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

	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

Relational data assumptions

GLM: $y_{ij} \sim \beta^T X_{ij} + e_{ij}$

Networks typically show evidence against independence of $e_{ij} : i \neq j$

Not accounting for dependence can lead to:

- biased effects estimation
- uncalibrated confidence intervals
- poor predictive performance
- inaccurate description of network phenomena

We've been hearing this concern for decades now:

Thompson & Walker (1982)	Beck et al. (1998)	Snijders (2011)
Frank & Strauss (1986)	Signorino (1999)	Erikson et al. (2014)
Kenny (1996)	Li & Loken (2002)	Aronow et al. (2015)
Krackhardt (1998)	Hoff & Ward (2004)	Athey et al. (2016)

Outline

- Nodal and dyadic dependencies in networks
  - Can model using the “A” in AME
- Third order dependencies
  - Can model using the “M” in AME
- Application

What network phenomena? Sender heterogeneity

Values across a row, say $\{y_{ij}, y_{ik}, y_{il}\}$, may be more similar to each other than other values in the adjacency matrix because each of these values has a common sender i

	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

What network phenomena? Receiver heterogeneity

Values across a column, say $\{y_{ji}, y_{ki}, y_{li}\}$, may be more similar to each other than other values in the adjacency matrix because each of these values has a common receiver i

	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

What network phenomena? Sender-Receiver Covariance

Actors who are more likely to send ties in a network may also be more likely to receive them

	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

What network phenomena? Reciprocity

Values of y_{ij} and y_{ji} may be statistically dependent

	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

Social Relations Model (The "A" in AME)

We use this model to form the additive effects portion of AME

$$
\begin{aligned}
y_{ij} &= \color{red}{\mu + e_{ij}} \\
e_{ij} &= a_i + b_j + \epsilon_{ij} \\
\{(a_1, b_1), \ldots, (a_n, b_n)\} &\sim N(0, \Sigma_{ab}) \\
\{(\epsilon_{ij}, \epsilon_{ji}) : \; i \neq j\} &\sim N(0, \Sigma_{\epsilon}), \text{ where}
\end{aligned}
$$

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

- μ baseline measure of network activity
- e_{ij} residual variation that we will use the SRM to decompose

Social Relations Model (The "A" in AME)

$$
\begin{aligned}
y_{ij} &= \mu + e_{ij} \\
e_{ij} &= a_i + b_j + \epsilon_{ij} \\
\{(a_1, b_1), \ldots, (a_n, b_n)\} &\sim N(0, \Sigma_{ab}) \\
\{(\epsilon_{ij}, \epsilon_{ji}) : \ i \neq j\} &\sim N(0, \Sigma_\epsilon), \text{ where}
\end{aligned}
$$

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

- row/sender effect (a_i) & column/receiver effect (b_j)

- Modeled jointly to account for correlation in how active an actor is in sending and receiving ties

Social Relations Model (The "A" in AME)

$$
\begin{aligned}
y_{ij} &= \mu + e_{ij} \\
e_{ij} &= a_i + b_j + \epsilon_{ij} \\
\{(a_1, b_1), \ldots, (a_n, b_n)\} &\sim N(0, \Sigma_{ab}) \\
\{(\epsilon_{ij}, \epsilon_{ji}) : \ i \neq j\} &\sim N(0, \Sigma_\epsilon), \text{ where}
\end{aligned}
$$

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

- σ_a^2 and σ_b^2 capture heterogeneity in the row and column means
- σ_{ab} describes the linear relationship between these two effects (i.e., whether actors who send [receive] a lot of ties also receive [send] a lot of ties)

Social Relations Model (The “A” in AME)

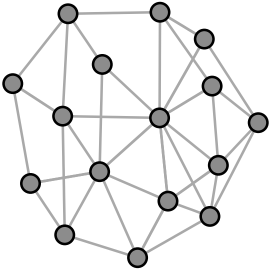
$$
\begin{aligned}
y_{ij} &= \mu + e_{ij} \\
e_{ij} &= a_i + b_j + \epsilon_{ij} \\
\{(a_1, b_1), \ldots, (a_n, b_n)\} &\sim N(0, \Sigma_{ab}) \\
\{(\epsilon_{ij}, \epsilon_{ji}) : \ i \neq j\} &\sim N(0, \Sigma_{\epsilon}), \text{ where}
\end{aligned}
$$

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

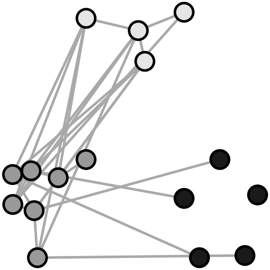
- ϵ_{ij} captures the within dyad effect
- Second-order dependencies are described by σ_{ϵ}^2
- Reciprocity, aka within dyad correlation, represented by ρ

Third Order Dependencies

Homophily



Stochastic Equivalence



To account for these patterns we can build on what we have so far and find an expression for γ:

$$y_{ij} \approx \beta^T X_{ij} + a_i + b_j + \gamma(u_i, v_j)$$

Latent Factor Model: The "M" in AME

Each node i has an unknown latent factor

$$\mathbf{u}_i, \mathbf{v}_j \in \mathbb{R}^k \quad i, j \in \{1, \ldots, n\}$$

The probability of a tie from i to j depends on their latent factors

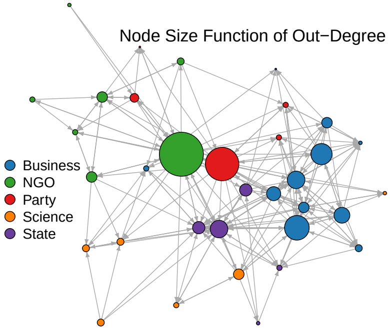
$$\begin{aligned} \gamma(\mathbf{u}_i, \mathbf{v}_j) &= \mathbf{u}_i^T D \mathbf{v}_j \\ &= \sum_{k \in K} d_k u_{ik} v_{jk} \end{aligned}$$

D is a $K \times K$ diagonal matrix

Can account for both stochastic equivalence and homophily

Swiss Climate Change Application

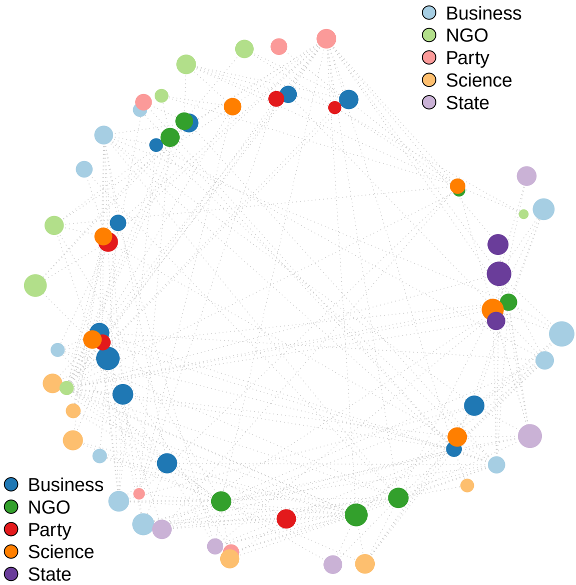
Cross-sectional network measuring whether an actor indicated that they collaborated with another during the policy design of the Swiss CO_2 act (Ingold 2008)



Parameter Estimates

	Expected Effect	Logit	MRQAP	LSM	ERGM	AME
Conflicting policy preferences						
Business vs. NGO	−	-0.86	-0.87*	-1.37*	-1.11*	-1.37*
Opposition/alliance	+	1.21*	1.14*	0.00	1.22*	1.08*
Preference dissimilarity	−	-0.07	-0.60	-1.76*	-0.44	-0.79*
Transaction costs						
Joint forum participation	+	0.88*	0.75*	1.51*	0.90*	0.92*
Influence						
Influence attribution	+	1.20*	1.29*	0.08	1.00*	1.09*
Alter's influence indegree	+	0.10*	0.11*	0.01	0.21*	0.11*
Influence absolute diff.	−	-0.03*	-0.06*	0.04	-0.05*	-0.07*
Alter = Government actor	+	0.63*	0.68	-0.46	1.04*	0.55
Functional requirements						
Ego = Environmental NGO	+	0.88*	0.99	-0.60	0.79*	0.67
Same actor type	+	0.74*	1.12*	1.17*	0.99*	1.04*

Latent Factor Visualization

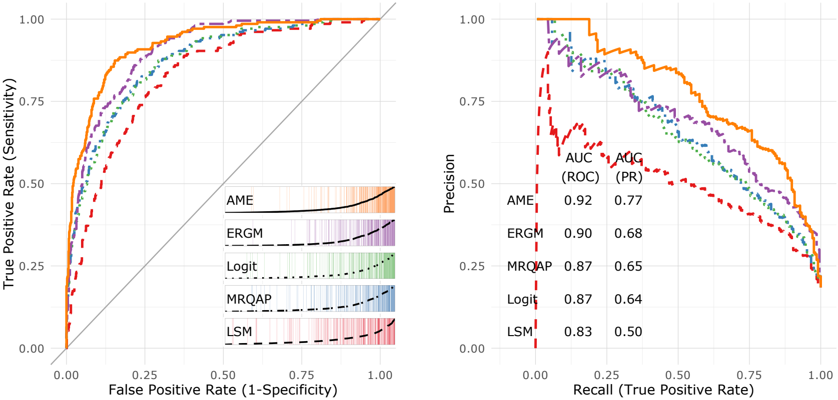


Out of Sample Performance Assessment

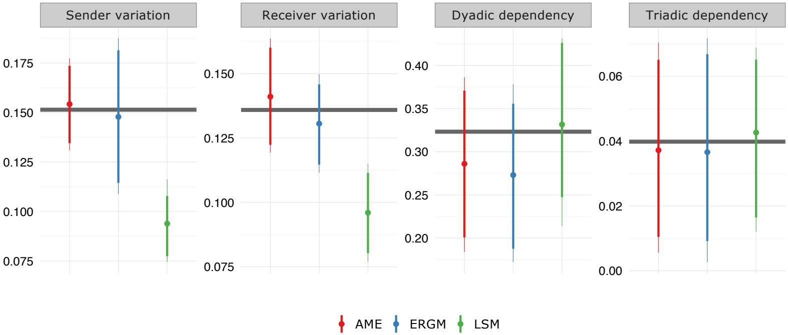
- Randomly divide the $n \times (n-1)$ data points into S sets of roughly equal size, letting s_{ij} be the set to which pair $\{ij\}$ is assigned.
- For each $s \in \{1, \ldots, S\}$:
  - Obtain estimates of the model parameters conditional on $\{y_{ij} : s_{ij} \neq s\}$, the data on pairs not in set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

This procedure generates a sociomatrix of out-of-sample predictions of the observed data

Performance Comparison

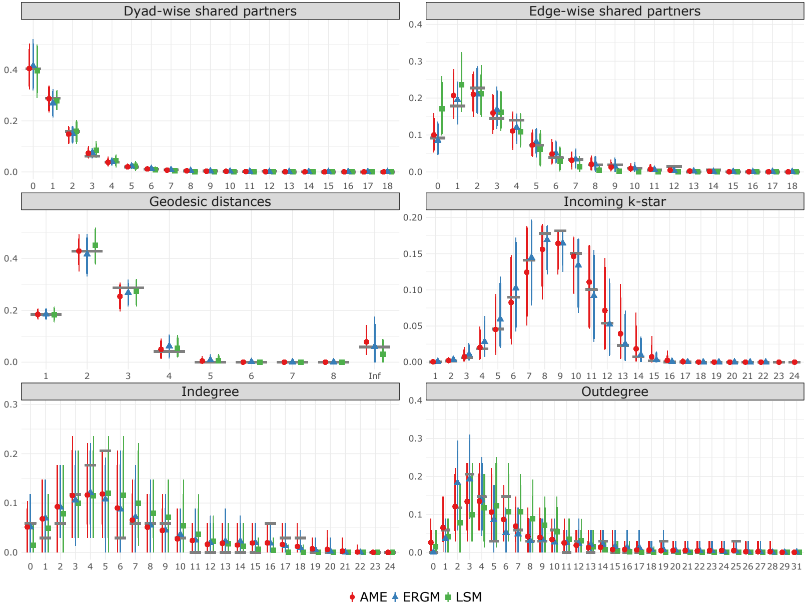


Network Dependencies

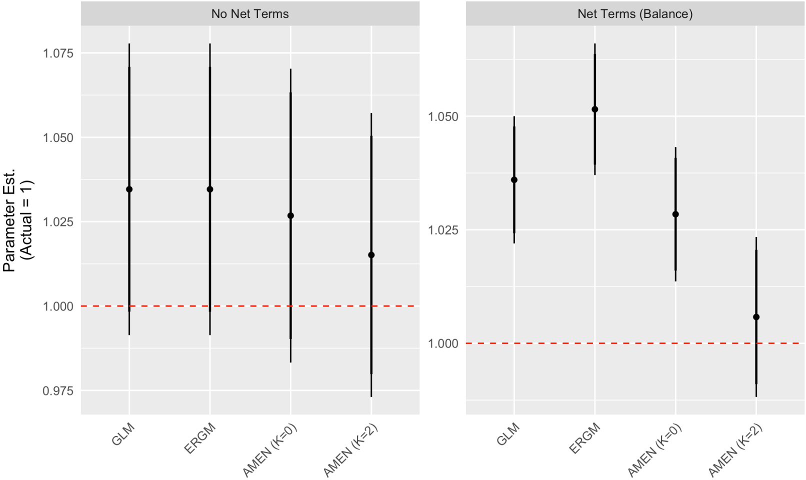


Thanks.

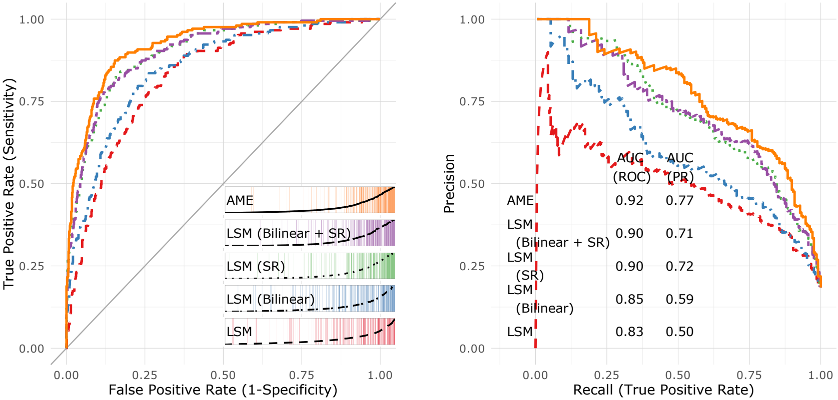
Standard Network Dependence Measures



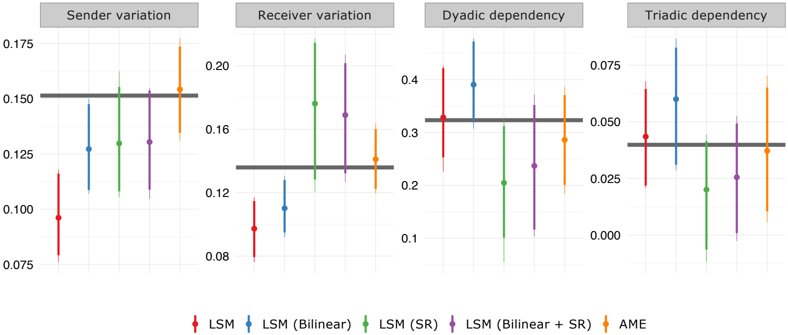
Simulation Comparison



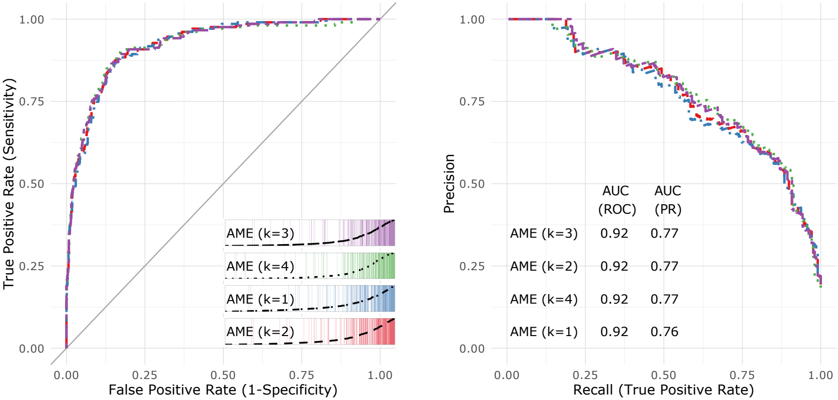
AMEN v LSM Performance



AMEN versus LSM Net Dependence



AMEN varying K Performance



AMEN varying K Net Dependence