Intermediate Social Network Theory

Matthew J. Denny
University of Massachusetts Amherst

March 22, 2015

From Description to Theory

- We have developed a **vocabulary for describing networks**.
- **Common patterns?**
- What **processes** underlay observed structure?
- **Structure $\Longleftrightarrow$ outcomes?**

Identifying A Relational Theory

Why Do We Care?

- What is your theory a theory of?
- Do you really need a network representation?
- Adding degree centrality to a regression – **NOT** a relational theory.
- **Occam's Razor** – the simplest explanation is best.

Positional Theories

Definition 1: *A* ***positional theory*** *is a theory about how the positions of nodes in a network affect their individual or group level outcomes, or how their positions in the network change over time.*

Positional Theories

- People with more friends have more social capital.
- People with more sexual partners are more likely to have HIV.
- Senators with more connections are more powerful.
- Network centrality is related to some outcome (degree, betweenness, closeness).

Relational Theories

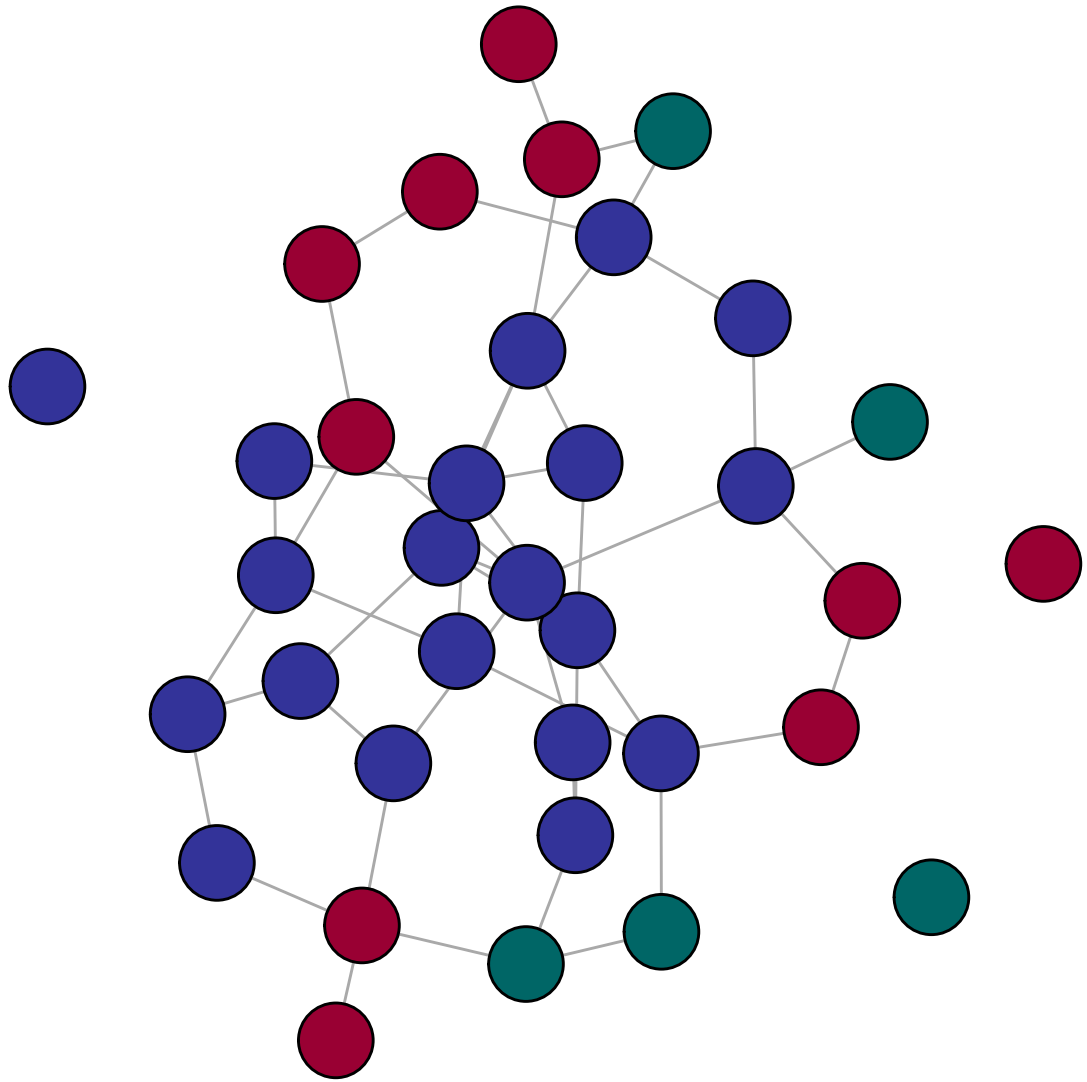
Definition 2: *A* ***relational theory*** *concerns the structure of the connections between nodes in which the state of a node is related to ties that do not involve that node.*

Relational Theories

- Small world networks are fault tolerant.
- Friendship networks between school children are race and gender homophilous.
- The international economic sanctions network is intransitive.
- Women are excluded from the ‘locus of control’ in organizations.

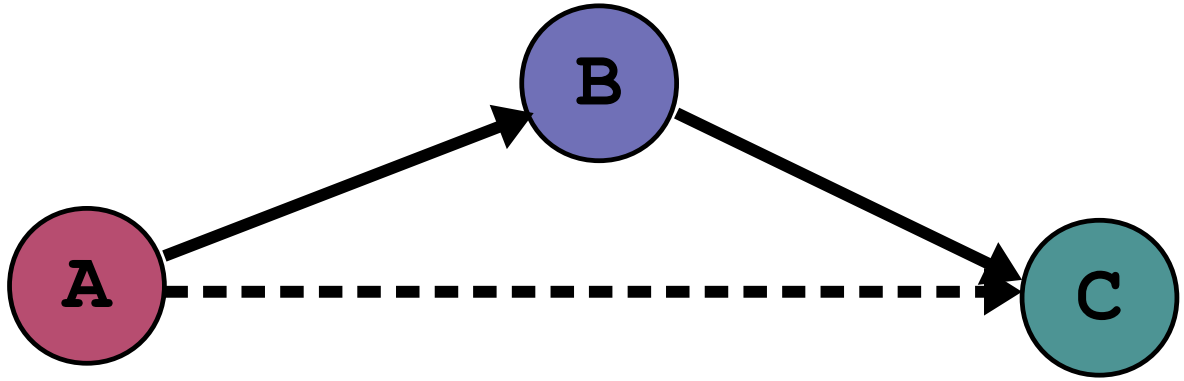
Building Blocks

The Network



Transitivity and Reciprocity

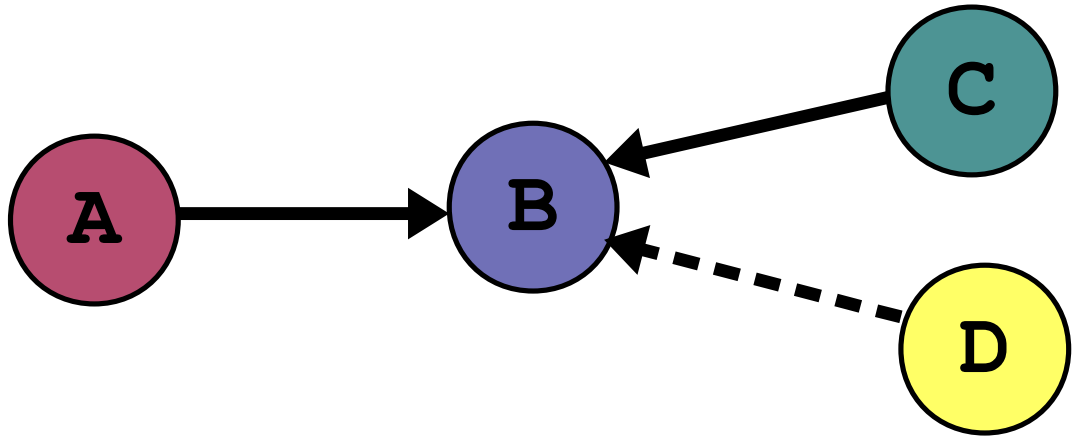
Transitivity – Clustering



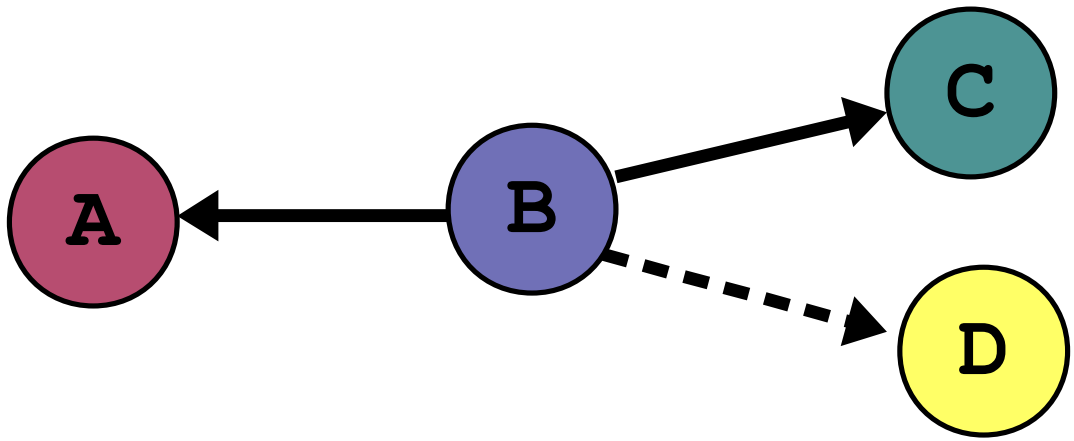
Reciprocity – Collaboration, Stability

Preferential Attachment

Popularity – Power, Path Dependence

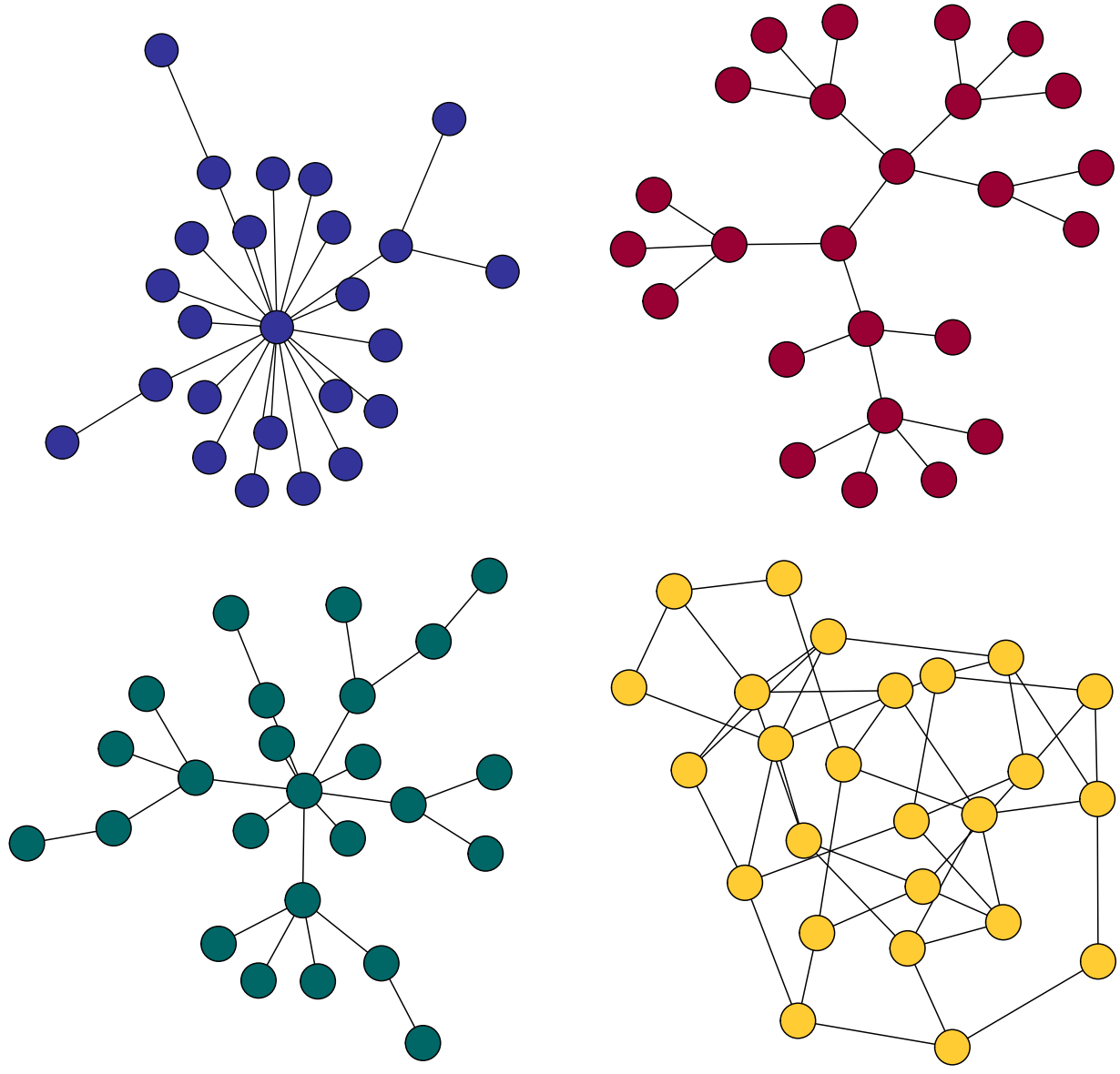


Sociality – Economies of Scale



Ill-Defined Concepts

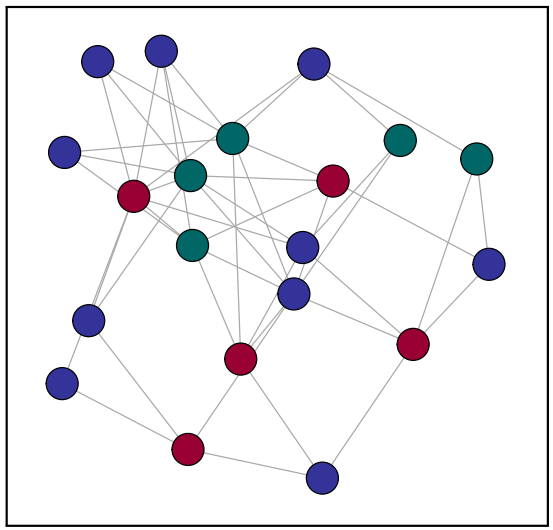
What is Hierarchy?

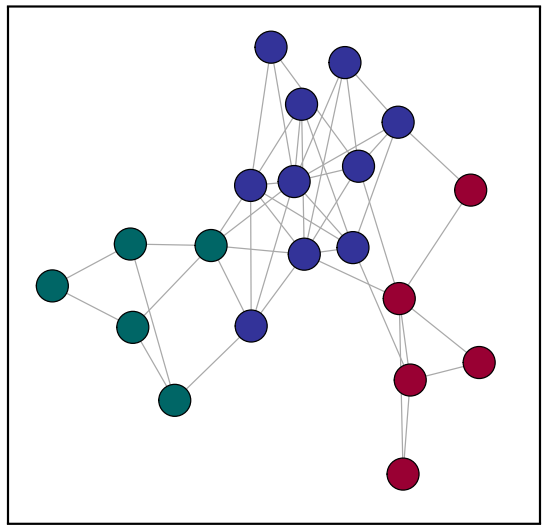


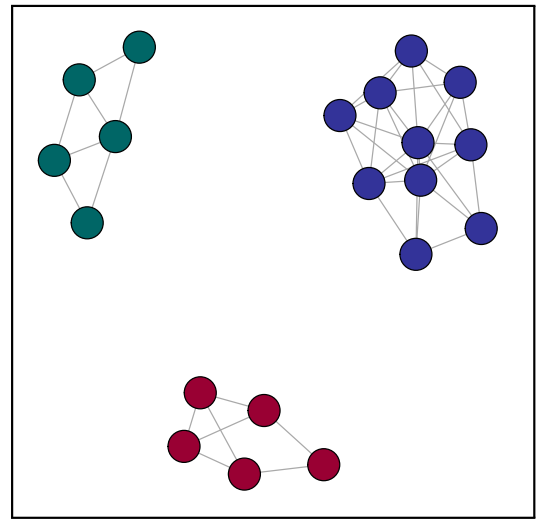
Hierarchy

- Physicists say it is a wide tree.
- Is it defined on "positions" or structure?
- **Width** and **Depth**.
- Is hierarchy a useful concept?

Compartmentalization







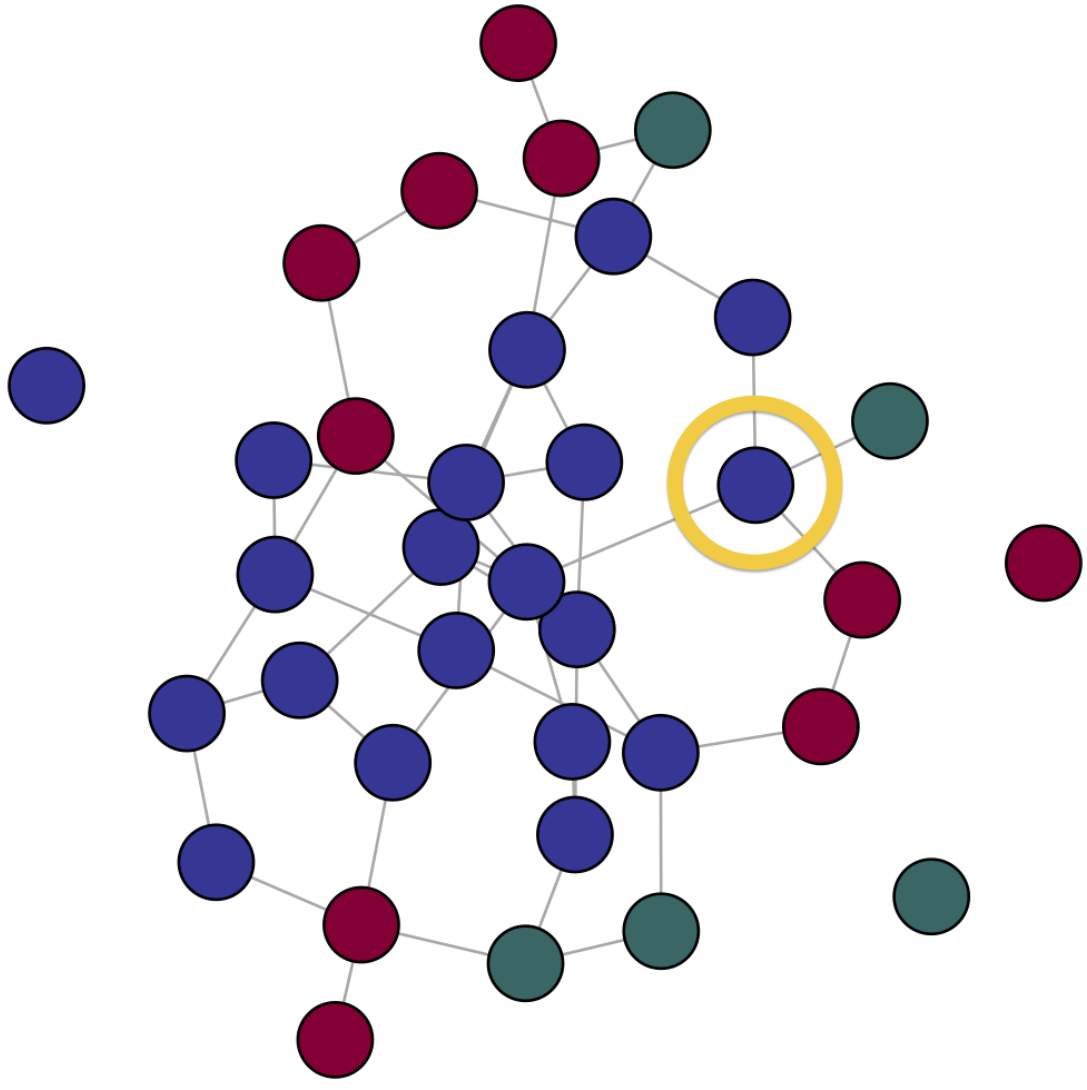
http://arxiv.org/abs/1407.2854

Levels of Analysis

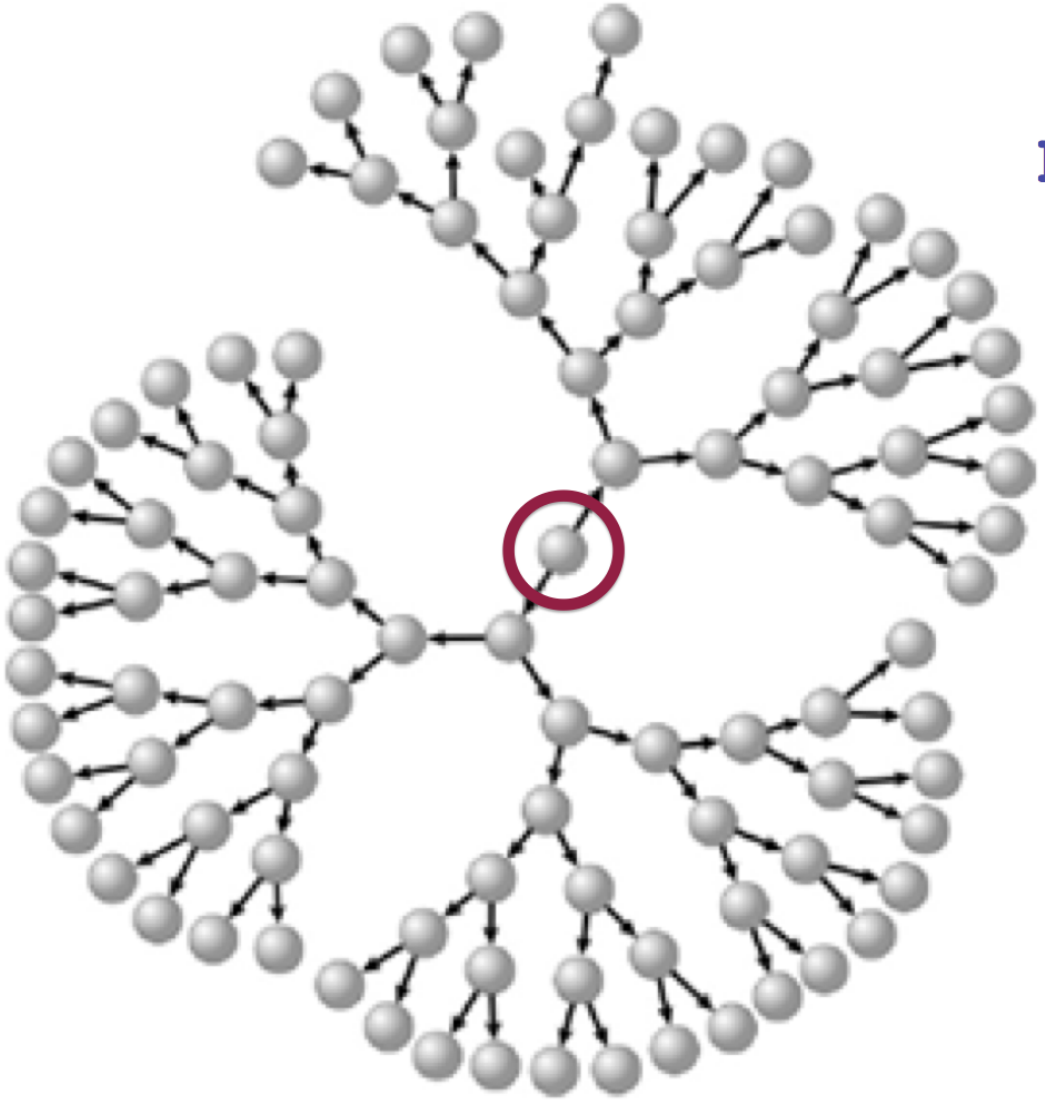
Levels of Analysis

- The **systems level** concerns characteristics of the entire network.
- The **group level** concerns differences and similarities in the network structure within, between and across groups.
- The **node level** concerns the patterns connections by individual nodes.

Levels of Analysis



Example: Information Diffusion



Efficiency

Fault Tollerance

Specifying A Relational Theory

A Rule of Three

A Rule of Three: *Researchers should subset (through matching or experimental design) their data until a regression with only three (at most) covariates explains the data.*

Achen, C. H. (2002). Toward A New Political Methodology: Microfoundations and ART. Annual Review of Political Science, 5(1), 423450. doi:10.1146/annurev.polisci.5.112801.080943

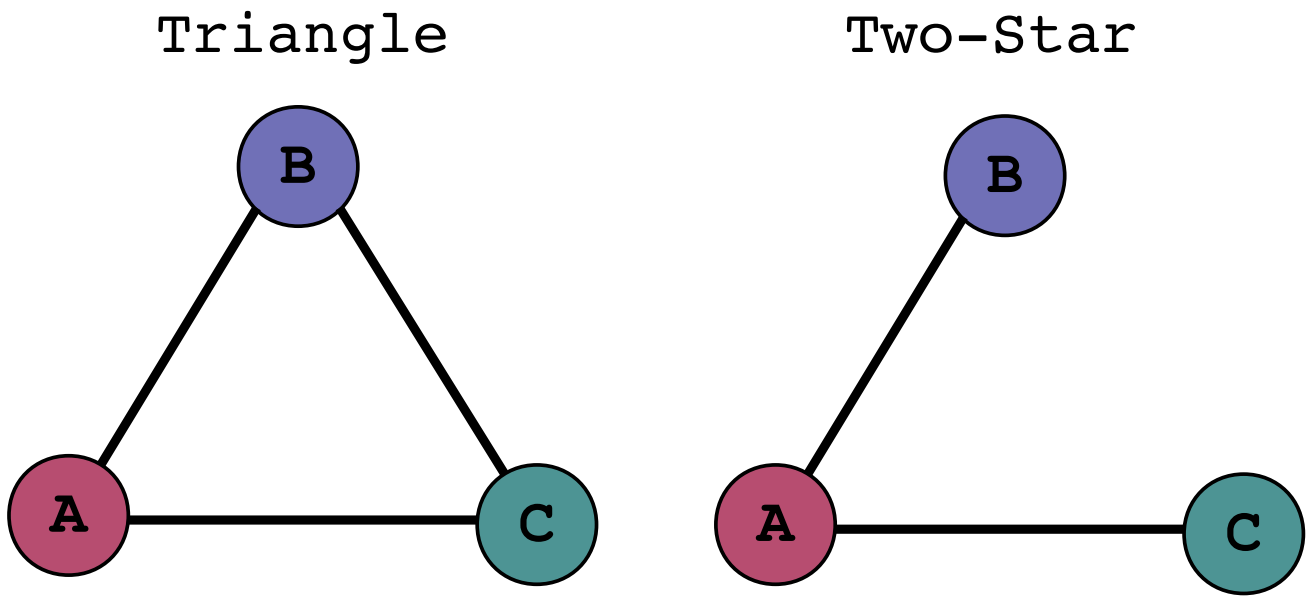
What about in the relational context?

A Relational Rule of Three: *A relational theory should seek to explain the observed network structure at all three levels of analysis, and should be parsimonious.*

Parsimony vs. Completeness

A Note on Observational Data

- How do we measure network properties?
- We count:



Multicolinearity and Omitted Variable Bias

- **Multicolinnearity** – a motivation for A.R.T.
- If counts are too highly correlated – inflated standard errors, sign switching.
- **Omitted Variable Bias**
- Multicolinearity will exacerbate, leading to biased estimates.

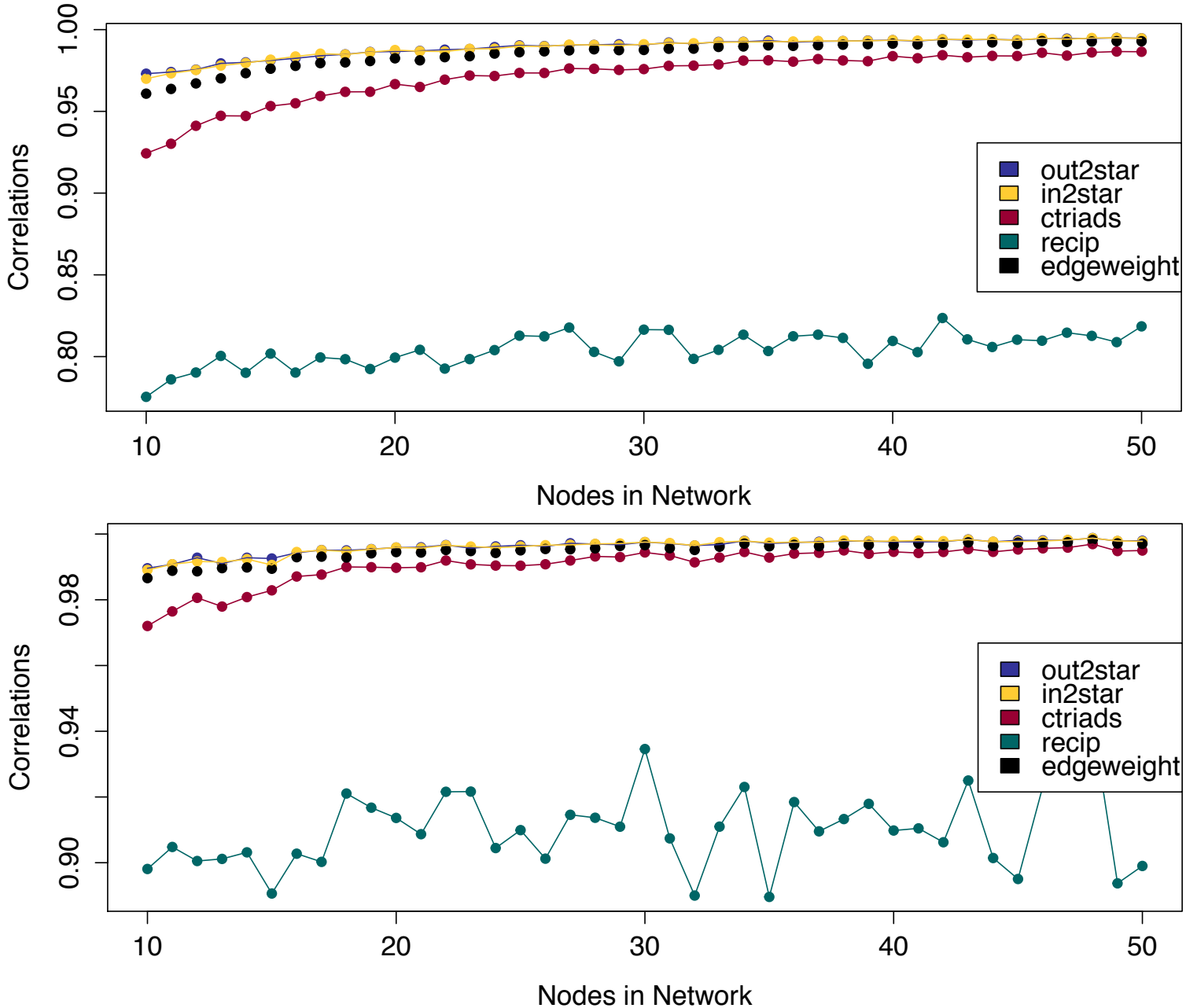
The Exponential Random Graph Model

- Let Y be a n-node network
- An ERGM is specified as:

$$\mathcal{P}(Y, \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}' \mathbf{h}(Y)\}}{\sum_{\text{all } Y^* \in \mathcal{Y}} \exp\{\boldsymbol{\theta}' \mathbf{h}(Y^*)\}}$$

- $\boldsymbol{\theta}$ is a parameter vector
- $\mathbf{h}(Y)$ is a vector of statistics on the network
- Object of inference: the probability of Y among all possible permutations of Y given the network statistics.

Null Model: High Correlation



Solution?

- **Develop A Strong Theory!**
- Theory is highly parsimonious + complete – no theoretical problem.
- Nuance vs. Interpretability
- In practice network models are tricky, may not be able to estimate.

Example – Beyond "Gravity" in International Trade

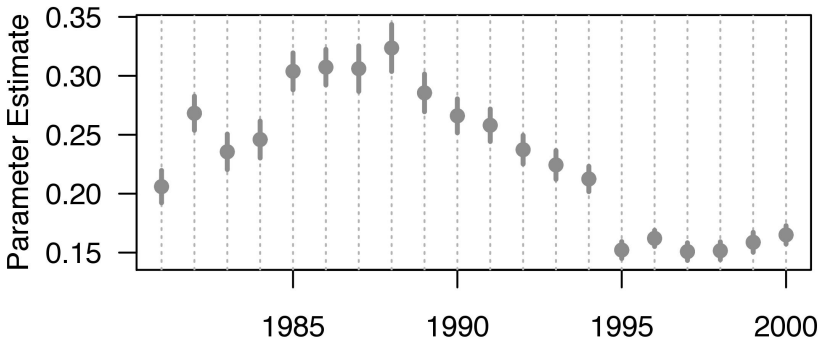
- Ward & Hoff. "Persistent Patterns of International Commerce". *Journal of Peace Research*, 2007.

- Yearly data on international trade flows from the UN Commodity Trade Statistics Database (1980-2001)

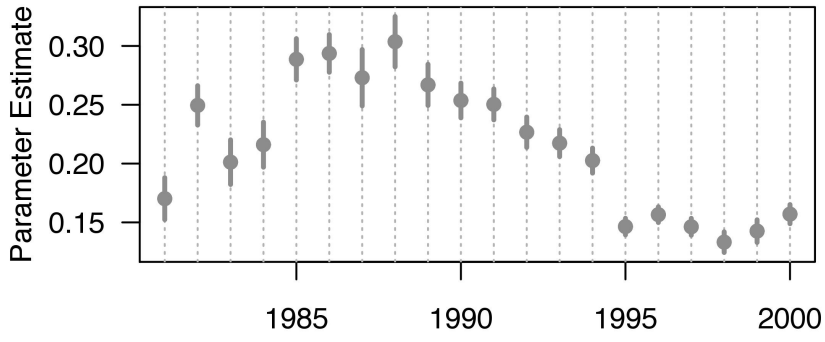
- What is our Theory?

(Generalized) ERGM Results

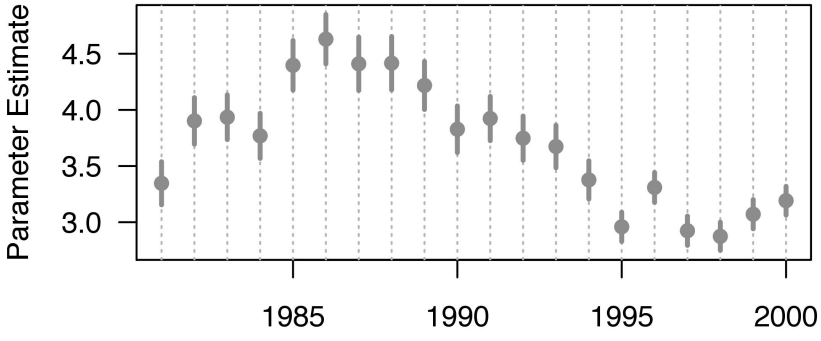
Sociality – (Exporters)



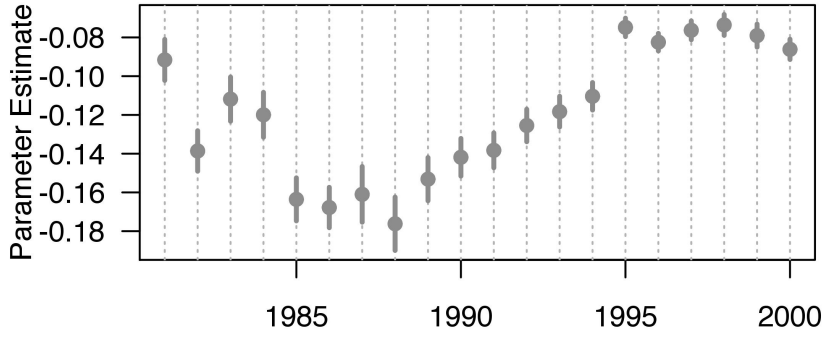
Popularity – (Importers)



Reciprocity



Transitivity



Unidentified Models

The Latent Space Model

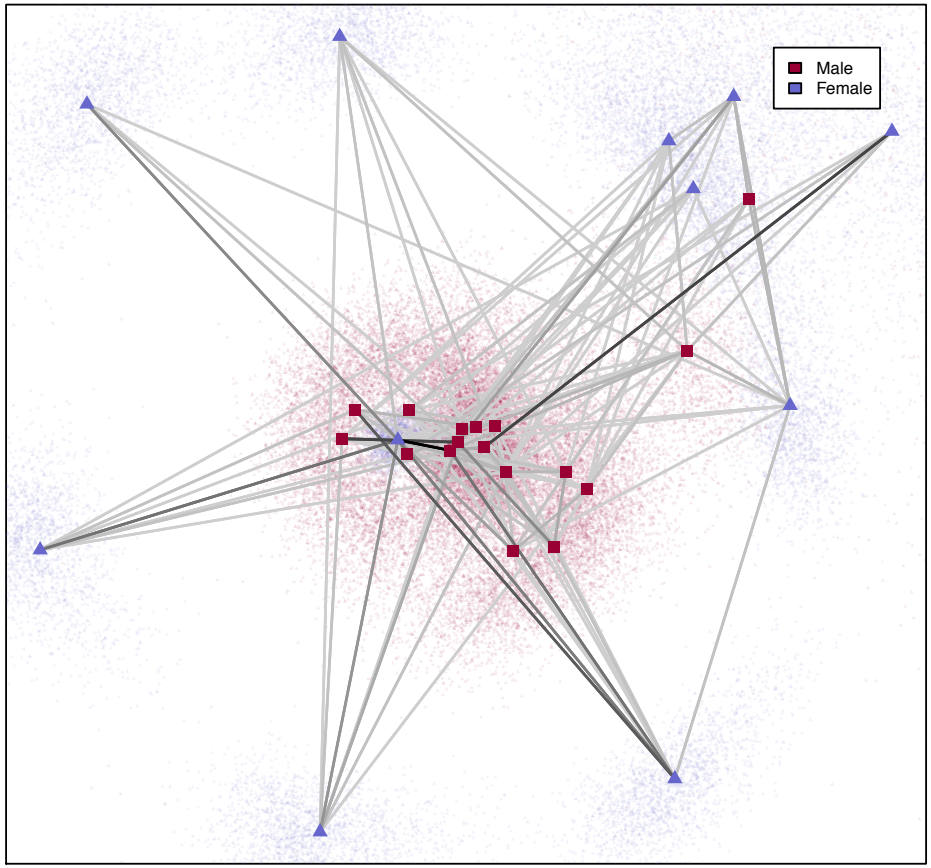
Under the latent space model, the log odds of a tie between two nodes i and j is defined as:

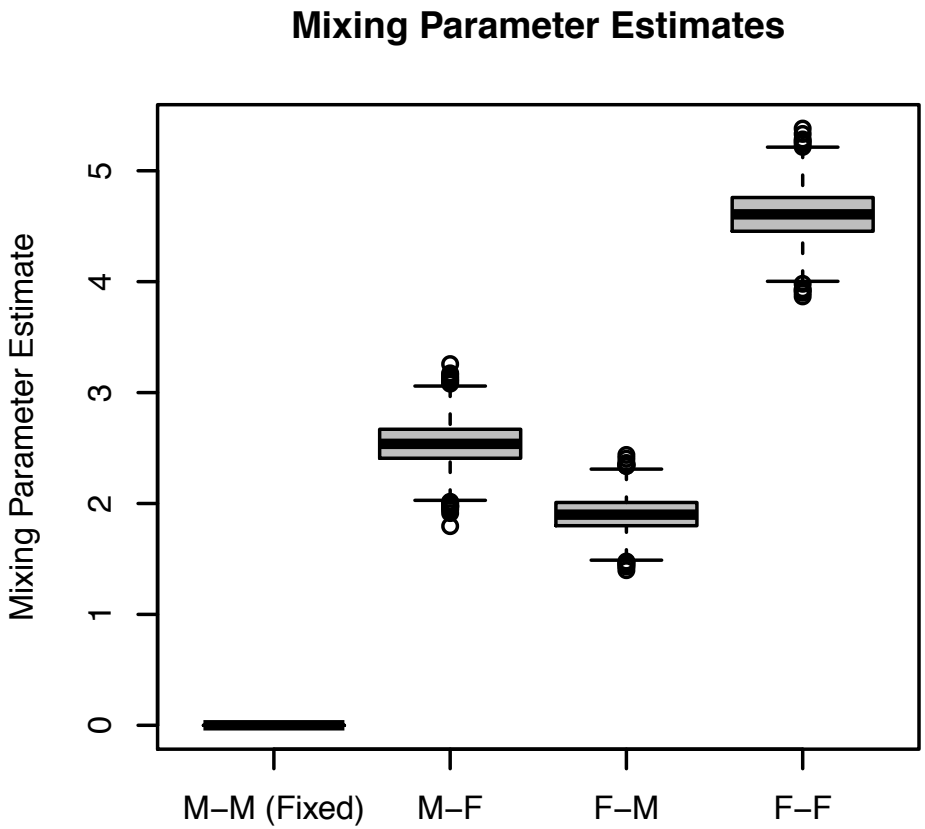
$$\eta_{i,j} = \text{log odds}(y_{i,j} = 1 | z_i, z_j, x_{i,j}, \alpha, \beta) \quad (1)$$

$$= \alpha + \beta' x_{i,j} - |z_i - z_j| \quad (2)$$

- α is an intercept term
- β is a vector of dyad specific covariate effects
- $|z_i - z_j|$ is the euclidean distance between nodes

Example: Gender Mixing





How Do We Interpret?

- Women are more likely to email women given network structure.
- Women are on the periphery in the network – less likely to communicate.
- Intercept and spread of latent positions
- Class of models is only weakly identified through an informative prior.

Relational Processes

Relational Processes

- Why does the network look the way it does?
- How will the network grow?
- How relatively important are different processes shaping the network?
- Many processes can lead to same observed structure.

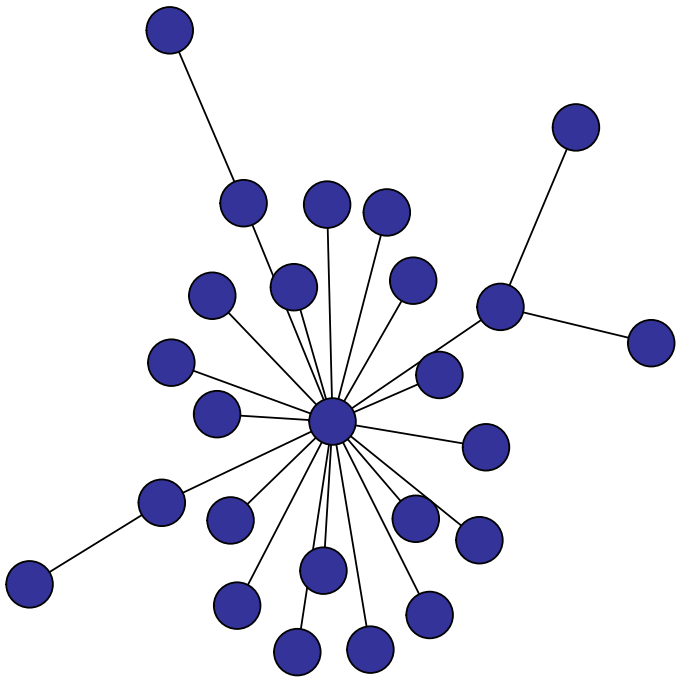
Example – Global Finance

> *In **hierarchical structures, one country stands at the center** of the system, and other states are on the periphery. Hence influence is unevenly distributed between a central hegemon and everyone else. In **flat structures, no country is substantially more central than another**. Hence influence is more evenly balanced between countries. Thus hierarchical and flat network topologies generate the same distributions of influence that existing IR structure-based models emphasize. [p.137]*

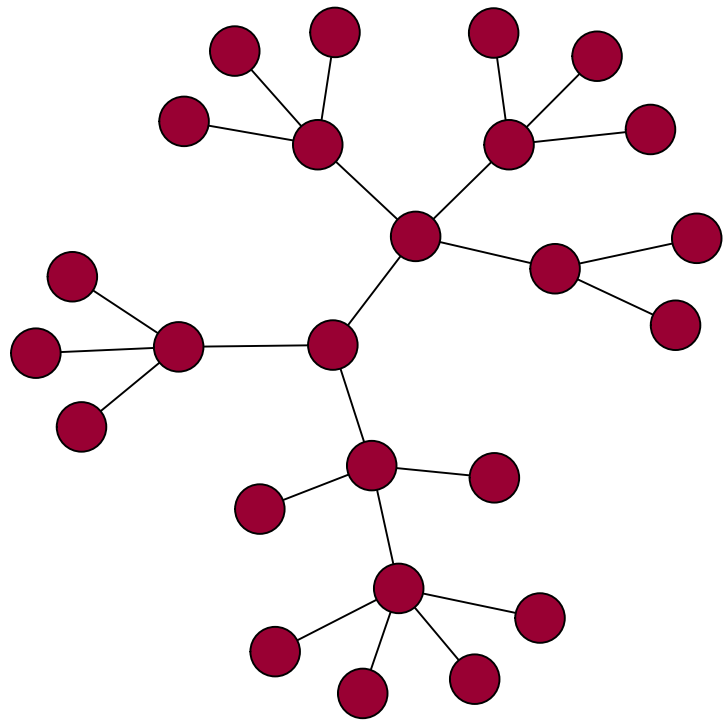
Oatley et al. (2013). The Political Economy of Global Finance: A Network Model. *PS*.

By What Process?

“Hierarchical” (Star)



“Flat” (Tree)

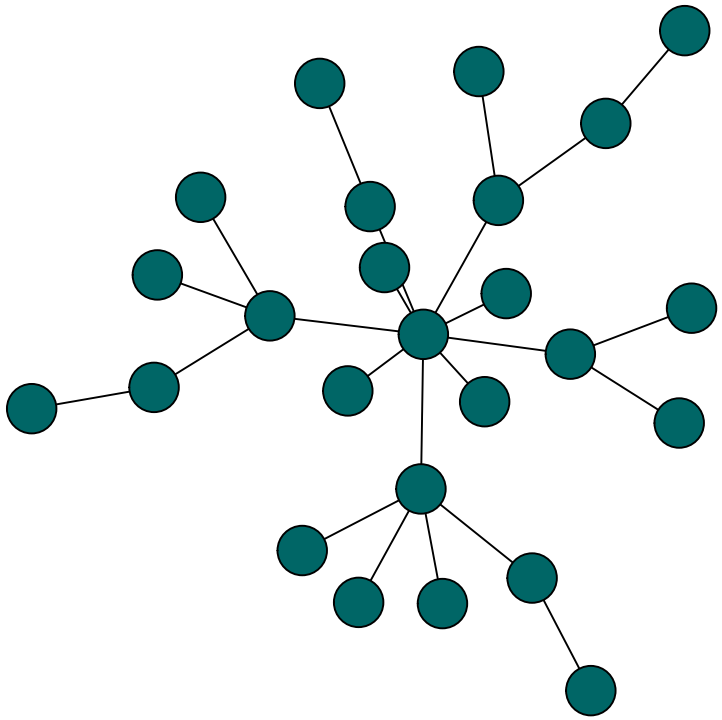


A Relational Theory of Global Finance

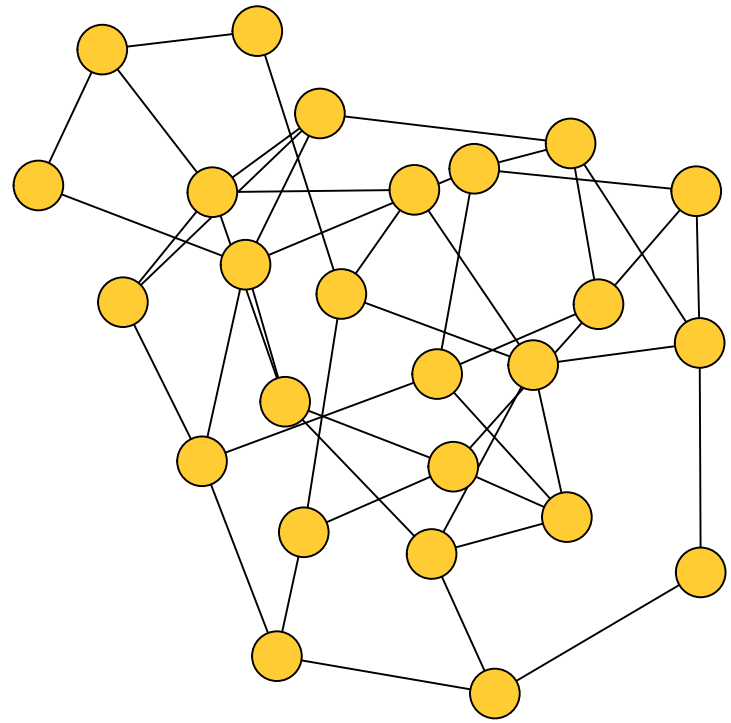
- **Hierarchy:** *Fitness with Preferential Attachment (FPA)* – (authors actually suggest this as the process)
- Degree centrality not important – **position** is.
- **Flat:** *Erdos-Renyi random graph model*
- If we think of financial crises as diffusion processes then Hierarchical structure is better.

Specified Process

Preferential Attachment



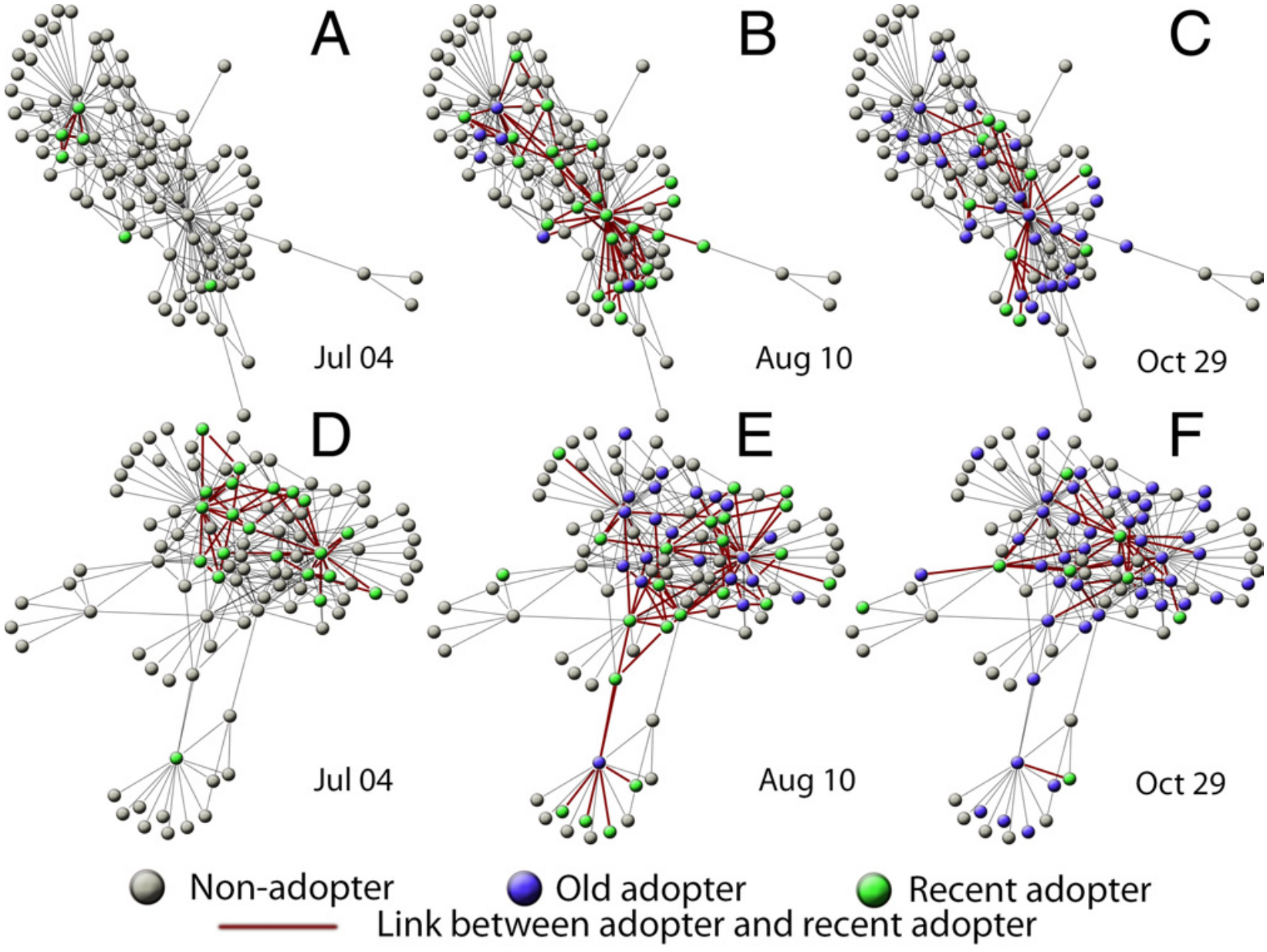
Random



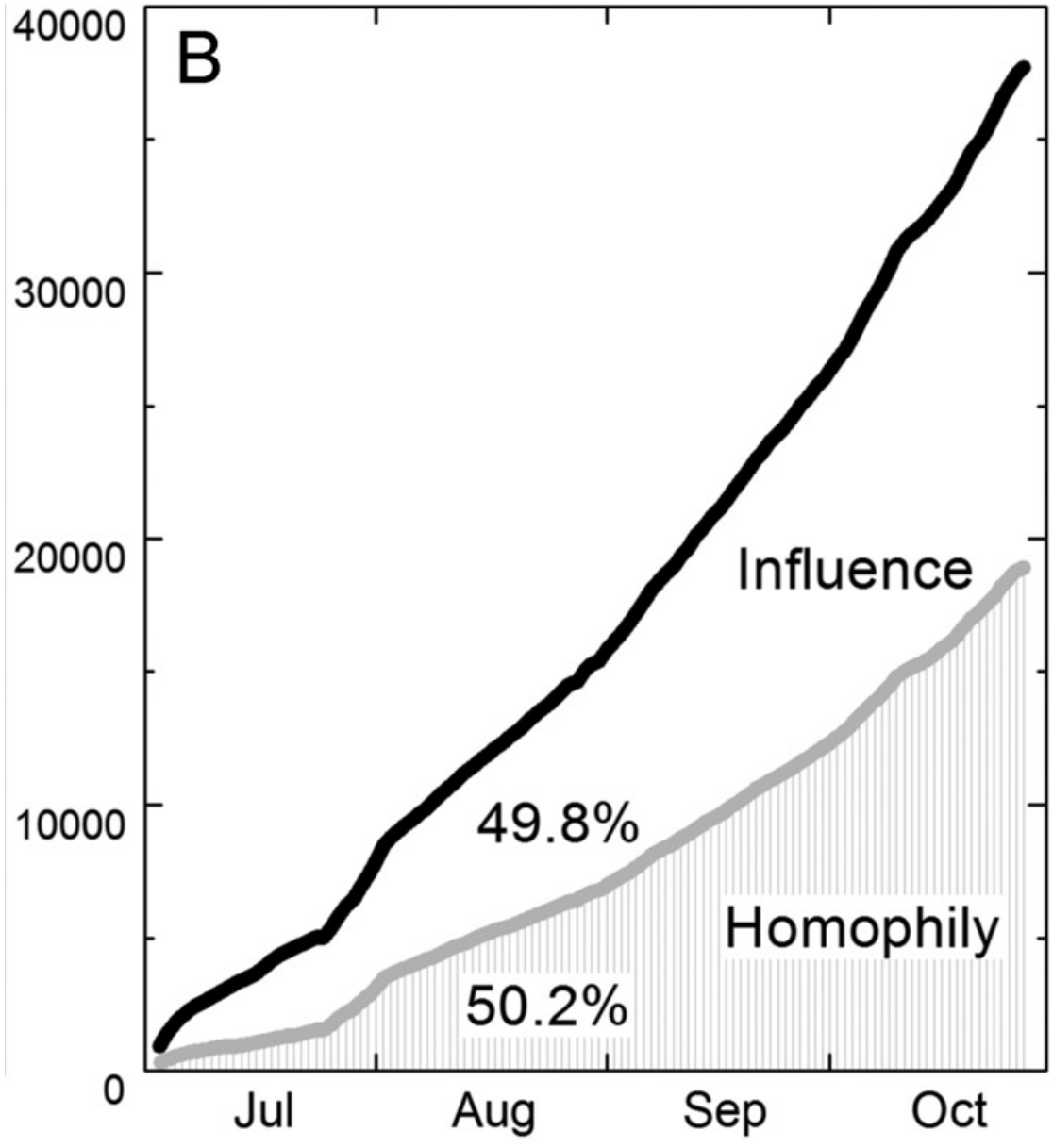
Influence and Homophily

- One of the big areas of research in network dynamics.
- Is smoking passed on to friends or do people who smoke just hang out with smokers?
- Hard to distinguish, can use experiments.
- Aral et al. (2009). "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks". *PNAS*

Yahoo Go!


A
B
C
Jul 04
Aug 10
Oct 29
D
E
F
Jul 04
Aug 10
Oct 29
Non-adopter
Old adopter
Recent adopter
Link between adopter and recent adopter

Yahoo Go!


B
40000
30000
20000
10000
0
Jul
Aug
Sep
Oct
Influence
49.8%
Homophily
50.2%

Participation Time!