AMEN for Latent Factor Models

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

April 6, 2017

Motivation

Relational data are composed of interactions between actors that are interdependent

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}

$\longrightarrow$

	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

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)

Social Relations Model

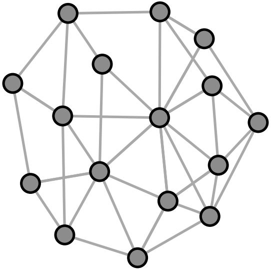
$$
\begin{aligned}
y_{ij} &= \mu + e_{ij} \\
e_{ij} &= a_i + b_j + \epsilon_{ij} \\
\{(a_1, b_1), \ldots, (a_n, b_n)\} &\overset{\text{iid}}{\sim} N(0, \Sigma_{ab}) \\
\{(\epsilon_{ij}, \epsilon_{ji}) : \ i \neq j\} &\overset{\text{iid}}{\sim} N(0, \Sigma_{\epsilon}), \text{ where} \\
\Sigma_{ab} = \begin{pmatrix}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

Third Order Dependencies

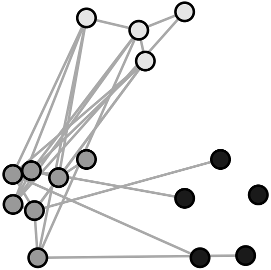
Homophily: “birds of a feather flock together”

Stochastic equivalence: nothing as pithy to say here, but this model focuses on community detection

Homophily



Stochastic Equivalence



Latent Variable Models

Latent class model

$$\alpha(u_i, u_j) = m_{u_i,u_j}$$
$$u_i \in \{1,\ldots,K\},\ i \in \{1,\ldots,n\}$$
$$M \text{ a } K \times K \text{ symmetric matrix}$$

Latent distance model

$$\alpha(\mathbf{u}_i, \mathbf{u}_j) = -|\mathbf{u}_i - \mathbf{u}_j| \tag{2}$$
$$\mathbf{u}_i \in \mathbb{R}^K,\ i \in \{1,\ldots,n\}$$

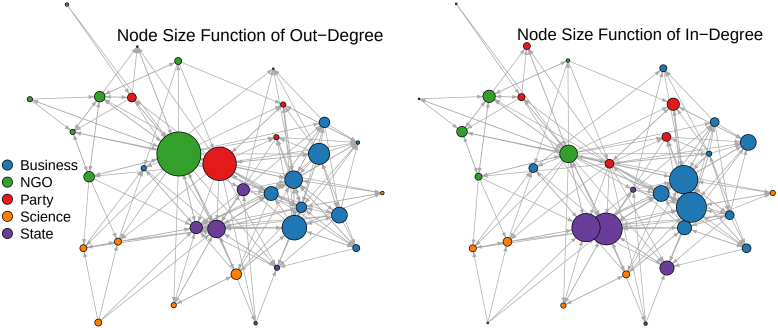
Latent factor model

$$\alpha(\mathbf{u}_i, \mathbf{u}_j) = \mathbf{u}_i^T \Lambda \mathbf{u}_j$$
$$\mathbf{u}_i \in \mathbb{R}^K,\ i \in \{1,\ldots,n\}$$
$$\Lambda \text{ a } K \times K \text{ diagonal matrix}$$

Putting it together: AME

$$
\begin{aligned}
y_{ij} &= g(\theta_{ij}) \\
\theta_{ij} &= \boldsymbol{\beta}^T \mathbf{X}_{ij} + e_{ij} \\
e_{ij} &= a_i + b_j + \epsilon_{ij} + \alpha(\mathbf{u}_i, \mathbf{v}_j) \text{ , where} \\
&\alpha(\mathbf{u}_i, \mathbf{v}_j) = \mathbf{u}_i^T \mathbf{D} \mathbf{v}_j = \sum_{k \in K} d_k u_{ik} v_{jk}
\end{aligned}
\tag{3}
$$

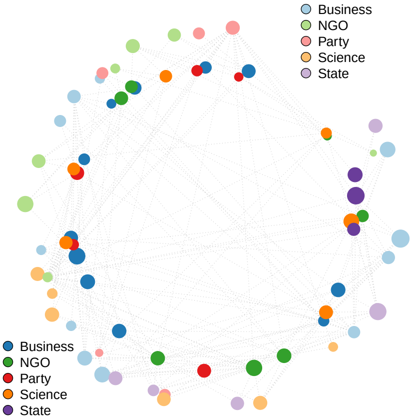
Swiss Climate Change Application



Parameter Estimates

Variable	Expected	Logit	MRQAP	LSM	ERGM	AME
Conflicting policy preferences						
Business v. NGO	−	−	−	+	−	−
Opposition/alliance	+	−	−	+	−	−
Preference dissimilarity	−	−	−	+	−	−
Transaction costs						
Joint forum participation	+	−	−	+	−	−
Influence						
Influence attribution	+	−	−	+	−	−
Alter's influence in-degree	+	−	−	+	−	−
Influence absolute diff.	−	−	−	+	−	−
Alter = Government Actor	+	−	−	+	−	−
Functional requirements						
Ego = Environment NGO	+	−	−	+	−	−
Same actor type	+	−	−	+	−	−
Endogenous dependencies: ERGM Specific Parameters						
Outdegree popularity	+	−	−	+	−	−
Twopaths	−	−	−	+	−	−
GWIdegree (2.0)	+	−	−	+	−	−
GWESP (1.0)	+	−	−	+	−	−
GWOdegree (0.5)	+	−	−	+	−	−

Latent Factor Visualization



Out of Sample Performance Assessment

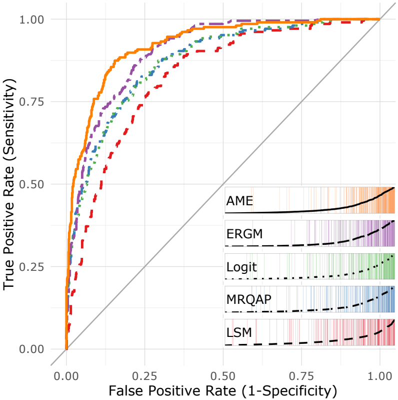
Randomly divide the $n \times (n-1)$ data points into S sets of roughly equal size, letting s_{ij} be the set to which pair $\{ij\}$ is assigned.

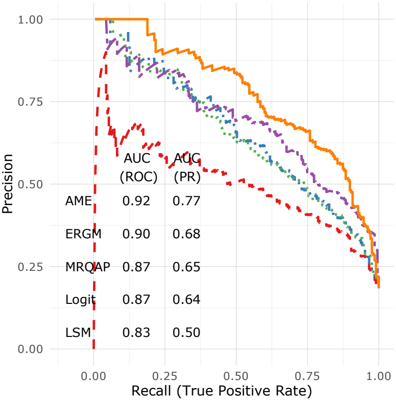
For each $s \in \{1, \ldots, S\}$:

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

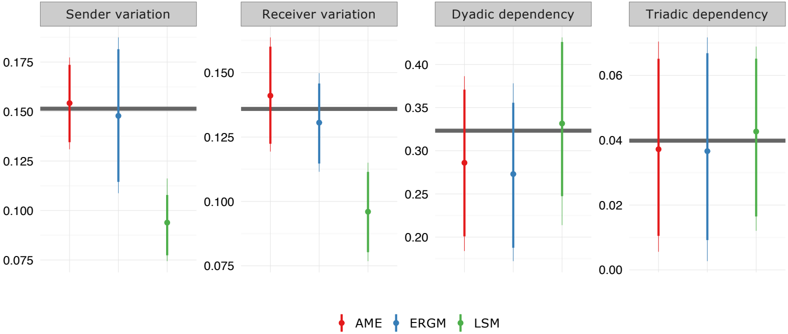
For pairs $\{kl\}$ in set s, let $\hat{y}_{kl} = E[y_{kl} | \{y_{ij} : s_{ij} \neq s\}]$, the predicted value of y_{kl} obtained using data not in set s.

Performance Comparison



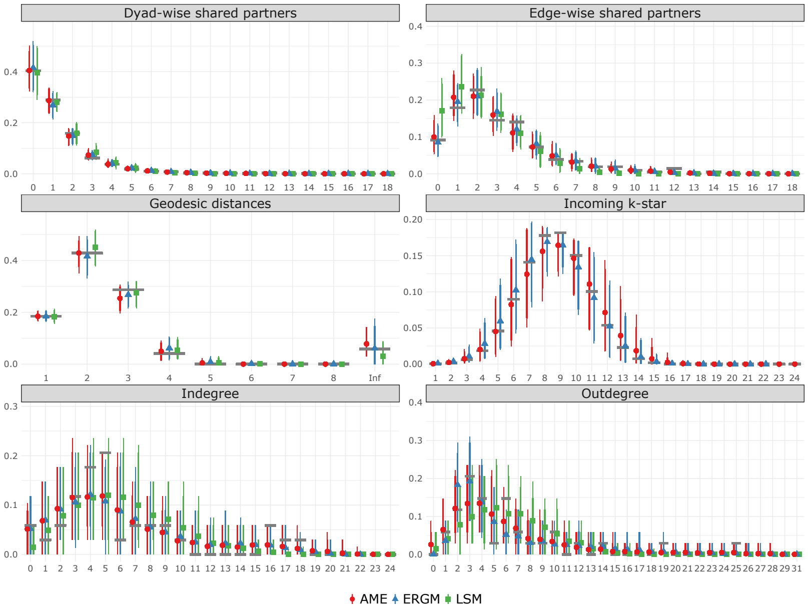


Network Dependencies

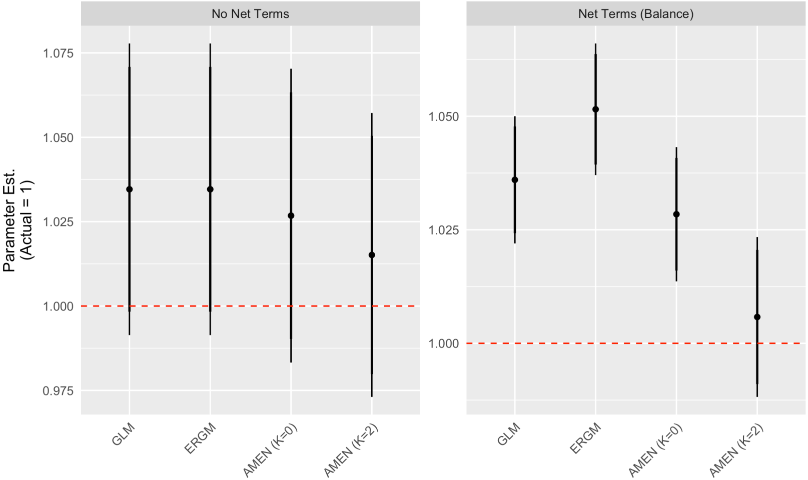


Thanks.

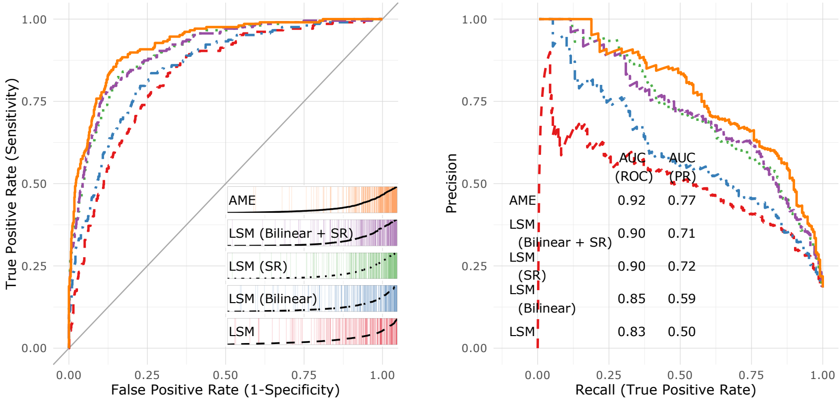
Standard Network Dependence Measures



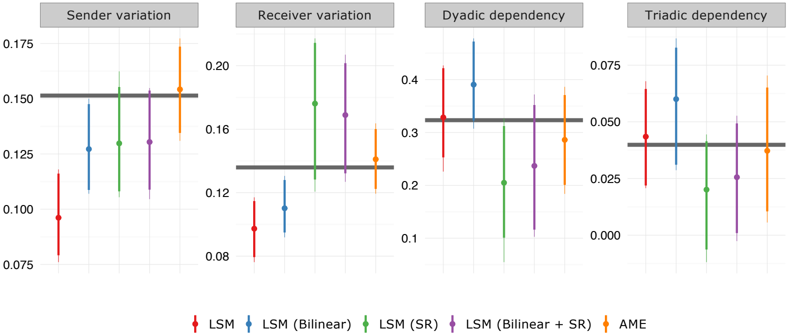
Simulation Comparison



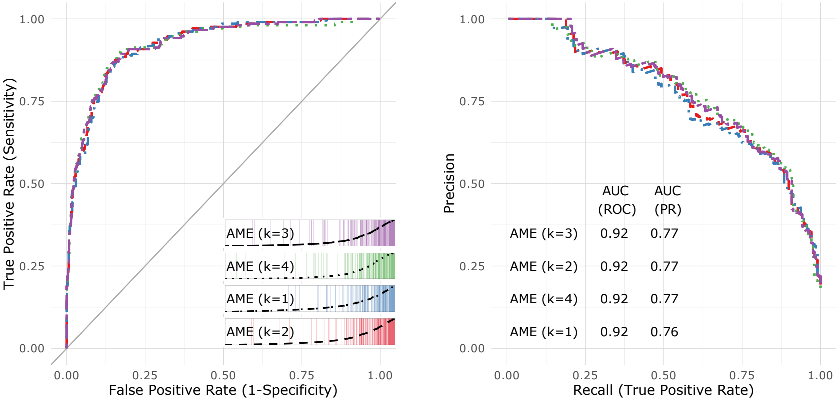
AMEN v LSM Performance



AMEN V LSM Net Dependence



AMEN v LSM Performance



AMEN V LSM Net Dependence