Complex Stochastic Weighted Graphs: Flexible Specification and Simulation

James D. Wilson[1], *Matthew J. Denny*[2], Shankar Bhamidi[1], Skyler Cranmer[3], Bruce Desmarais[2]

[1]UNC Chapel Hill, [2]UMass Amherst, [3]OSU | Preprint: http://arxiv.org/abs/1505.04015 | Contact: mdenny@polsci.umass.edu, www.mjdenny.com

Project Objectives

- Develop Metropolis-Hastings for GERGM estimation.
- Develop weighting to avoid likelihood degeneracy.
- Illustrate applications of GERGM.

Generalized Exponential Random Graph Model

- Flexible generative model for networks Y with m edges, n vertices.
- Extends ERGM to real valued networks.
- Accepts arbitrary network statistics $\mathbf{h}(x)$ controlled by $\boldsymbol{\theta}$ parameters.
- Works by restricting observed network Y to X, $x \in [0,1]$, through use of a transformation function T, parameterized by $\boldsymbol{\Lambda}$.
- Computational challenge is to approximate normalizing constant Z.

Exponential Random Graph Model (ERGM):

$$f_X(x,\boldsymbol{\theta}) = \frac{\exp\left(\boldsymbol{\theta}'\mathbf{h}(x)\right)}{\int_{[0,1]^m}\exp\left(\boldsymbol{\theta}'\mathbf{h}(z)\right)dz}, \qquad x \in [0,1]^m \tag{1}$$

Generalized ERGM (GERGM):

$$f_Y(y,\boldsymbol{\theta},\boldsymbol{\Lambda}) = \frac{\exp\left(\boldsymbol{\theta}'\mathbf{h}(T(y,\boldsymbol{\Lambda}))\right)}{\int_{[0,1]^m}\exp\left(\boldsymbol{\theta}'\mathbf{h}(z)\right)dz}\Pi_{ij}t_{ij}(y,\boldsymbol{\Lambda}), \quad y \in \mathbb{R}^m \tag{2}$$

Flexible Network Statistics:

Network Statistic	Parameter	Value
Reciprocity	θ_R	$\left(\Sigma_{i<j}x_{ij}x_{ji}\right)^{\alpha_R}$
Cyclic Triads	θ_{CT}	$\left(\Sigma_{i<j<k}\left(x_{ij}x_{jk}x_{ki} + x_{ik}x_{kj}x_{ji}\right)\right)^{\alpha_{CT}}$
In-Two-Stars	θ_{ITS}	$\left(\Sigma_i\Sigma_{j<k\neq i}x_{ji}x_{ki}\right)^{\alpha_{ITS}}$
Out-Two-Stars	θ_{OTS}	$\left(\Sigma_i\Sigma_{j<k\neq i}x_{ij}x_{ik}\right)^{\alpha_{OTS}}$
Edge Density	θ_E	$\left(\Sigma_{i\neq j}x_{ij}\right)^{\alpha_E}$
Transitive Triads	θ_{TT}	$\left(\Sigma_{i<j<k}\left(x_{ij}x_{jk}x_{ik} + x_{ij}x_{kj}x_{ik} + x_{ij}x_{jk}x_{ik}\right) + \Sigma_{i<j<k}\left(x_{ji}x_{jk}x_{ki} + x_{ji}x_{jk}x_{ik} + x_{ji}x_{kj}x_{ik}\right)\right)^{\alpha_{TT}}$

Degeneracy and the In-Two-Stars Model

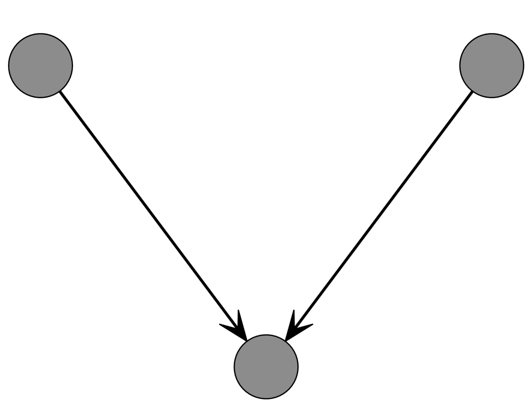


Figure: In-Two Star

- We can model **preferential attachment** using the In-Two-Stars Model.
- Functional Form:

$$f_X(x,\boldsymbol{\theta},\alpha) = \frac{\exp\left(\theta_E h_E(x) + \theta_{ITS}h_{ITS}(x)\right)}{C(\theta_E,\theta_{ITS})} \tag{3}$$

Degeneracy Issues

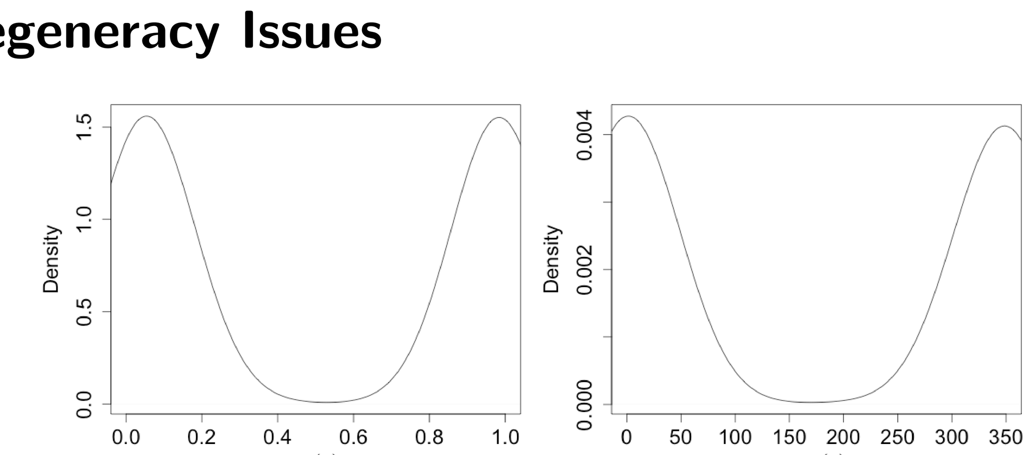


Figure: Statistics from 400K simulated In-Two-Stars networks with no exponential weighting. Networks are typically degenerate and either empty or complete.

α-Weighting Can Avoid Degeneracy

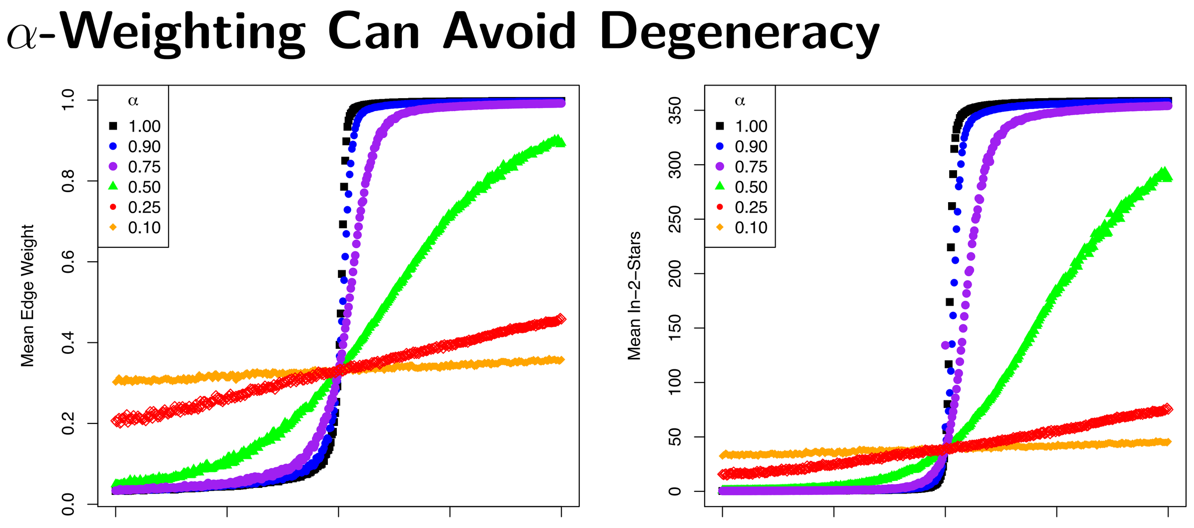


Figure: Statistics from 1K simulated networks with α - weighting on the in-two-stars statistic across values of θ_{ITS}.

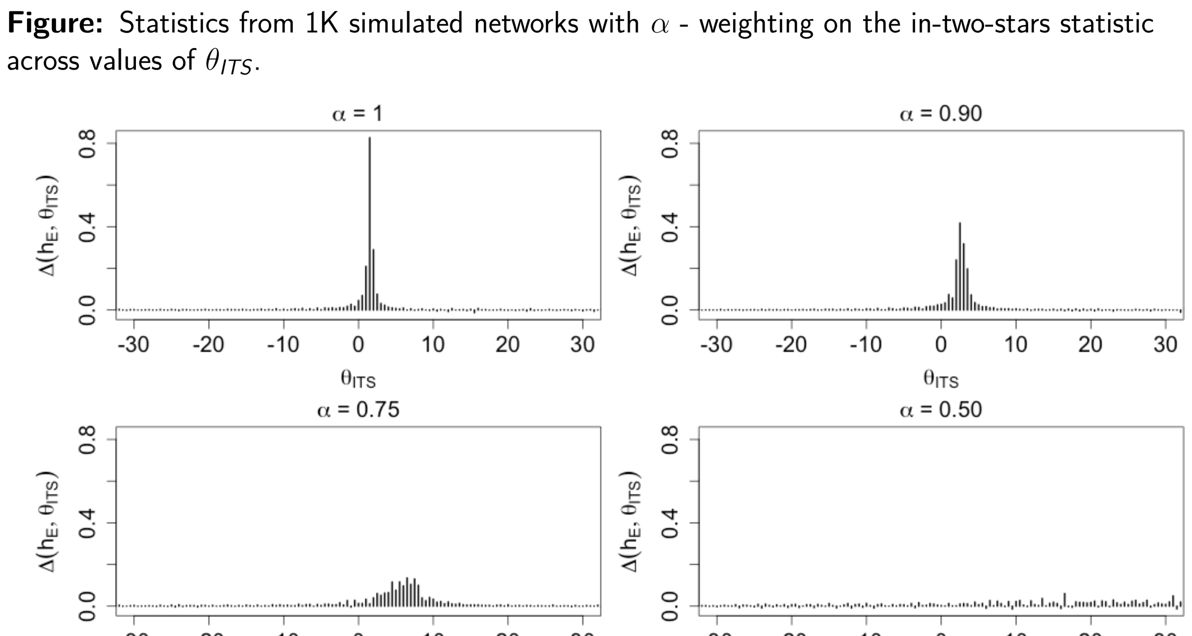


Figure: First differences of edge density statistic across θ_{ITS} for the simulated networks. Networks with weighting of $\alpha = 0.5$ witness a smooth transition of edge density across θ_{ITS}.

A General Metropolis-Hastings Procedure

The $t+1$st sample $x^{(t+1)}$ is generated as follows using a truncated multivariate normal [0,1] proposal distribution $q_\sigma(\cdot|x_{i,j}^{(t)})$:

1. For $i,j \in [n]$, generate proposal edge $\tilde{x}_{i,j}^{(t)} \sim q_\sigma(\cdot|x_{i,j}^{(t)})$ independently.

2. Set

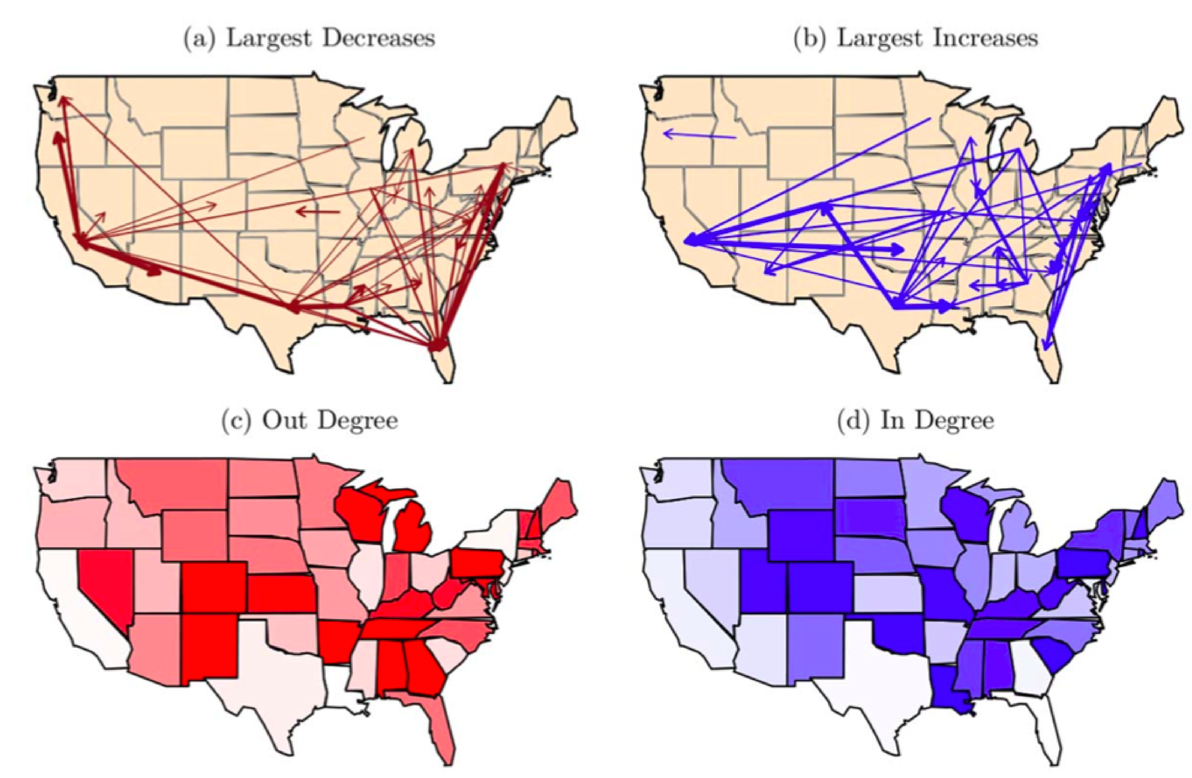
$$x^{(t+1)} = \begin{cases}\tilde{x}^{(t)} = (\tilde{x}_{i,j}^{(t)})_{i,j\in[n]} & \text{w.p.} \quad \rho(x^{(t)},\tilde{x}^{(t)}) \\ x^{(t)} & \text{w.p.} \quad 1-\rho(x^{(t)},\tilde{x}^{(t)})\end{cases}$$

where

$$\rho(x,y) = \min\left(\exp\left(\boldsymbol{\theta}'(\mathbf{h}(y)-\mathbf{h}(x))\right)\prod_{i=1}^{m}\frac{q_\sigma(x_i|y_i)}{q_\sigma(y_i|x_i)}, 1\right) \tag{4}$$

Application: U.S. Migration Data

The Data



GERGM Estimation

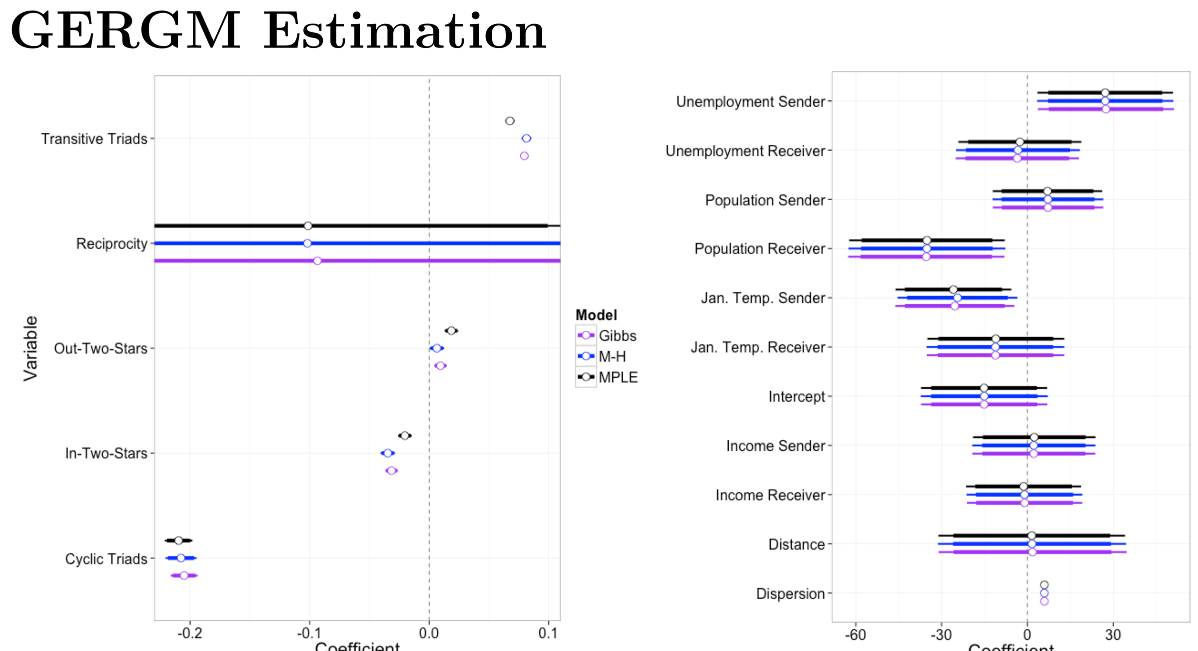


Figure: Parameter estimates for the U.S. migration network with 90 and 95% confidence intervals.

Goodness of Fit:

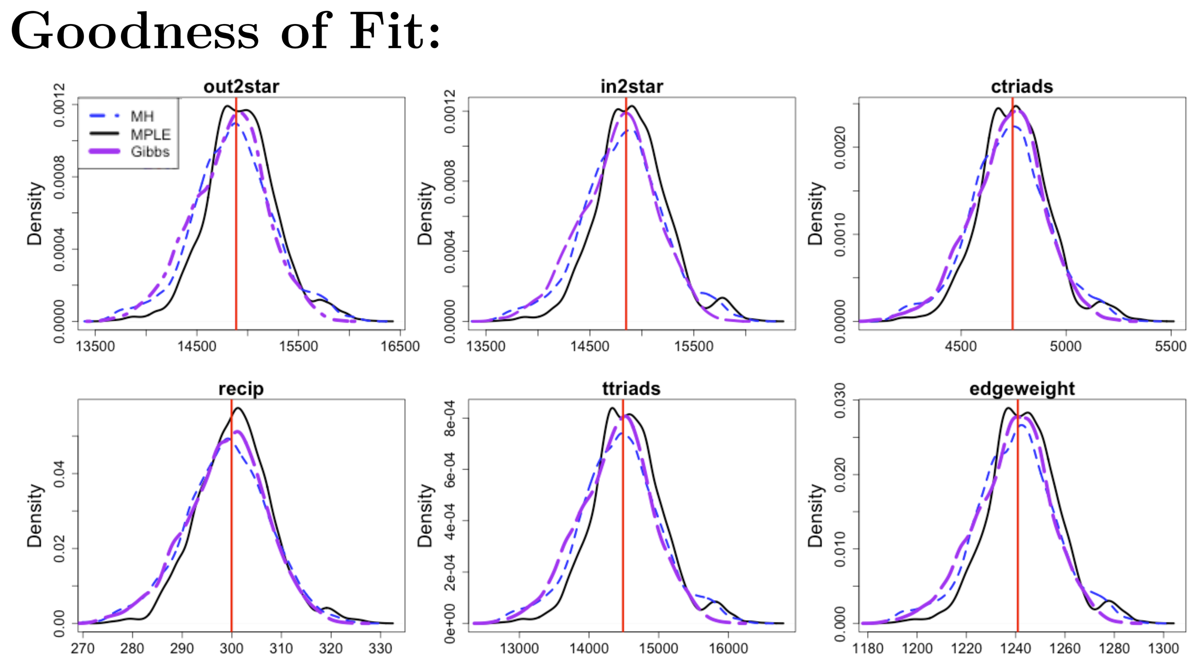


Figure: Goodness of fit evaluation of estimated models on the U.S. migration network. Statistics are calculated from 10,000 simulated networks from the fitted model.

Application: International Lending

Bank for International Settlements (BIS) aggregate (private and public) international lending volumes for 18 large economies.

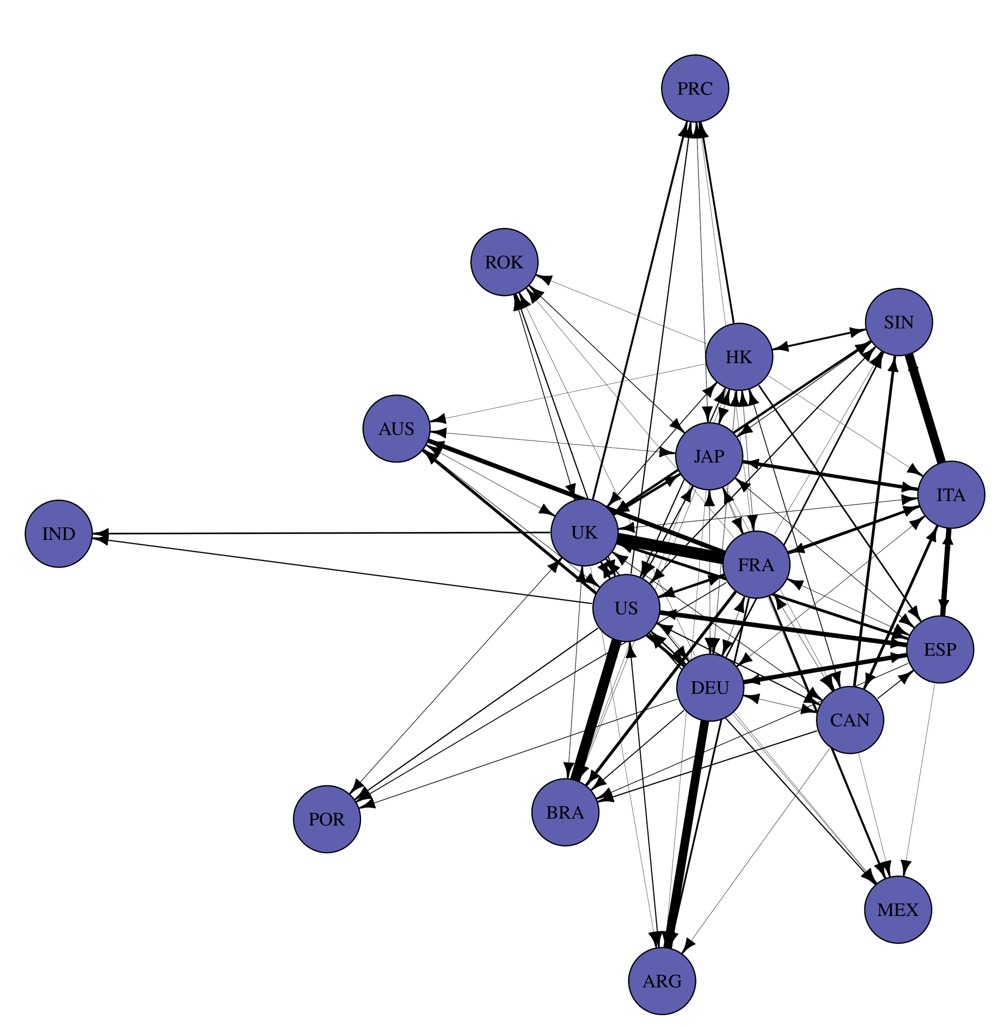


Figure: International Lending Network – 1980. ** Note that 2000 and 2005 networks have been procrustes-transformed against the 1980 network for comparability. Further note that all edges less than 0.5% of the maximum edge value ($\sim$ \$200 million) have been removed for clarity.

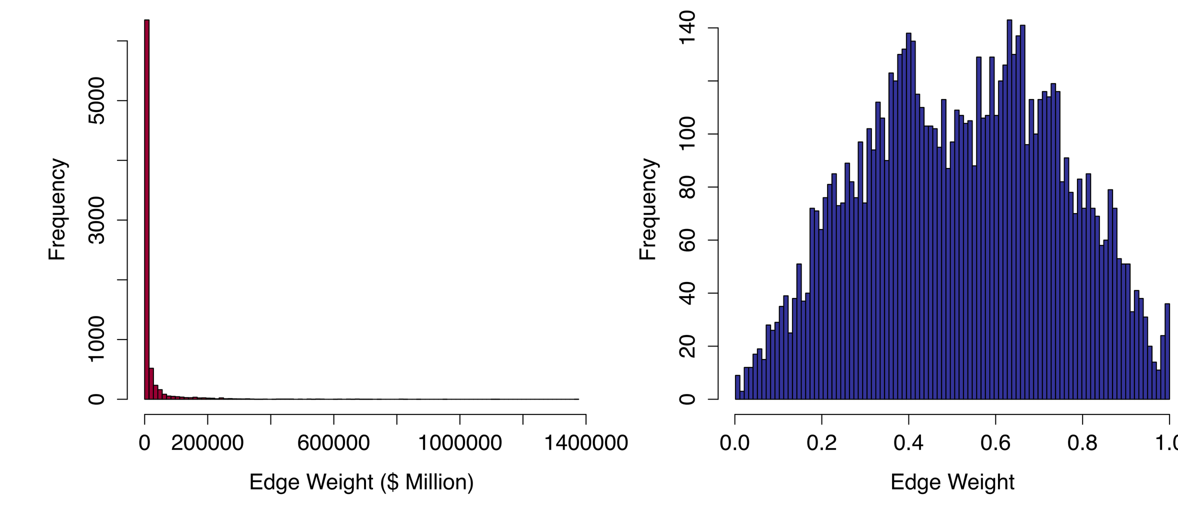


Figure: International lending volumes are heavy tailed (left panel), logging and normalizing edge values (right panel) reduces potential for degeneracy, but α weighting must be used to specify a two parameter model for these data (as determined by exhaustive grid search).

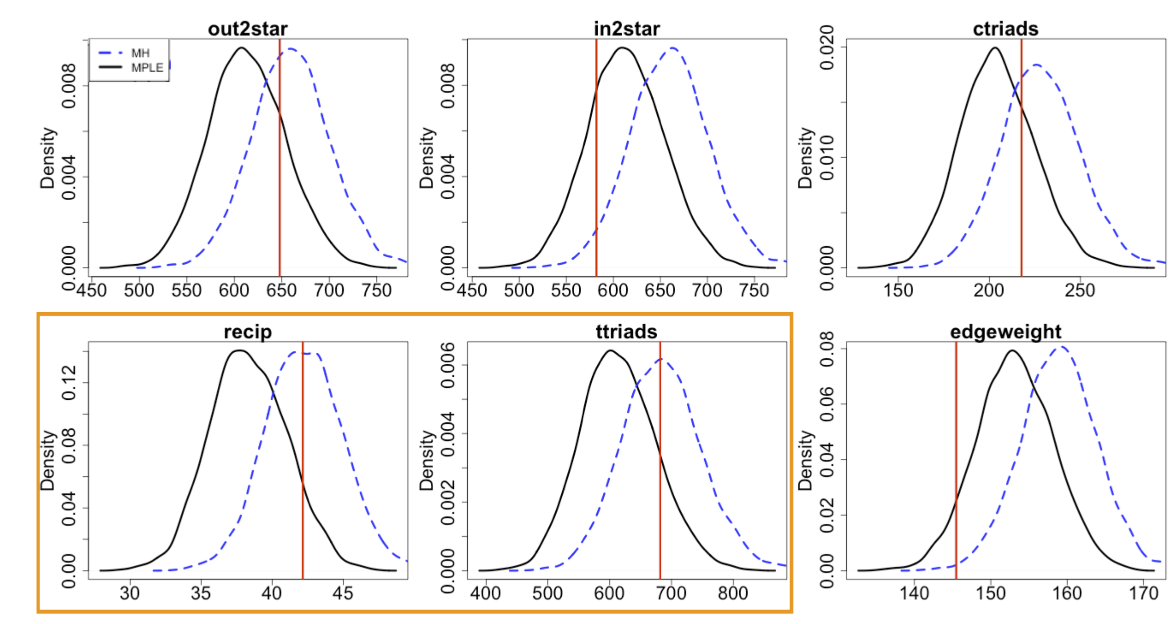


Figure: Goodness of fit evaluation of estimated model for 1980 (reciprocity and transitivity – highlighted in yellow) on the international lending network generated using 10K simulated networks from the fitted model (with $\alpha = 0.1$). Results indicate positive significant reciprocity effect.

Extension: Lending Over Time

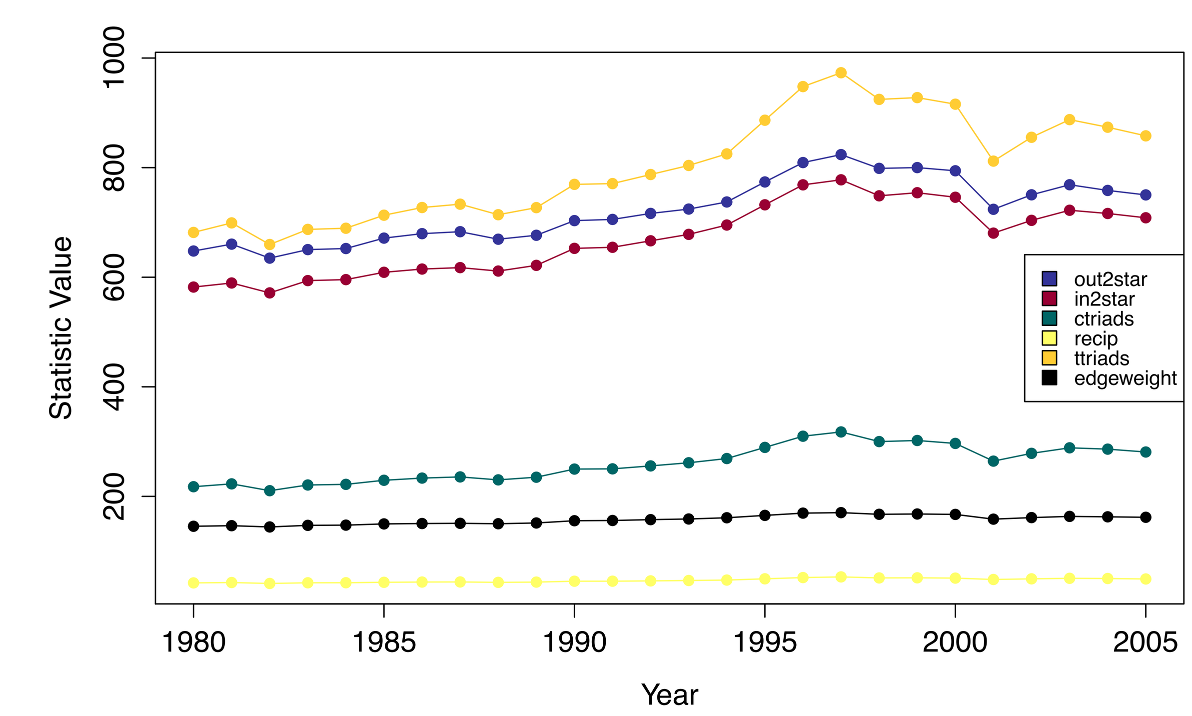


Figure: Time series plots of log-normalized network statistics (no α weighting).

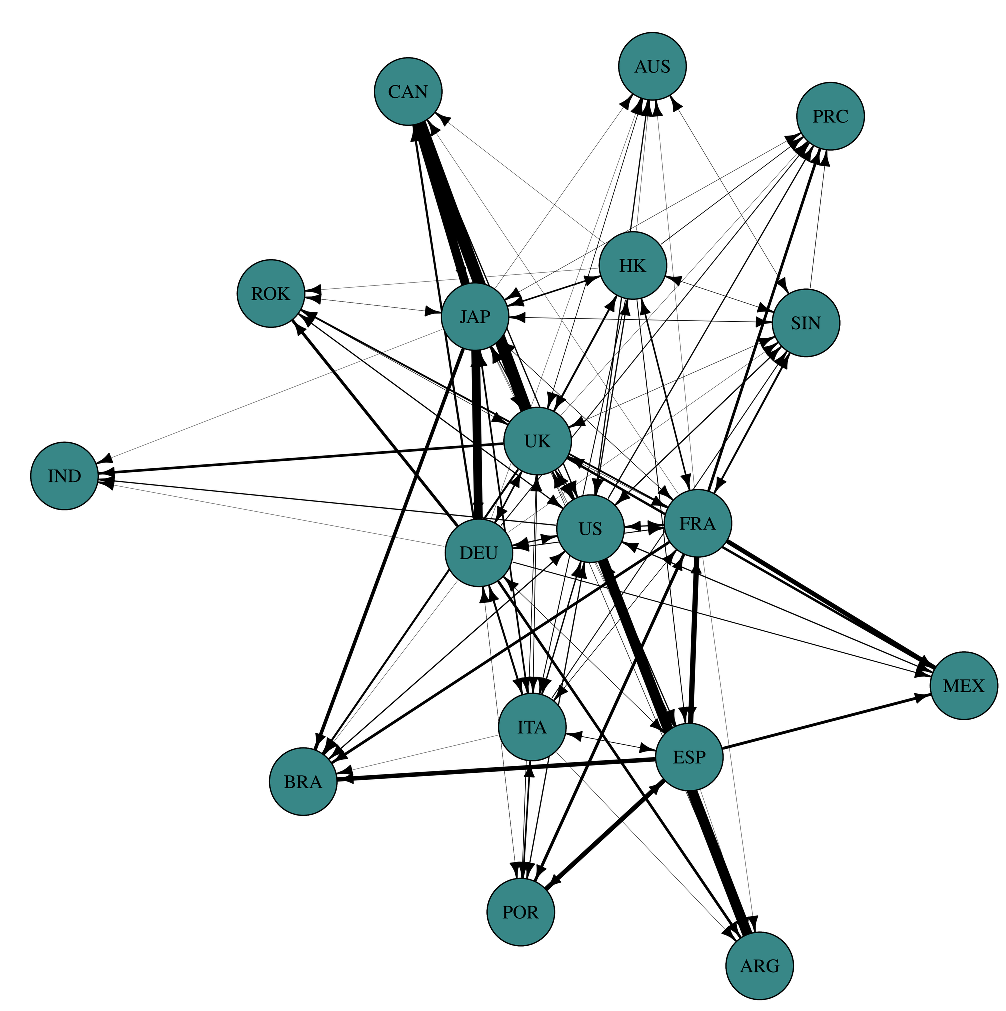


Figure: International Lending Network – 2000

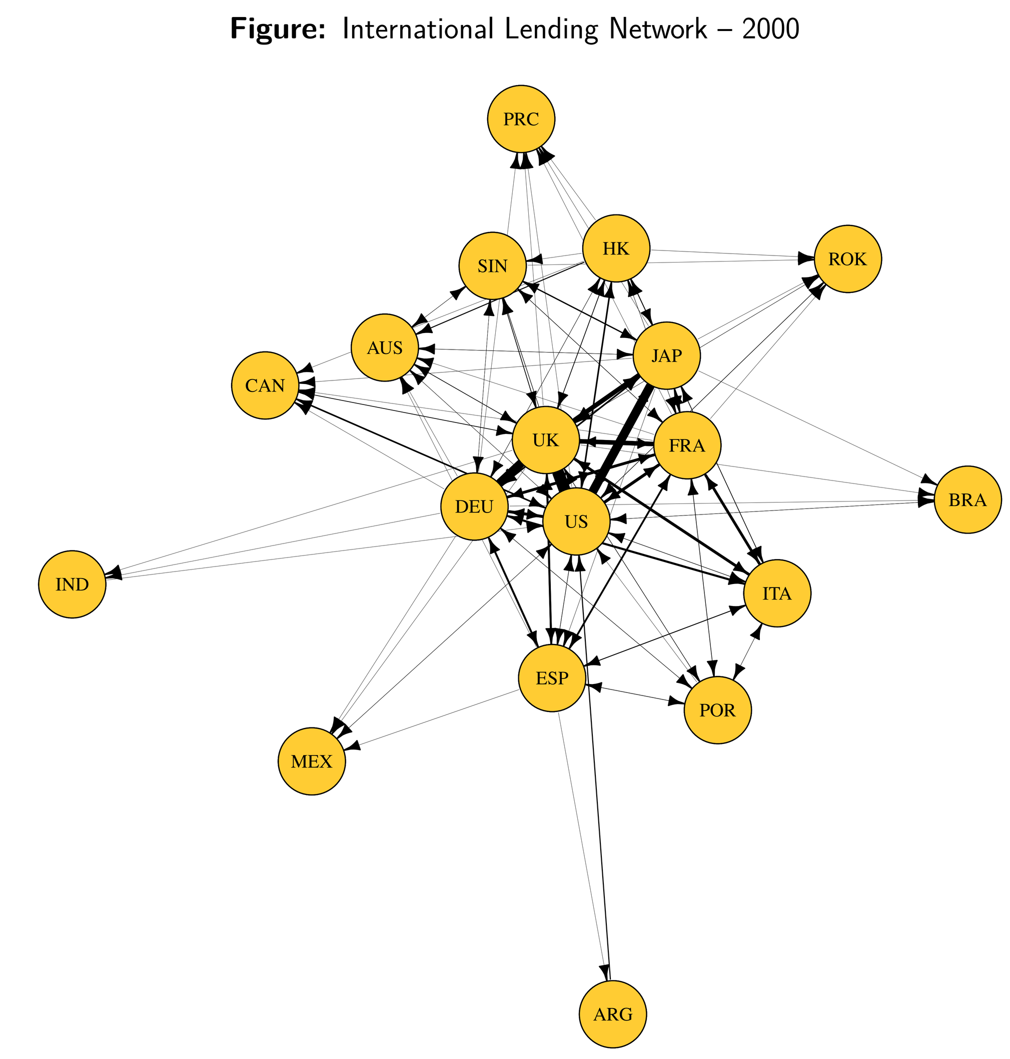


Figure: International Lending Network – 2005

UMassAmherst

Previous Work:

Desmarais, B. A., & Cranmer, S. J. (2012). "Statistical inference for valued-edge networks: the generalized exponential random graph model". *PloS One.* http://dx.plos.org/10.1371/journal.pone.0030136

S.B. was partially supported by NSF-DMS grants 1105581 and 1310002 and SES grant 1357622. B.D. was partially supported by US NSF grant SES-1357606 and by NSF grant CISE-1320219.