Lecture 8:
Triadic census, p*/ERGM, and MTML

Noshir Contractor

Jane S. & William J. White Professor of Behavioral Sciences

Professor of Ind. Eng. & Mgmt. Sciences, McCormick School of Engineering
Professor of Communication Studies, School of Communication &
Professor of Management & Organizations, Kellogg School of Management,
Director, Science of Networks in Communities (SONIC) Research Laboratory
nosh@northwestern.edu

NORTHWESTERN
UNIVERSITY

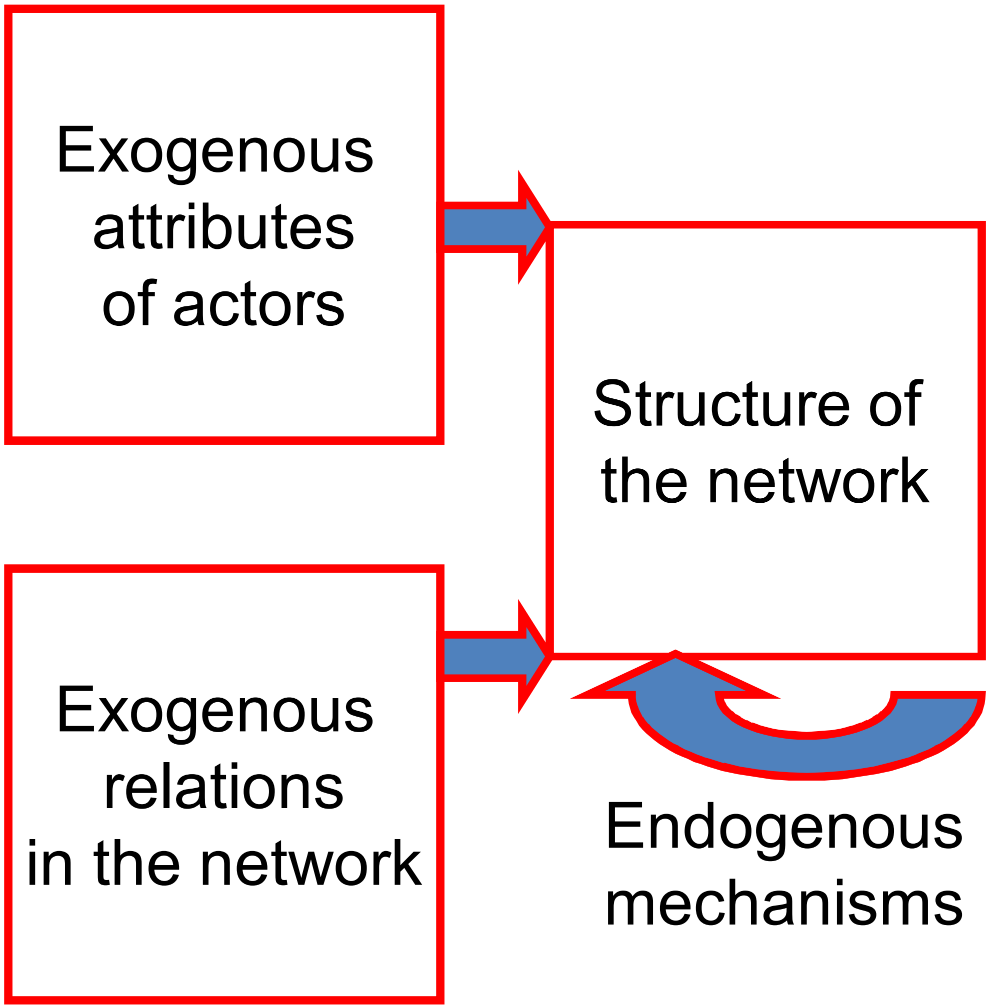
Multi-level, multi-theoretical models

- Networks are already multi-level (nodes, attributes, dyads, triads, etc.)
  - Only one level of analysis possible, thus only one theory tested
  - Structure influences attributes which influences structure
  - Need to control for exogenous and endogenous influences
- Specifying multiple parameters in ERGM allows simultaneous testing of multiple units of analysis
  - Evaluate relative significance, effect size, effect on model fit

Why do actors create, maintain, dissolve, and reconstitute network links?

- Theories of self-interest
- Theories of mutual interest and collective action
- Theories of social and resource exchange
- Theories of contagion
- Theories of balance
- Theories of homophily
- Theories of proximity
- Theories of co-evolution

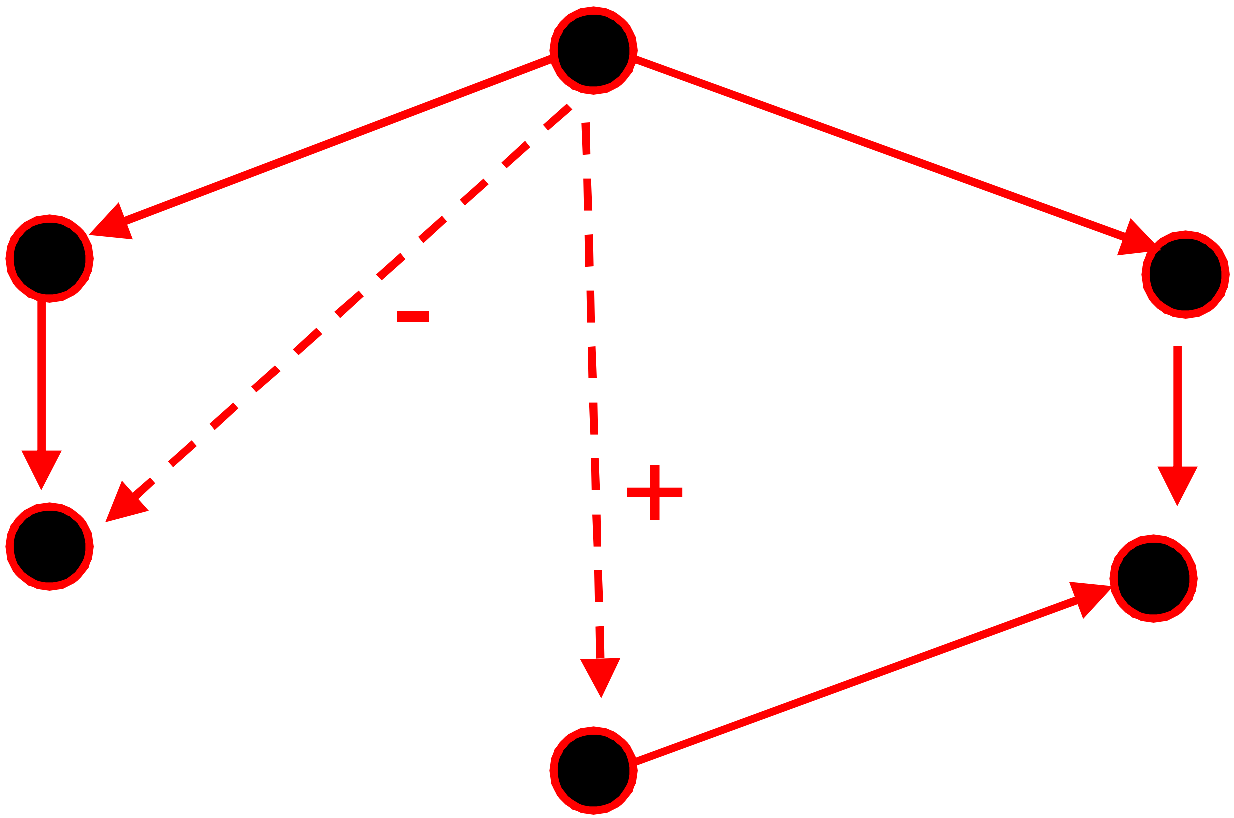
CMDR Structuring Processes



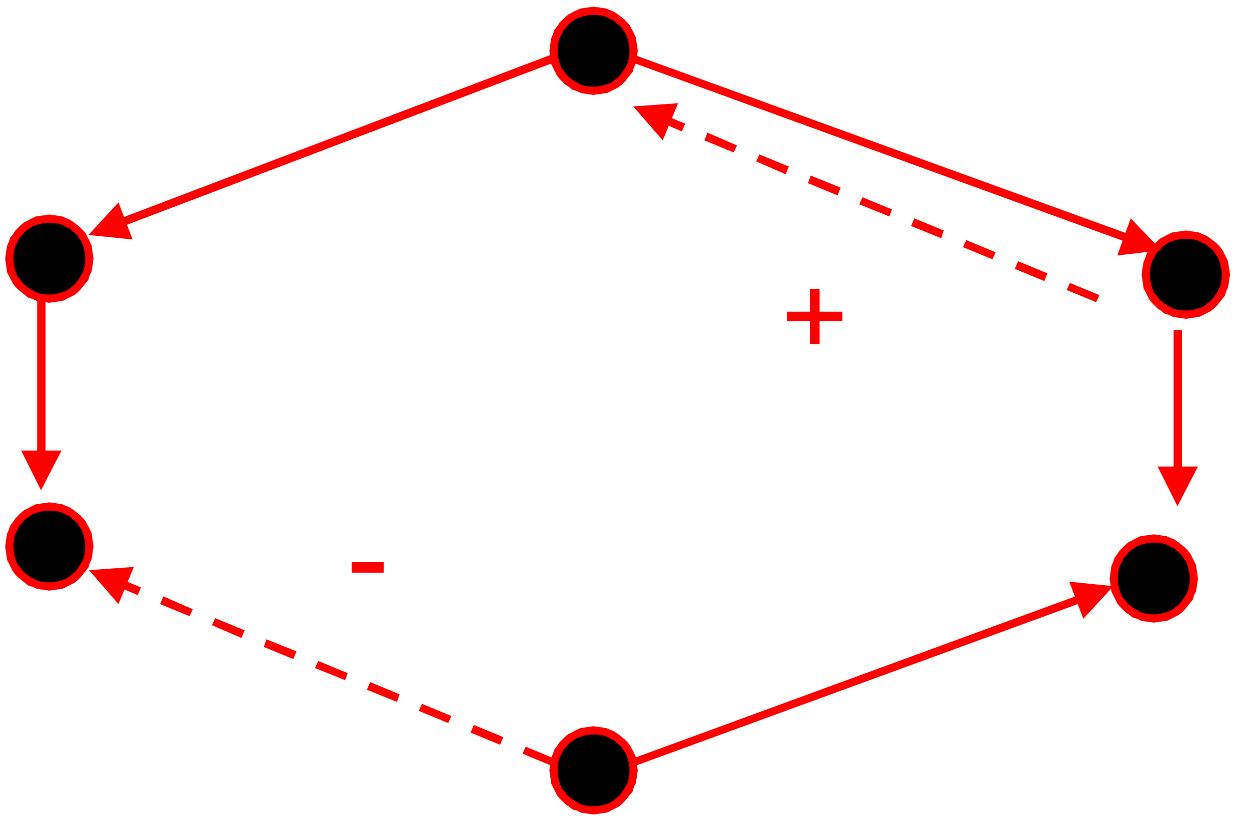
Network is structured by itself (Endogenous)

- Actor level
- Dyad level
- Triad level
- Subgroup level
- Global level

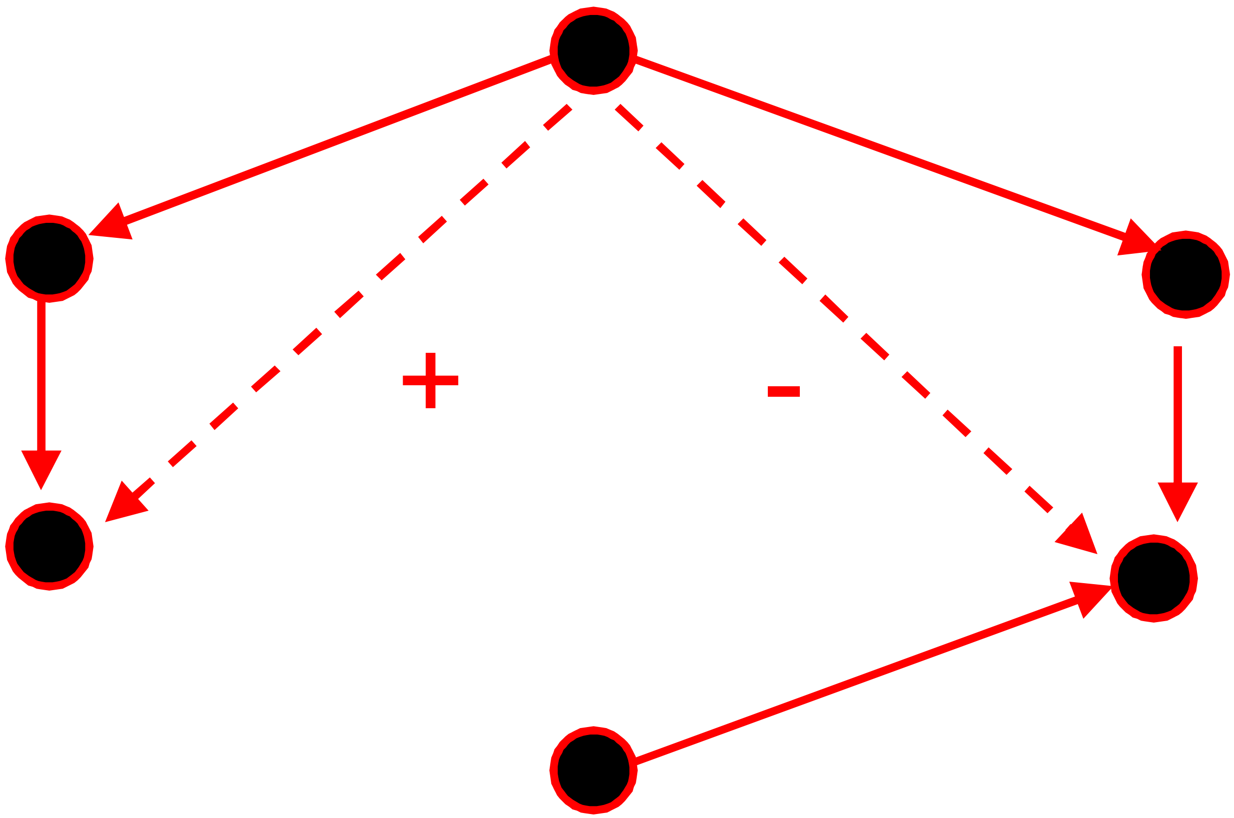
Endogenous Actor Level: Theory of Structural Holes



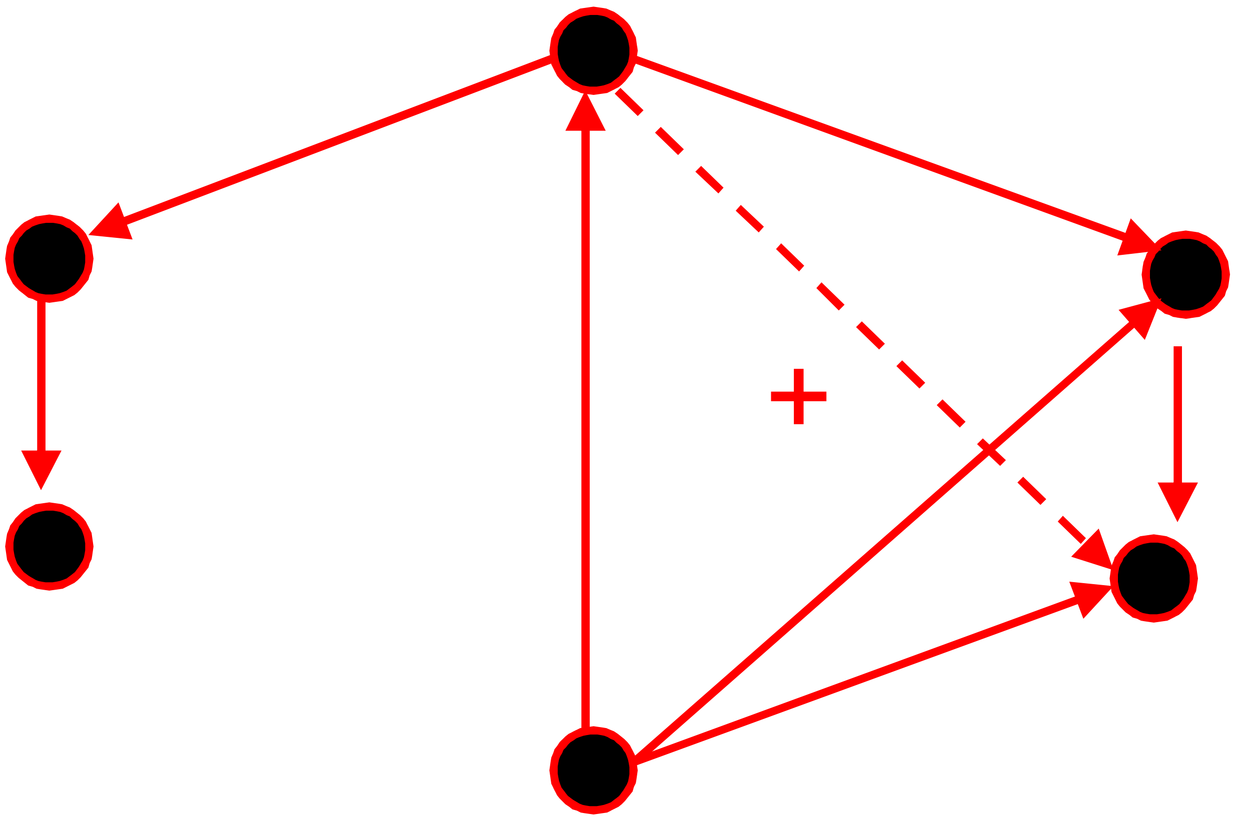
Endogenous Dyad Level: Exchange Theory



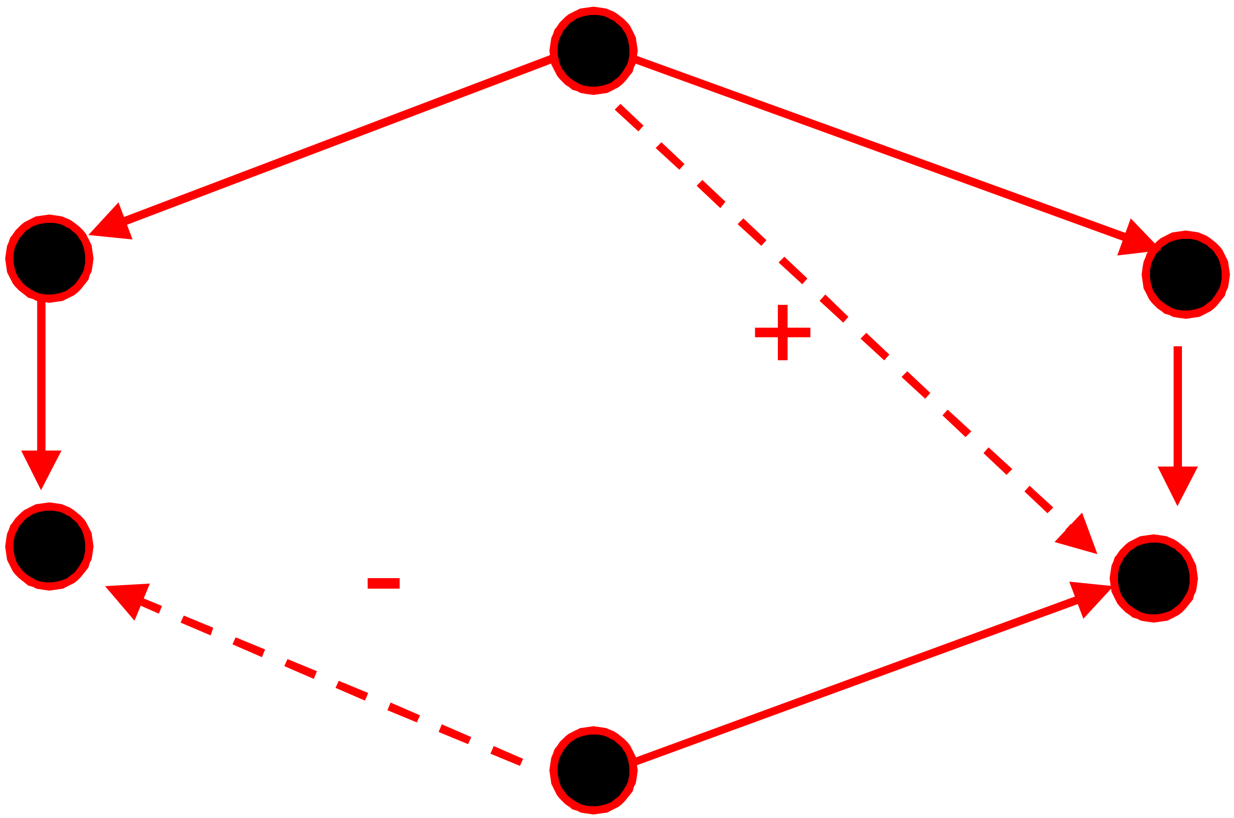
Endogenous Triad Level: Theories of Balance



Endogenous Subgroup Level: Theories of Cohesion



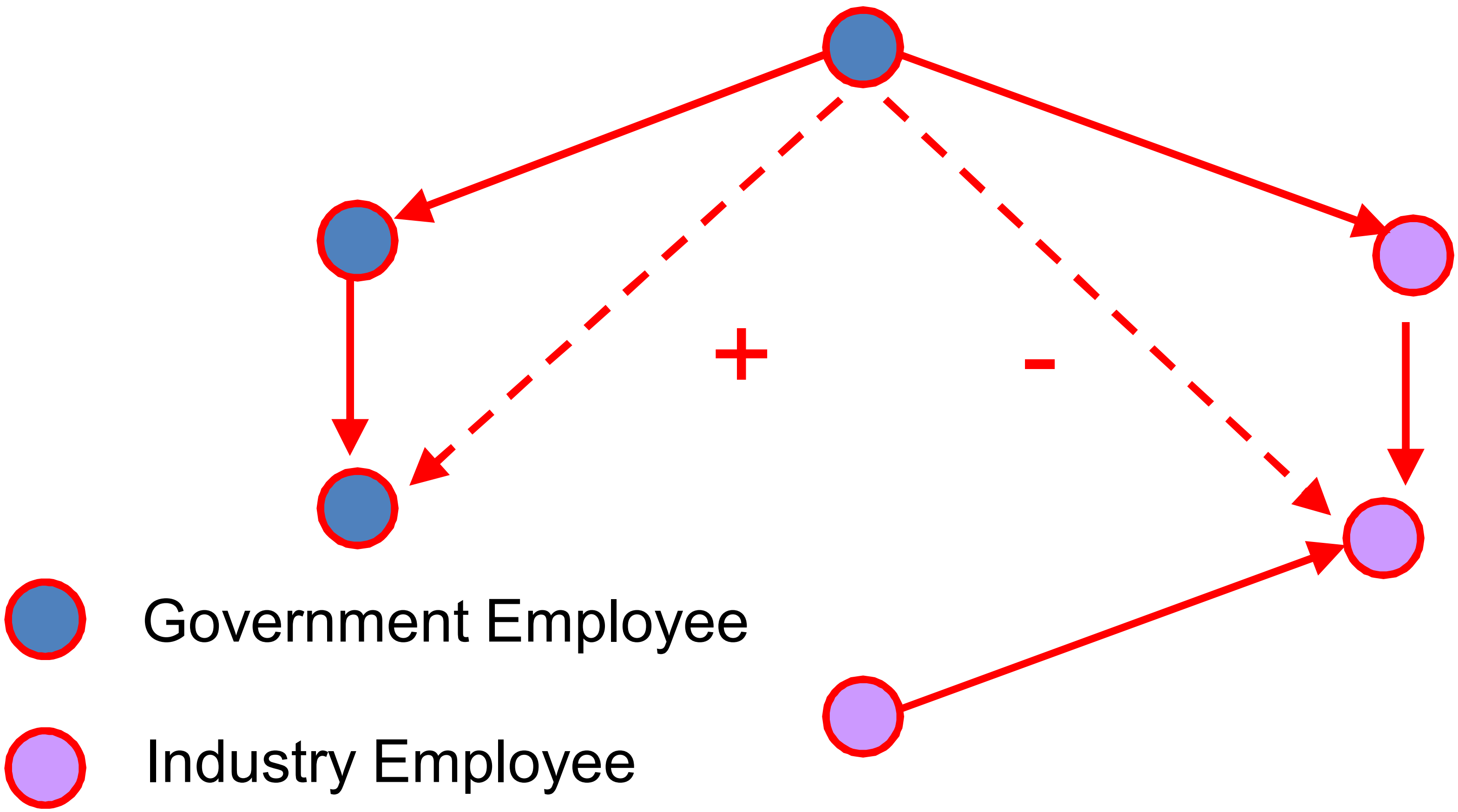
Endogenous Global Level Collective Action Theory



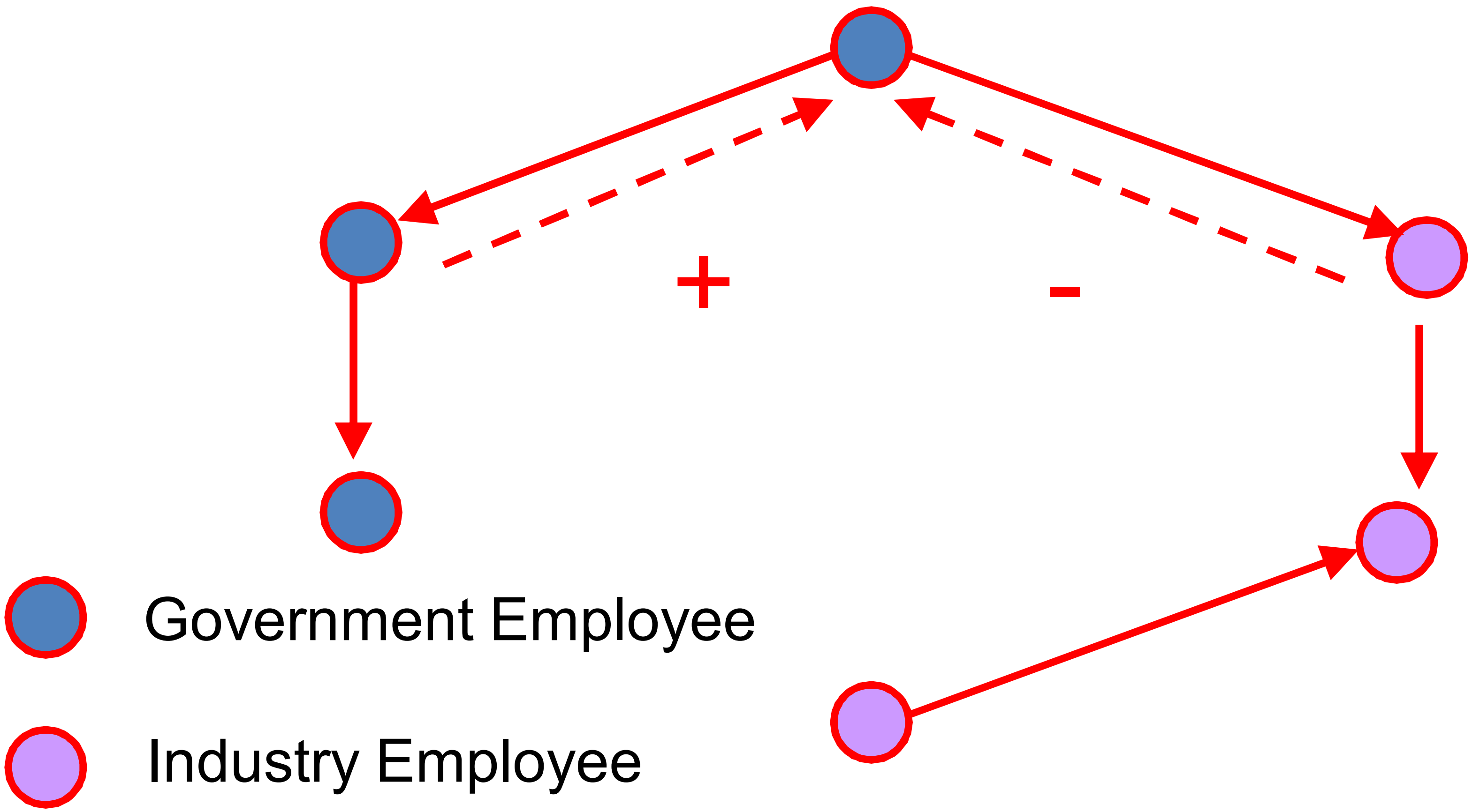
Exogenous attributes

- Actor level
- Dyad level
- Triad level
- Subgroup level
- Global level

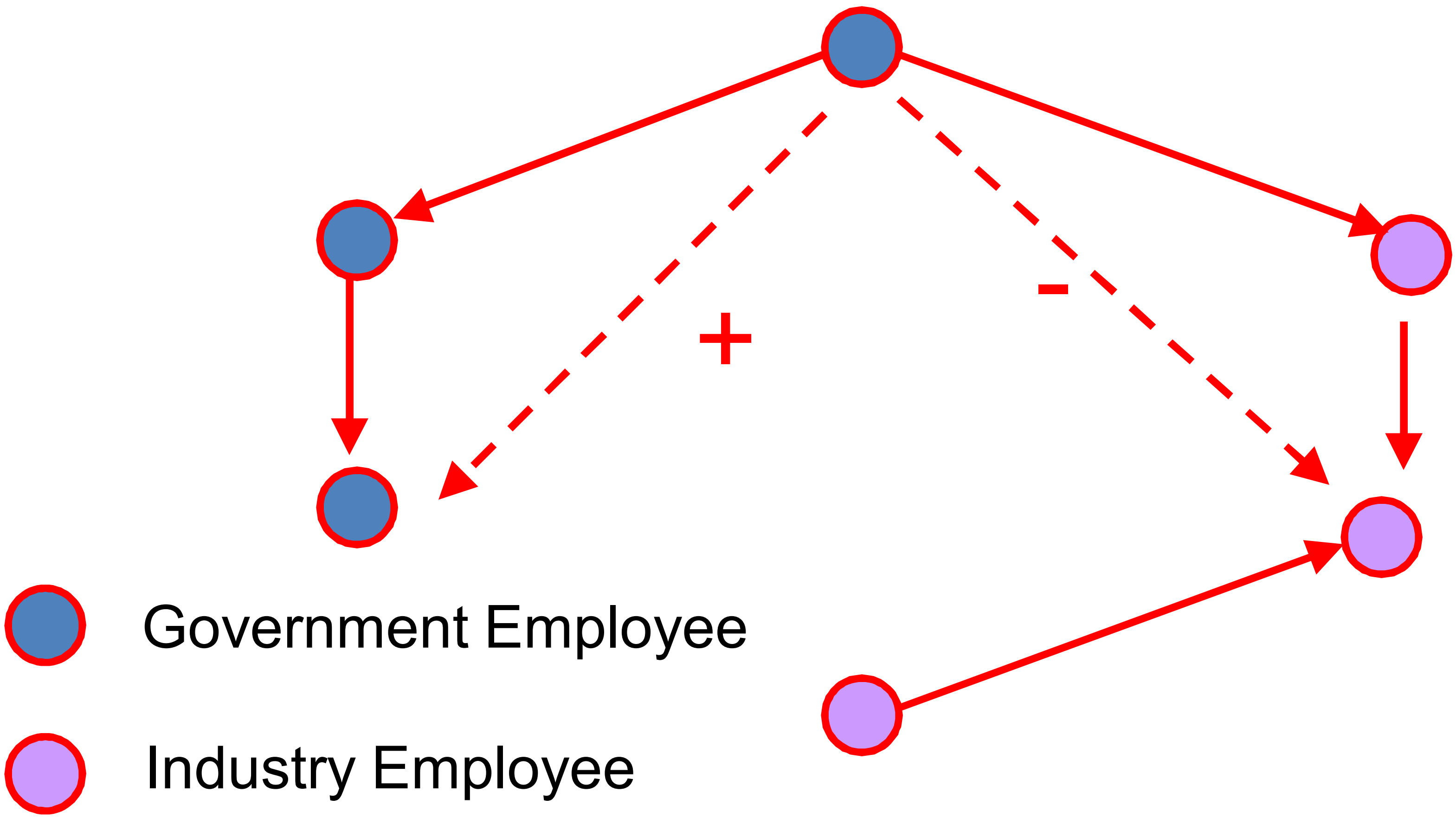
Exogenous Attribute Actor Level: Theories of Homophily



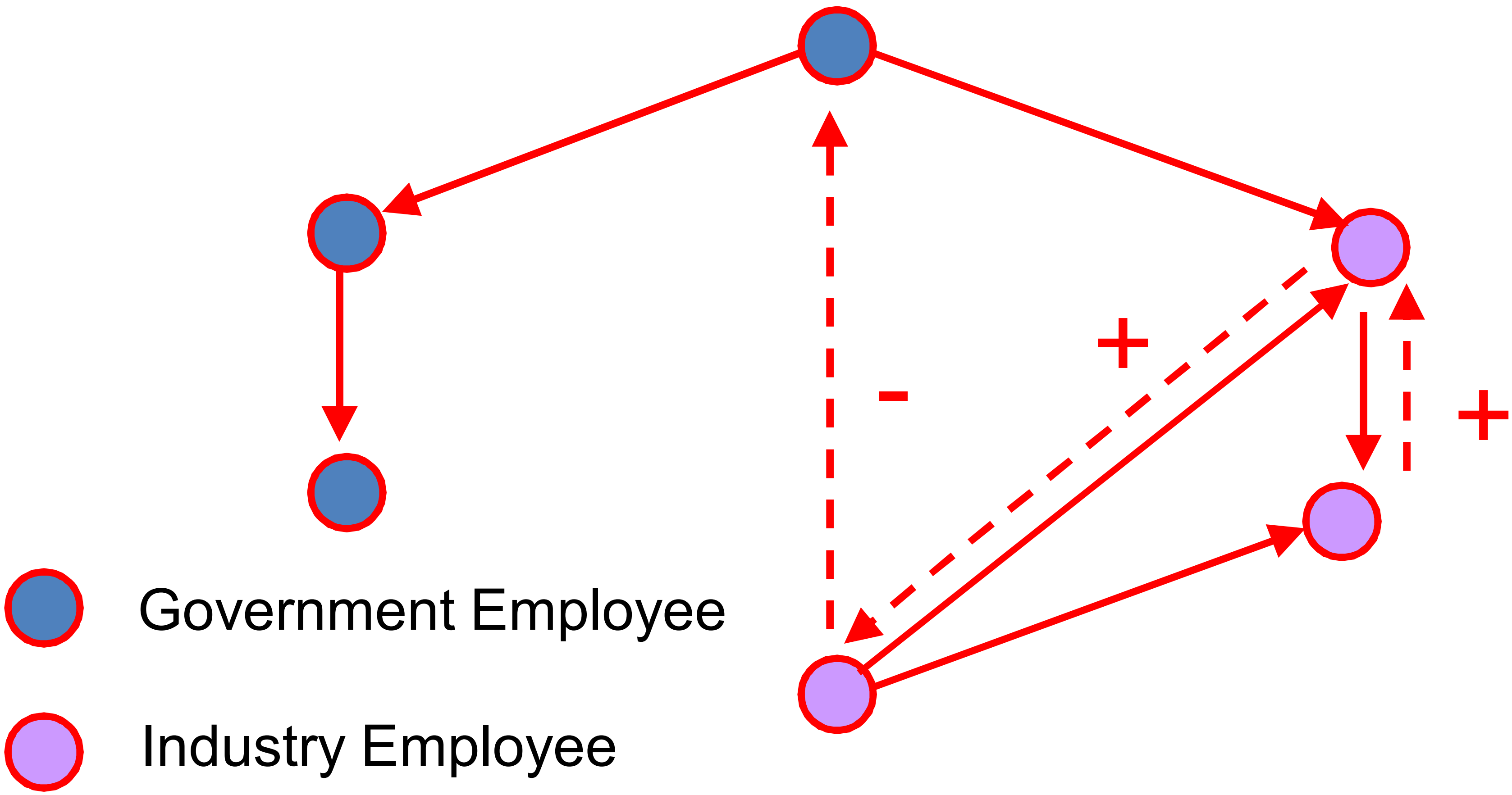
Exogenous Attribute Dyad Level: Resource Dependency Theory



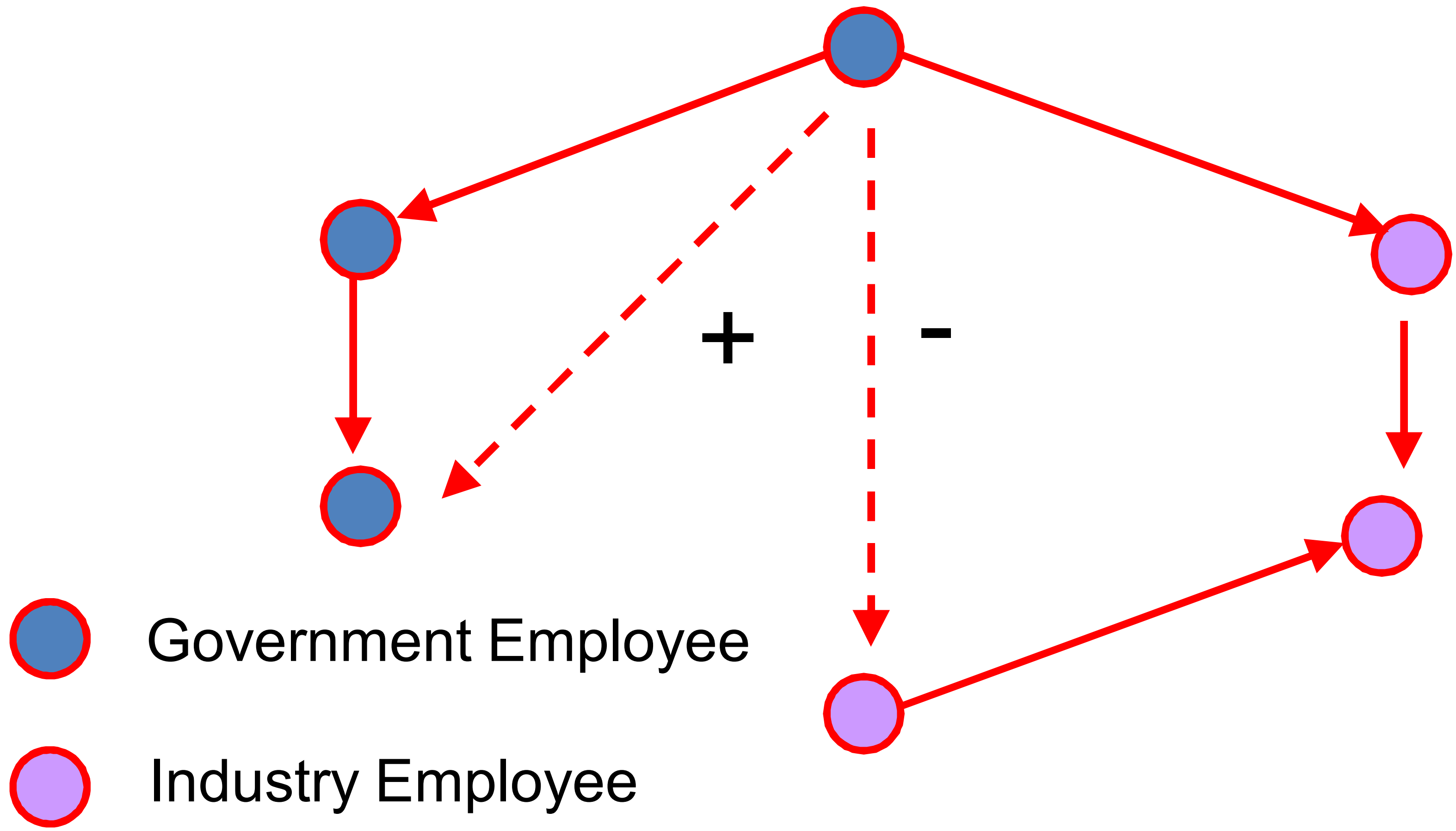
Exogenous Attribute Triad Level: Balance Theory



Exogenous Attribute Subgroup Level: Theories of Cohesion



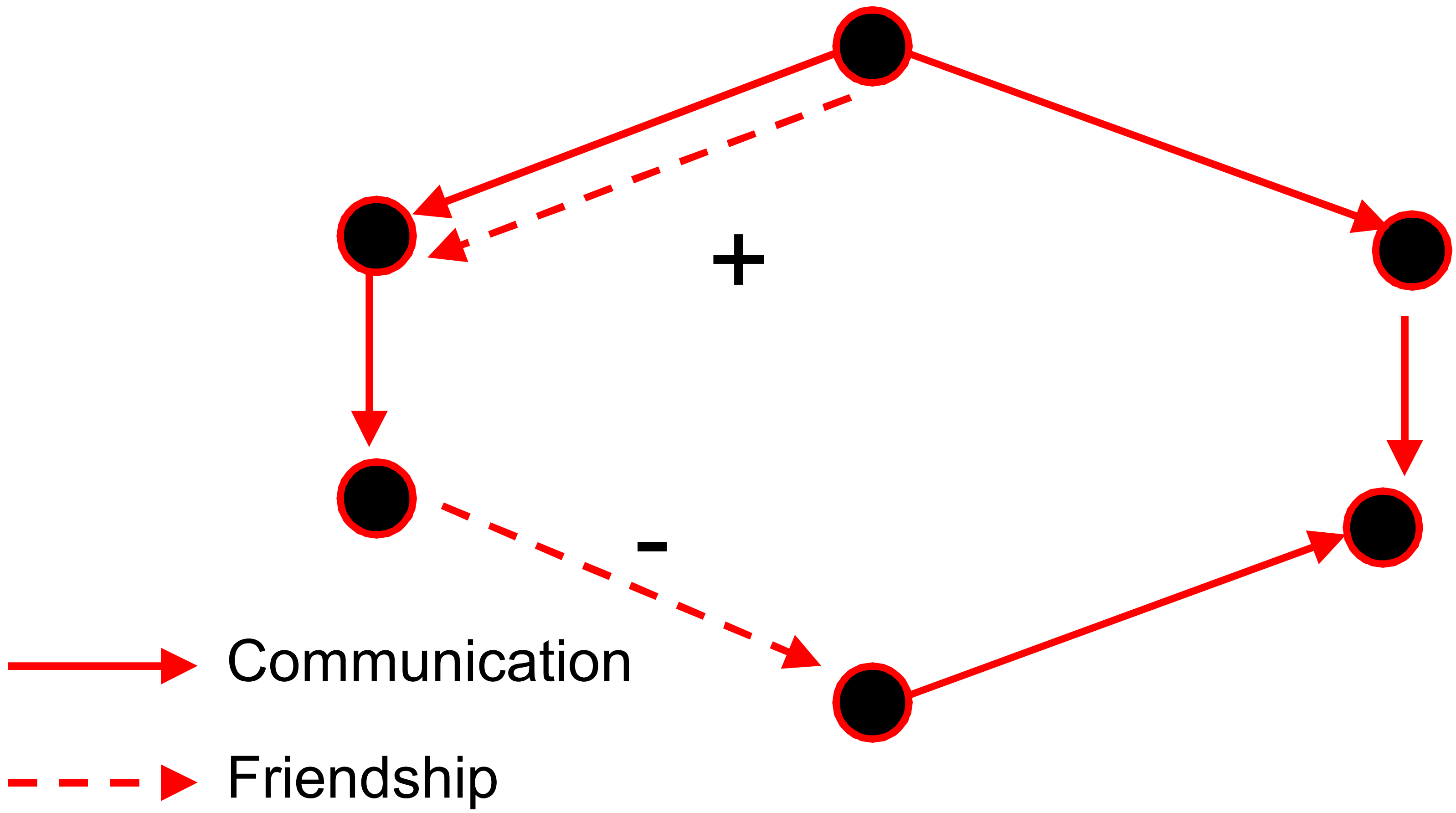
Exogenous Attribute Global Level: Theories of Collective Action



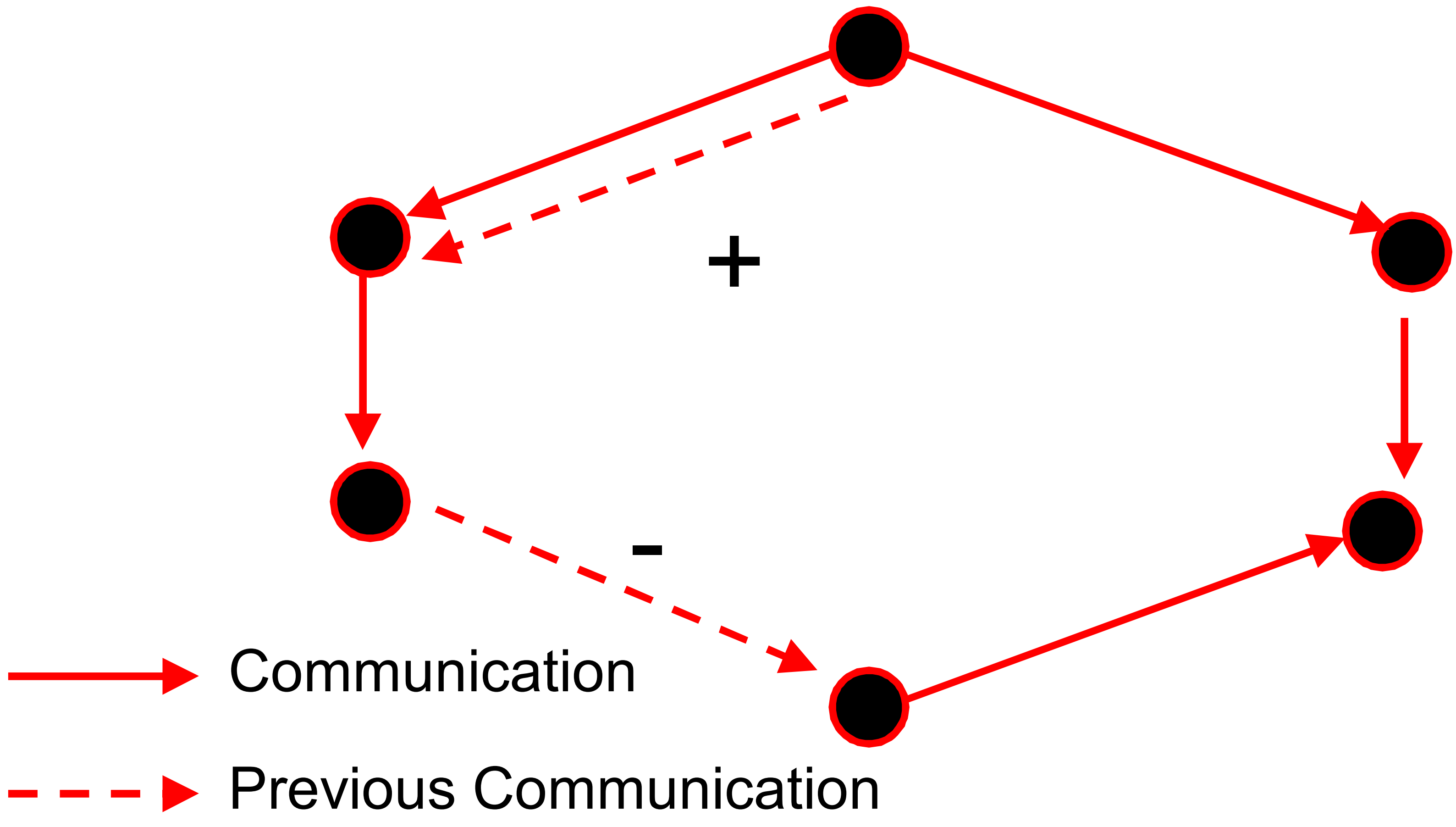
Network is structured by Exogenous Relations

- Advice
- Friendship
- Money

Exogenous Relation: Cognitive Theories



Exogenous Relation: Evolutionary Theories



Why a statistical approach?

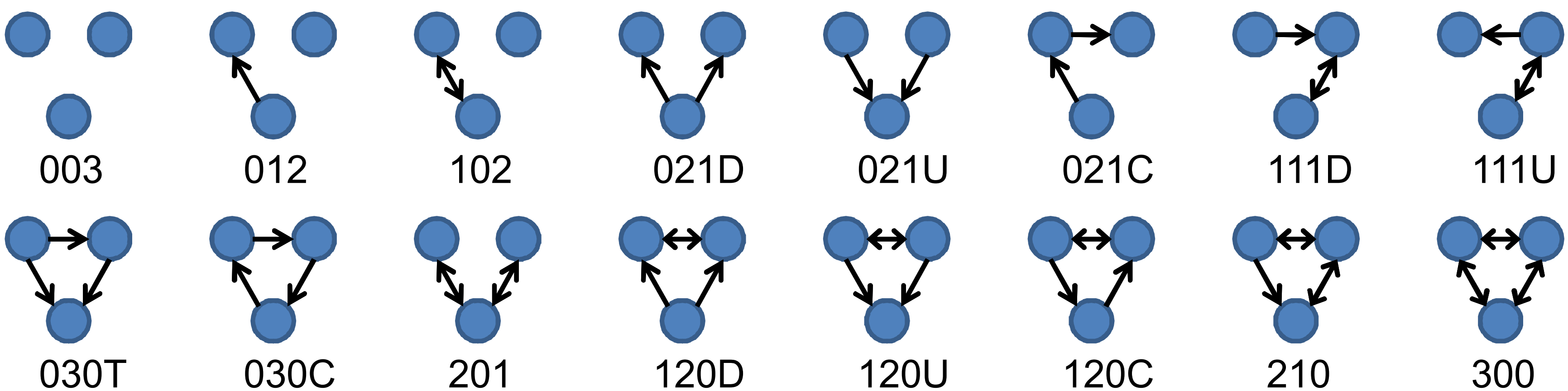
Descriptive vs. generative goals

- ***Descriptive**:* numerical summary measures
  - Nodal level: e.g., centrality, geodesic distribution
  - Configuration level: e.g., triadic census
  - Network level: e.g., centralization, clustering, small world, core/periphery

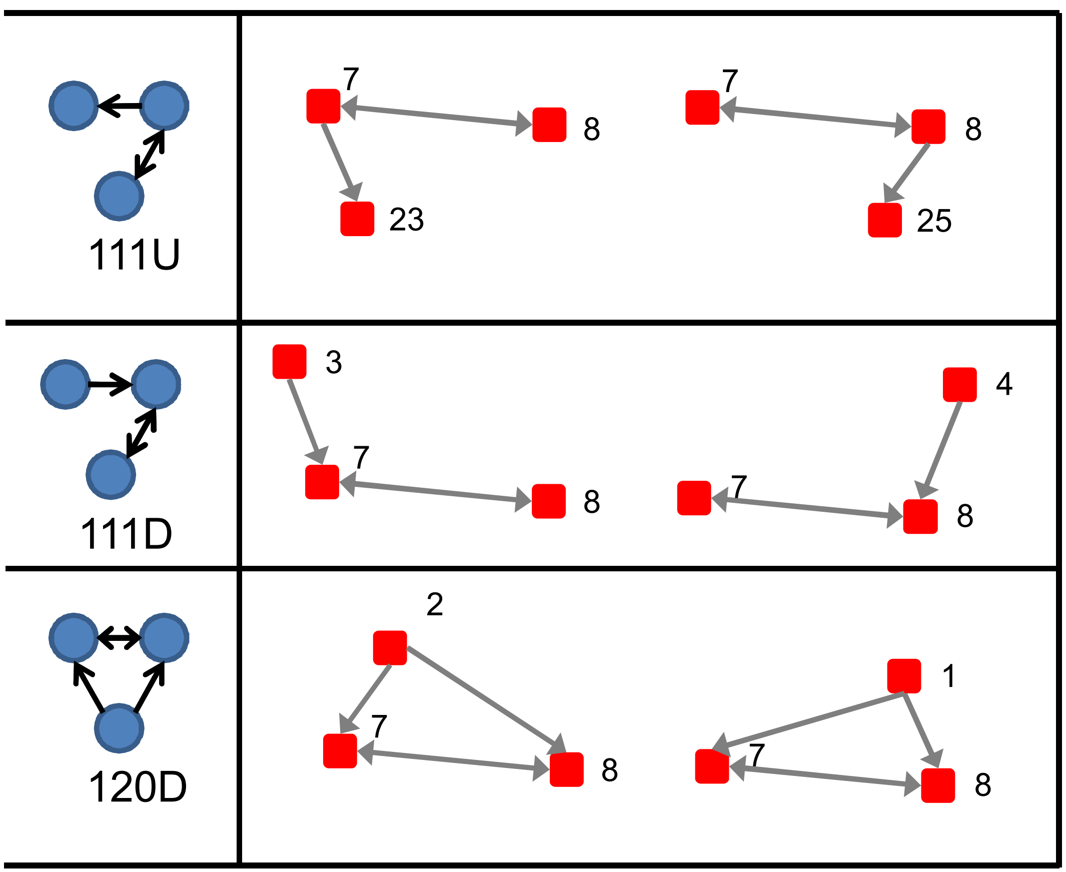
- ***Generative***: micro foundations for macro patterns
  - Recover underlying dynamic process from x-sectional data
  - Test alternative hypotheses
  - Extrapolate and simulate from model

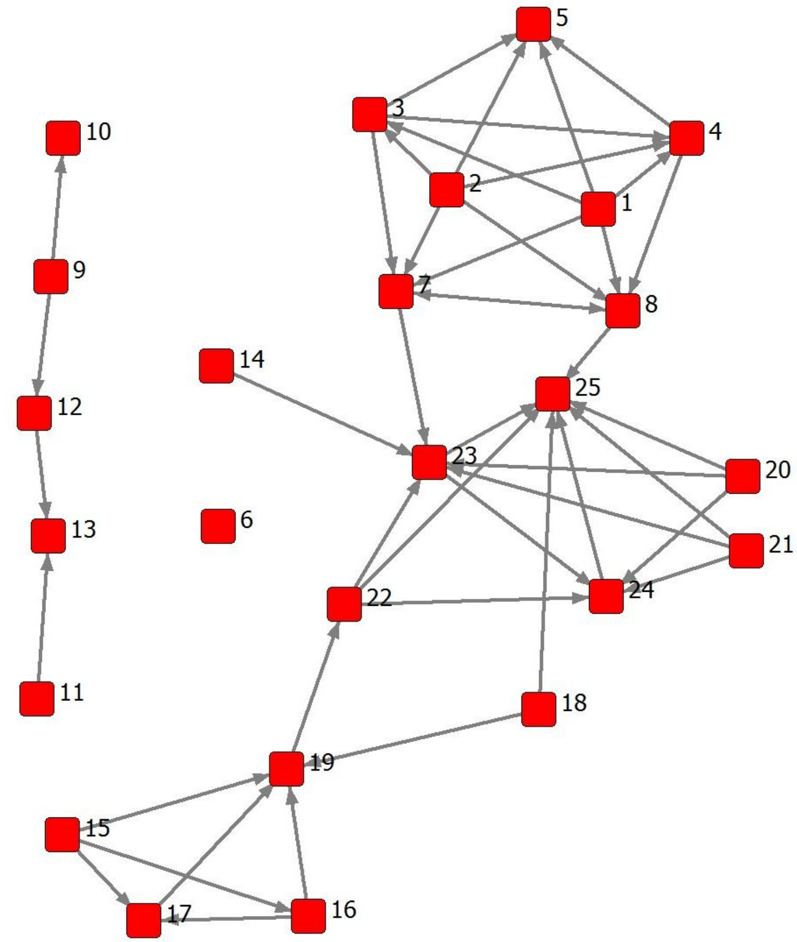
Triad census & network motifs

- 16 possible triads types in a directed network
  - (Reciprocated ties, unreciprocated ties, null ties)
- **Triad census**: frequency of each structure in a network
  - Compare frequencies in observed network, measuring deviations against frequencies in random networks

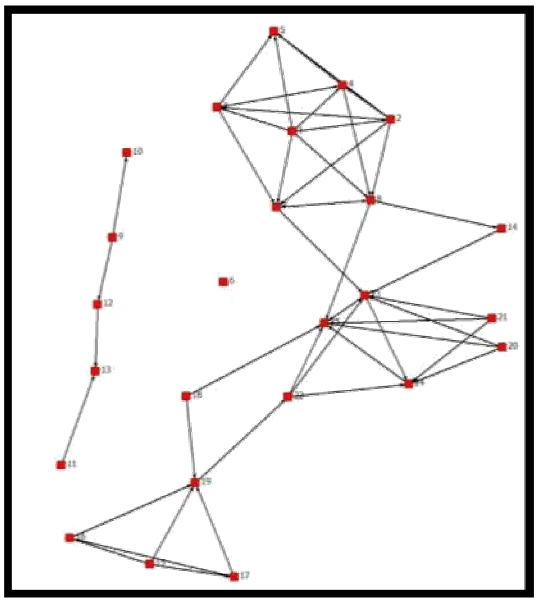


Triad census

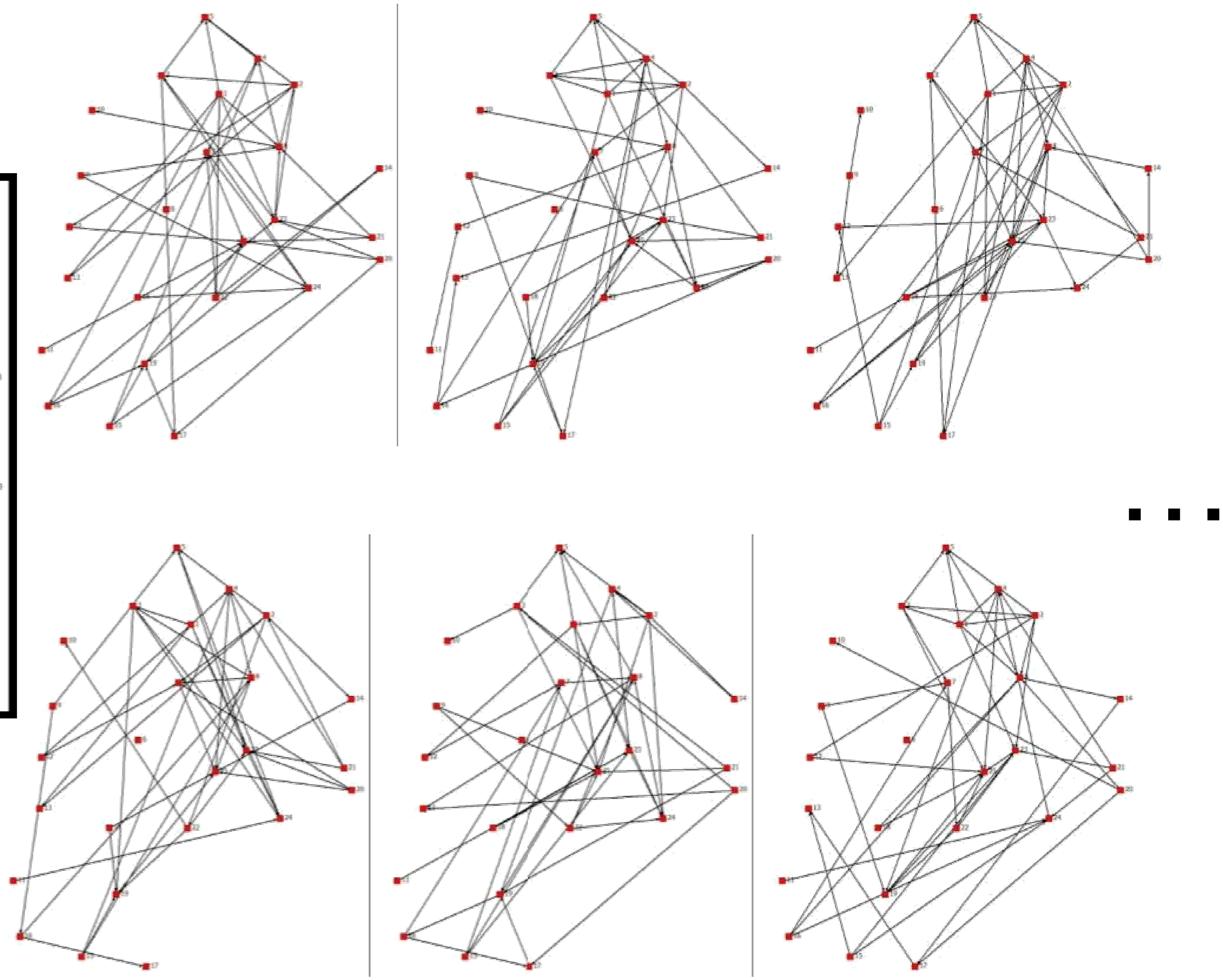


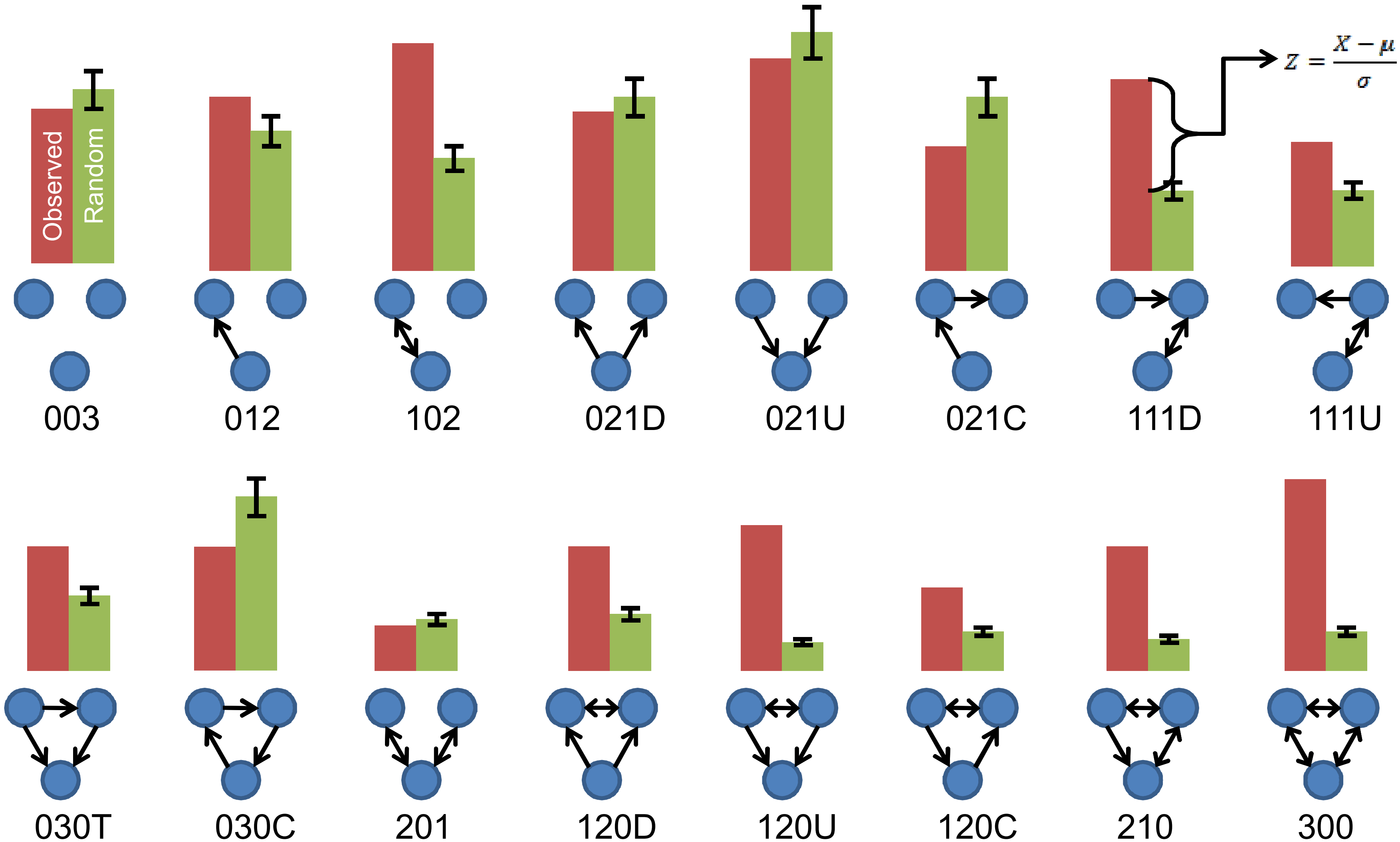


H_0: Random networks



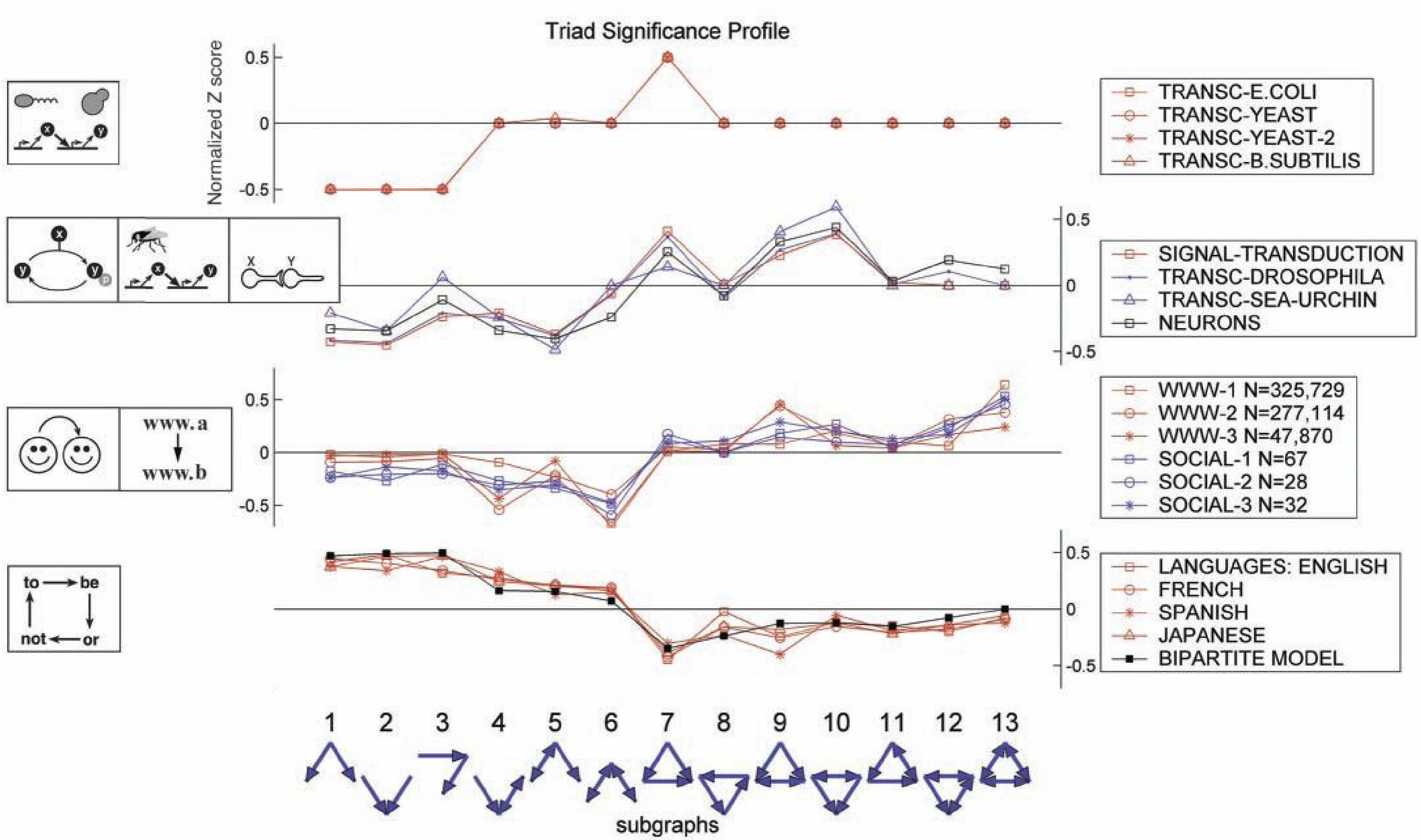
Observed





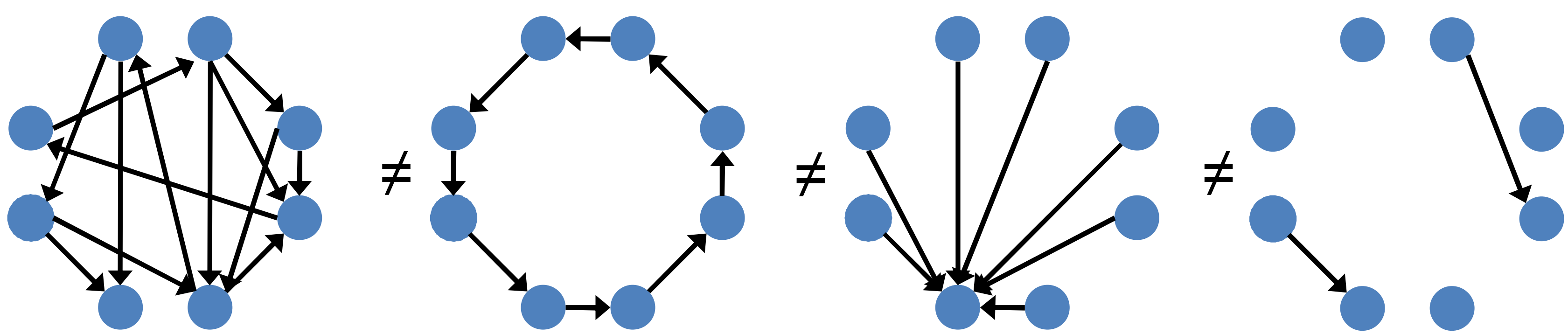
Observed
Random
$Z = \frac{X - \mu}{\sigma}$
003
012
102
021D
021U
021C
111D
111U
030T
030C
201
120D
120U
120C
210
300

Triad census - Network families



Statistical conditioning

- Observed graph of size *n* is only one possible realization of $2^{n(n-1)}$ possible (directed) networks
  - 100 node directed network = 2^{9900} graph permutations >>> ~10^{80} atoms in universe
- Most randomly-generated networks don't look like many observed networks → invalid null models for hypothesis testing



Statistical conditioning

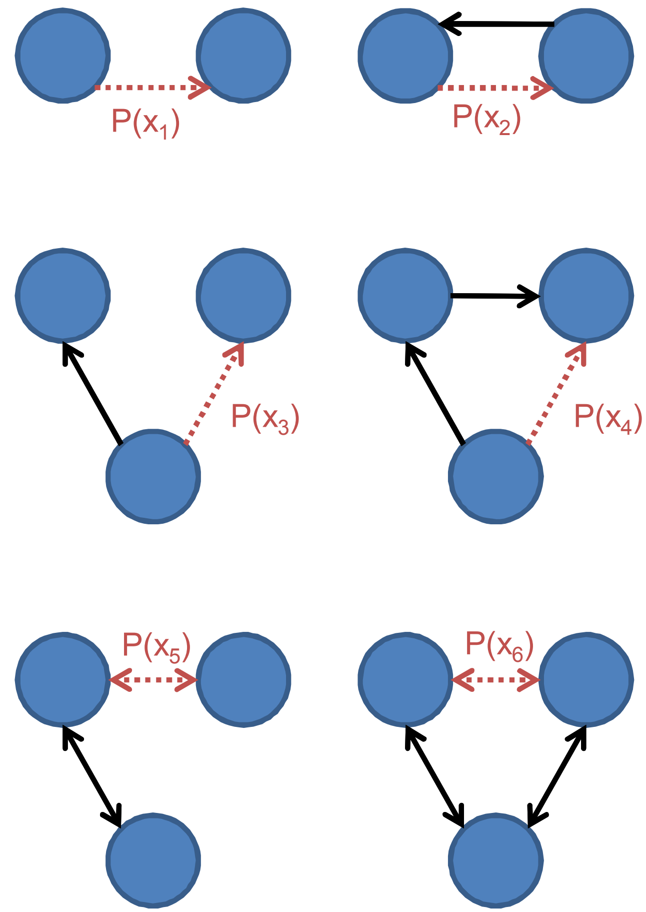
- Deriving a distribution of a random variable based on specific graph properties
- Restrict attention to only those random graphs possessing certain properties
  - Density
  - In/out-degree
  - Diameter
  - Clustering
  - Mutual/asymmetric/null dyads
  - *See*: W&F pp.535-555

Statistical theory testing

- Probabilistic view needed to statistically test propositions about a theory
  - Allow the data to show some error → a distribution of possible outcomes for a given model
  - Evaluate the significance of network structures' presence or absence
  - Control for alternative processes which could give rise to similar network configurations
  - Specify local rules which result in similar global patterns
- How should networked be modeled and what should their distributions look like?

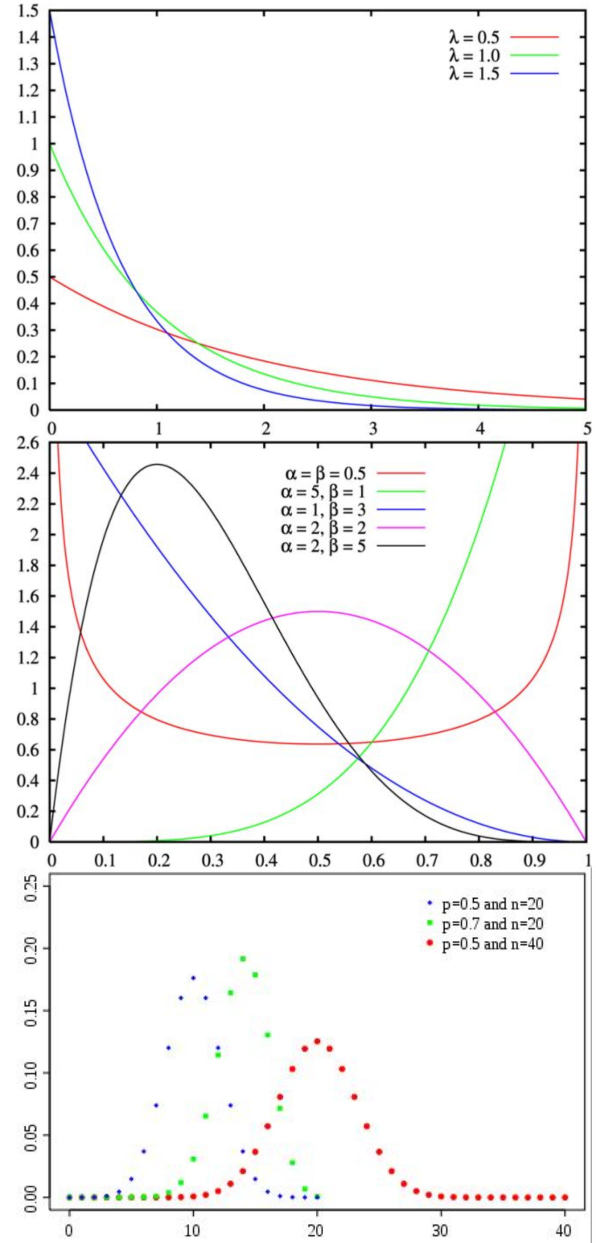
Dyadic independence?

- These models still assume probability of links forming are essentially independent and identically distributed
- Substantial evidence that tie formation depends on the presence/absence of other ties
  - Reciprocity: $P(x_1) << P(x_2)$
  - Transitivity: $P(x_3) << P(x_4)$
  - Balance: $P(x_5) << P(x_6)$
- Statistical models need to account for these dependencies



Probability distribution

- The observed network is one realization out of a set of other possible networks
- Any of the $2^{n(n-1)}$ graphs are possible
- A statistical model assigns a probability to all possible networks



Uniform\Bernoulli Distribution

- Every graph realization is equally likely to occur
- Uniform probability function: $P(X = x) = \frac{1}{2^{n(n-1)}}$
  - Probability that a (labeled) directed graph with n actors equals observed graph
  - All elements of network are independent of all other elements
  - All of the networks occur with equal probability
    - Actors "choose" about ½ of the other actors
    - Expected degree for each actor in network: (n – 1)/2
    - Expected density: 0.50

Bernoulli Process on KRACK?

- KRACKFR
  - n = 21 nodes
  - l_{obs} = 102 edges
  - $\mathbf{P}$ = ½ (uniform)
  - l_{max} = (21)•(21-20) = 420
  - $H_0 : P \bullet l_{max} = 210$
  - $Z = (l_{obs} - l_{max})/\sigma_{dist}$
    = (102-420)/10.25
    = -10.5
  - $p << 0.001$
- KRACKAD
  - n = 21 nodes
  - l_{obs} = 190 edges
  - $\mathbf{P}$ = ½ (uniform)
  - l_{max} = (21)•(21-20) = 420
  - $H_0 : P \bullet l_{max} = 210$
  - $Z = (l_{obs} - l_{max})/\sigma_{dist}$
    = (190-420)/10.25
    = -1.95
  - $p \approx 0.051$
- **Conclusion**: Uniform Bernoulli processes did not generate these networks
- **Major caveats**: Bernoulli only considers density, lower $\mathbf{P}$ would **not** have nullified H_0

$$\hat{P}_{KRACKFR} = 0.243 \qquad \hat{P}_{KRACKAD} = 0.452$$

Why a statistical approach?

Descriptive vs. generative goals

- ***Descriptive:*** numerical summary measures
  - Nodal level: e.g., centrality, geodesic distribution
  - Configuration level: e.g., triadic census
  - Network level: e.g., centralization, clustering, small world, core/periphery

- ***Generative***: micro foundations for macro patterns
  - Recover underlying dynamic process from x-sectional data
  - Test alternative hypotheses
  - Extrapolate and simulate from model

ERGMs are a hybrid:

Traditional generalized linear models (statistical)

- *but*
    - Unit of analysis: relation (dyad)
    - Observations may be dependent (like a complex system)
    - Complex nonlinear and threshold effects
    - Estimation is different

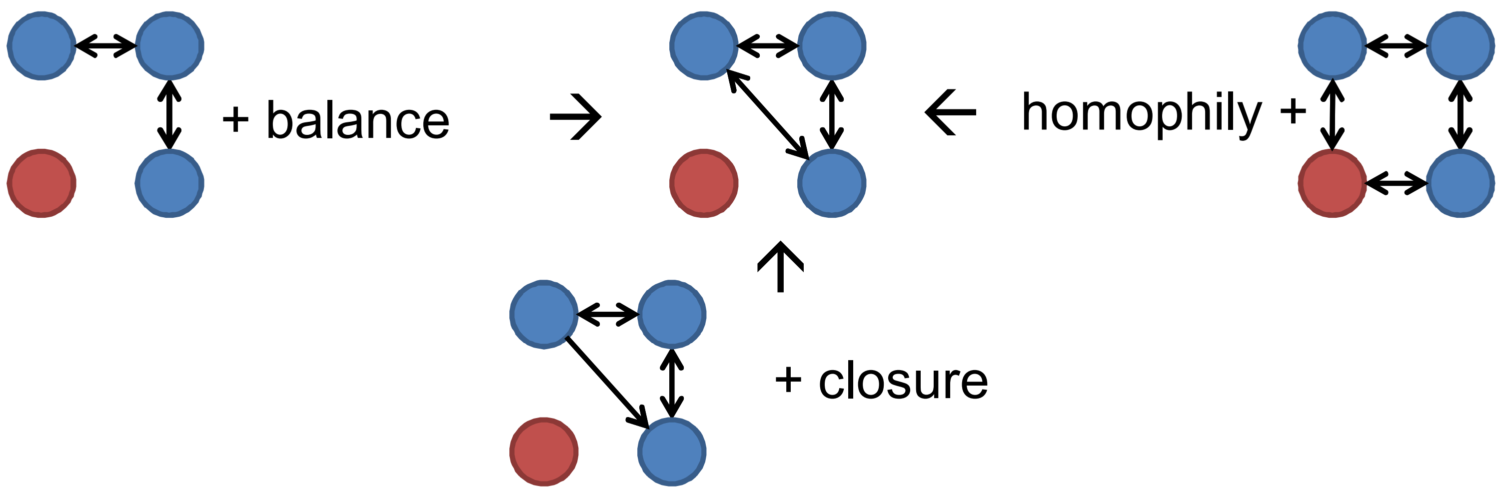
Agent-based models (mathematical)

- *but*
    - Can estimate model parameters from data
    - Can test model goodness of fit

Modified /w permission from – **statnet** Development Team "Introduction to ERGM." Sunbelt 2009.
http://csde.washington.edu/statnet/Resources/Sunbelt2009/ERGM%20intro.ppt

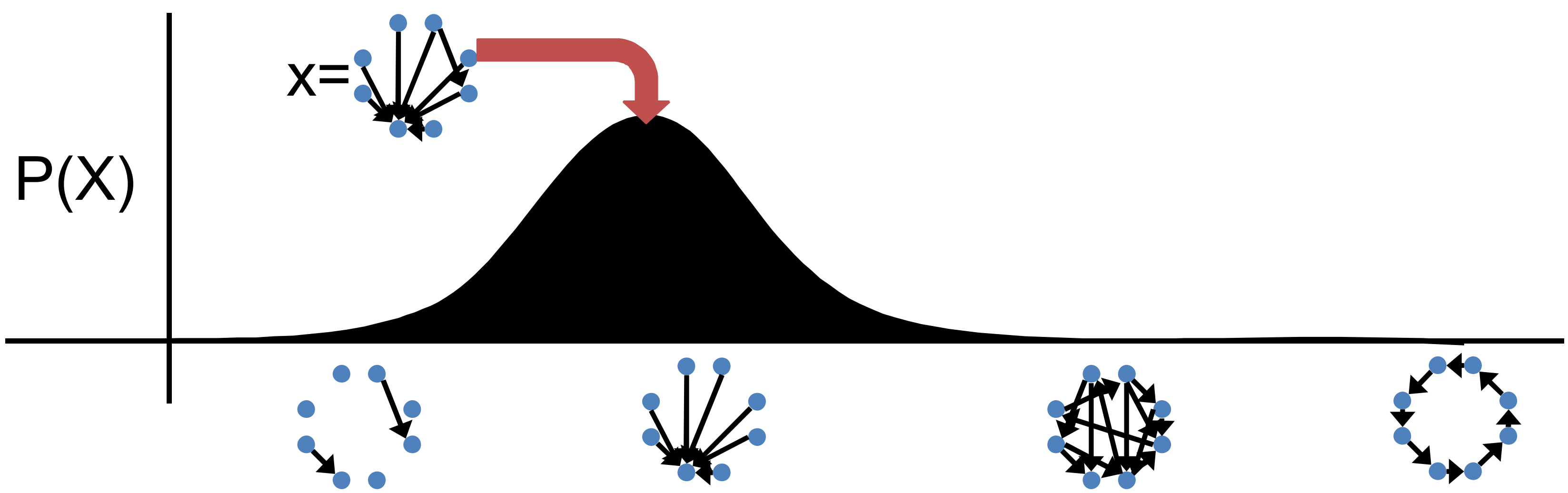
Motivation

- Models of network ties reflect outcome of unobserved social processes that are local, interactive, and stochastic
- Likelihood that a tie exists?
  - Violates traditional statistical independence assumption
  - Presence/absence of tie influenced by presence of other ties and attributes
  - Different micro processes can lead to similar macro structures



Motivation

- Networks exhibiting the structural features of interest should have higher probabilities than networks which do not exhibit structural features of interest
- The observed network is a particular graph in this distribution and has a corresponding probability



Motivation

- Sample graphs at random from the distribution according to probabilities
- Compare the sampled graphs to observed one on any characteristic of interest
- If the model is a good fit, then the sampled graphs will resemble the observed network
- The modeled structural effects *potentially* explain the emergence of the network

Robins, Pattison, Kalish, & Lusher (2006)

Exponential Random Graph Model

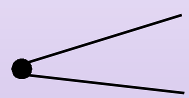
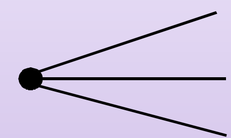
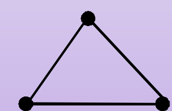
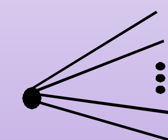
$$P(Y = y) = \frac{1}{k(\theta)} \exp\left(\theta^T g(y)\right)$$

- $\boldsymbol{Y}$ – the network realization, similar to a random variable
- $\boldsymbol{y}$ – the observed network
- $\boldsymbol{g(y)}$ – a vector of network statistics
- $\boldsymbol{\theta}$ – a vector of coefficients corresponding to $\boldsymbol{g(y)}$
- $\boldsymbol{k(\theta)}$ – a normalizing factor calculated by summing up $\exp(\theta^T g(y))$ over all possible network configurations

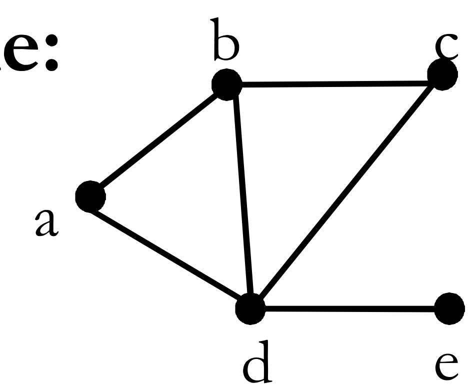
p*/ERG modeling outline

- Specify model parameters that should govern evolution of graph
  - Reciprocity, transitivity, closure, homophily, etc.
- Using observed data, use MLE to estimate coefficients for parameters to generate model
- Simulate other random networks based on this model
  - Generate sample of random networks following model rules
    - Markov Chain Monte Carlo algorithms
  - Resulting networks should "look like" observed network
- Compare the goodness of fit of observed network to model networks

Undirected Structural Parameters

Edge			
2-Star		3-Star	
Triangle		K-Star	

Example:



Edge: 6

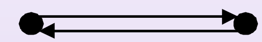
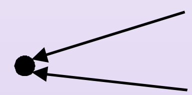
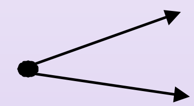
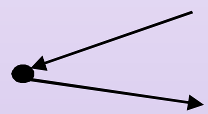
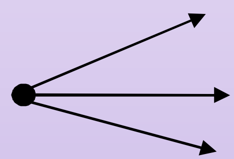
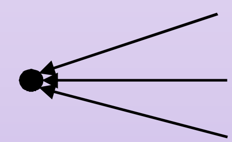
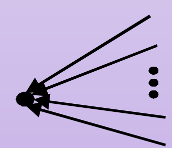
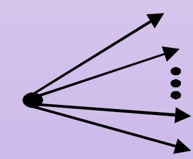
2-Star: 1+3+1+6+0=11

3-Star: 0+1+0+4+0=5

4-Star: 0+0+0+1+0=1

Triangle: 2

Directed Structural Parameters

Arc		Reciprocity	
2-In-Star		2-Out-Star	
Mixed-2-Star			
3-Out-Star		3-In-Star	
K-In-Star		K-Out-Star	

Variables and coefficients

- $\left(\theta^T g(y)\right)$ superficially similar to a multiple regression model's coefficients and variables:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i + \ldots + \beta_p x_{pi} + \varepsilon_i$$

- Observed network is "DV", thetas are "coefficients", parameters are "IVs"
  - ERGM with an arc and reciprocity parameter
  - Arc/edge/density parameter is "intercept"

$$P(Y = y) = \frac{1}{k(\theta)} \exp(\theta_1 * arc + \theta_2 * reciprocity)$$

$$\approx Network \sim \theta_1 * arc + \theta_2 * reciprocity$$

Estimation

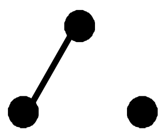
- In linear regression, typically estimate coefficients by ordinary least squares
- In ERGM, we don't know what parameters values to use to assign probabilities
- Using the observed network as a guide, estimate best values using maximum likelihood criterion
  - Maximum likelihood estimation (MLE) also used in logistic regression & multilevel/longitudinal models
  - If $\theta = 0$, then parameter occurs at same rate as chance
  - If $\theta > 0$, then parameter occurs more often than expected by chance
  - If $\theta < 0$, then parameter occurs less often than expected by chance

A Simple Example of MLE

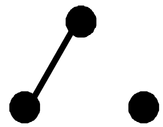
- Model:

$$Network \sim \theta_1 * arc$$

- Observed Network *y*:

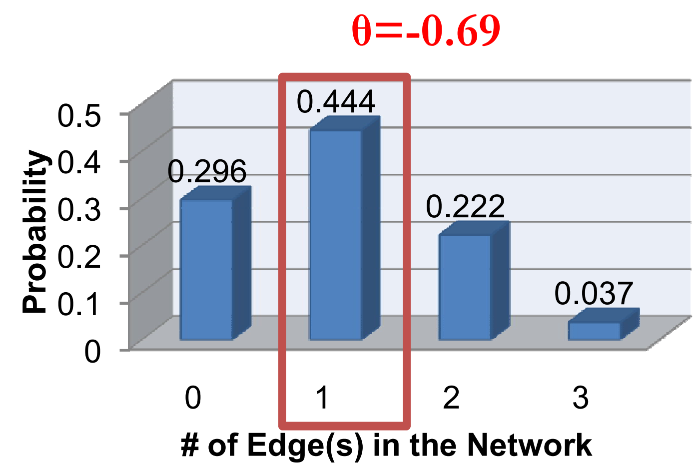


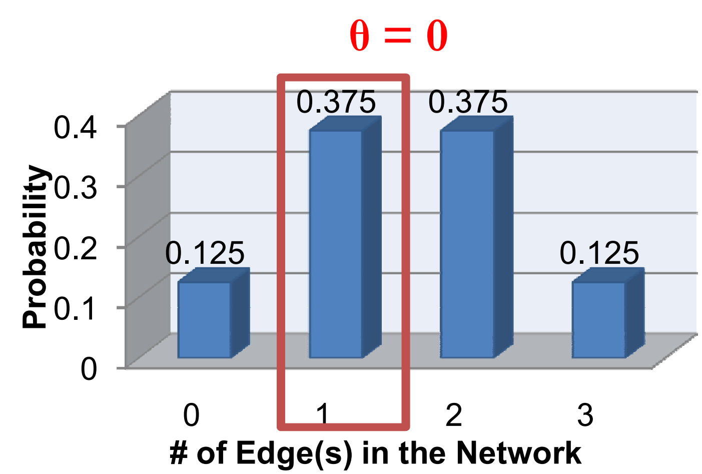
- Goal: Choose θ in such a way that the most probable number of edges is that which occurs in the observed network

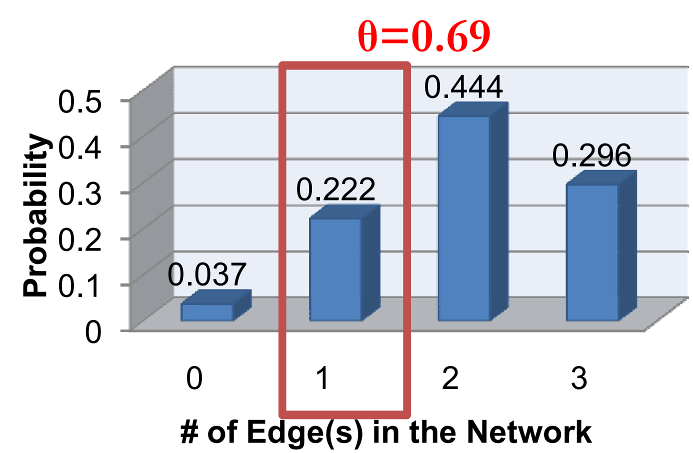
Given the observed Network *y*: 

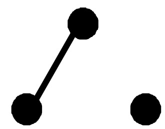
- If θ can be chosen from the following 3 cases, θ=-0.69 is preferred because it gives the highest probability for the observed network

θ	P(Y=*y*)
-0.69	0.444
0	0.375
0.69	0.222



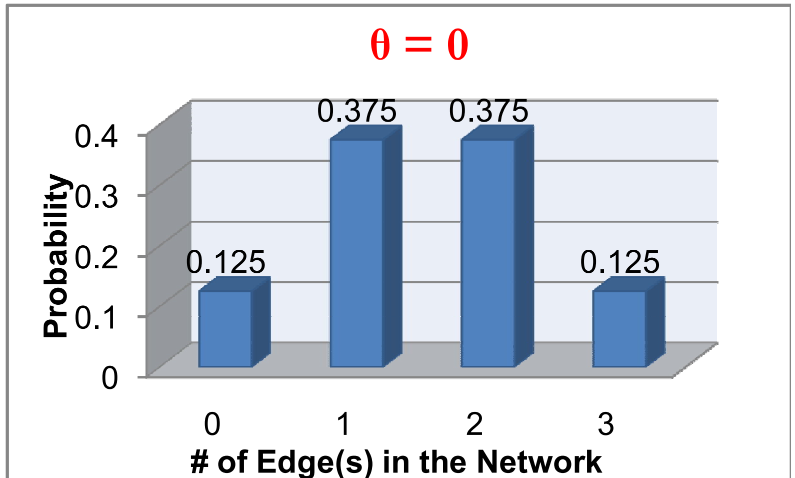


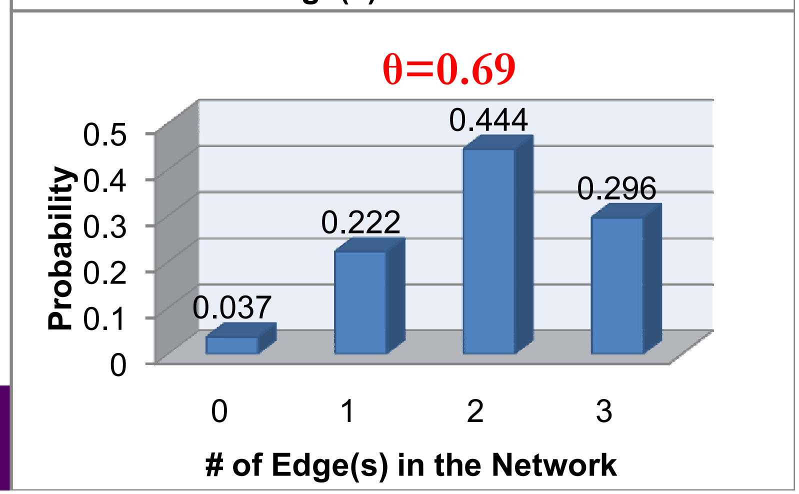


Given the observed Network *y*: 

- If *θ* can only be chosen from the following 2 cases, *θ=0* gives the higher probability and should be adopted.
- However, using the model with *θ=0,* another network configuration has the same probability as *y*, which indicates that this is not a good model.

θ	P(Y=*y*)
0	0.375
0.69	0.222





Goodness of Fit (GOF)

- Model Diagnostic: How well did the model fit the observed data?
- Basic Idea and Procedure
  - Generate a large sample of networks from the fitted model using ERGM simulation
  - Count the number of network statistics of interest for the observed network and each simulated network
  - Do statistical analysis on observed values and simulated values

Statistics in GOF

- Any valid network statistics can be included in the GOF test
- For example:
  - All statistics in the model such as edge, triangle, stars for undirected networks; arc, in-stars, out-stars, transitive triad, etc. for directed networks
  - Any statistics not included in the model but of interest
  - Other global statistics, such as geodesic distance distribution, degree distribution, if supported by the software

PNet GOF Output

- An Example of PNet GOF Output

  observation, sample mean (standard error), t-statistic

  t-statistics = (observation - sample mean)/standard deviation

  # Arc: 304.0000 Mean= 306.4990 (53.4642) t = -0.0467

  # Reciprocity: 85.0000 Mean= 85.0240 (29.6302) t = -0.0008

  # 2-In-Star: 2005.0000 Mean= 2028.2550 (687.4843) t = -0.0338

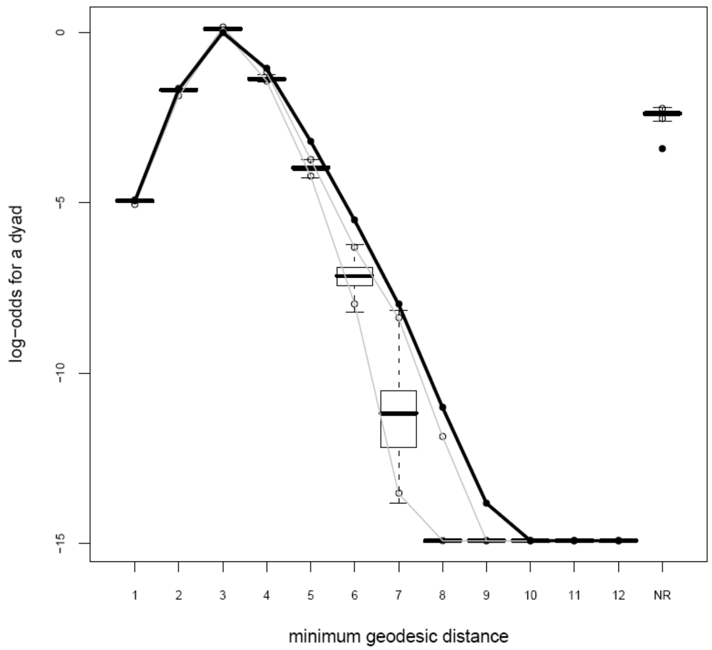
  # 2-Out-Star: 2253.0000 Mean= 2281.2280 (628.9484) t = -0.0449

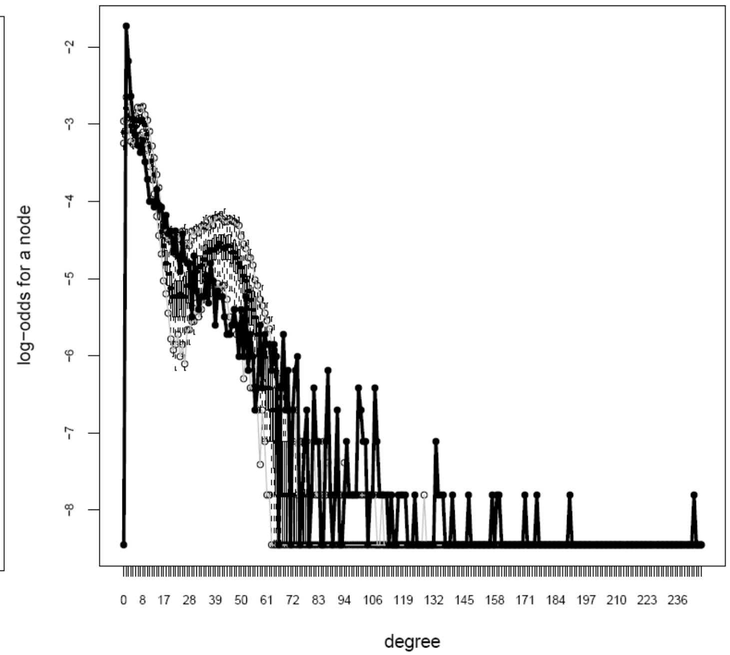
  # 3-In-Star: 9174.0000 Mean= 9261.2210 (4487.3457) t = -0.0194

  # 3-Out-Star: 12347.0000 Mean= 12430.2220 (3964.0944) t = -0.0210

- Criteria for good fit
  - t < 0.1 for network statistics included in the model
  - t < 1 for network statistics not included in the model

Statnet GOF Output Graphics





- Black dots and lines are observed network; Grey dots and lines denote simulated networks
- If the model is a good fit, black line should fall between or close enough to the grey lines

Dynamics - Attribute tendencies

- **Selection**: Immutable characteristics
  - Race, ethnicity, gender, etc.
  - Simple process: Attributes remain fixed and influence how connections are formed
  - Homophily: “birds of a feather flock together”
- **Influence**: alterable characteristics
  - Interests, activities, infectiousness, etc.
  - Complex process: Feedback and interactions between ego’s attributes, network structure, alters’ attributes
  - Diffusion and coevolution

Longitudinal network analysis

- Given longitudinal/panel data for a network, develop a statistical model for the processes driving changes in both network structure and actor attributes
- Pearson, Steglich, Snijders (2005):
  - Changes occur between observations
  - Changes depend on current configuration of structure & attributes – change is gradual and continual, not sudden
  - Actors' behavior based on local information from neighbors rather than considering entire network

Different from ERGM?

- ERGM packages like ‘statnet’ use Monte Carlo Markov Chain Maximum Likelihood Estimation (MCMCMLE) algorithms to model evolution of the network
- Longitudinal network analysis packages like SIENA/RSiena operate under a different set of assumptions: “stochastic actor-oriented models”
  - Like ERGM, specify a model with structural and behavioral parameters with varying levels of influence & significance
  - Like ERGM, estimate model parameters based on observed data
  - Like ERGM, test model goodness-of-fit by simulating networks and comparing characteristics

Co-evolution

- Network & Substance Use **[Pearson, et al. 2005]**

Model	Parameter estimate	SE	p-value	Interpretation
Outdegree	**-1.98**	**0.22**	**<0.001**	**Costly friendship ties**
reciprocity	**2.29**	**0.11**	**<0.001**	**Prefer reciprocation**
distance-2	**-1.08**	**0.07**	**<0.001**	**Prefer network closure**
Rate period 1	**12.72**	**1.49**		**Rate of network change (1)**
Rate period 2	**9.65**	**1.33**		**Rate of network change (2)**
gender homophily	**0.78**	**0.11**	**<0.001**	**Prefer same sex friends**
ego	0.12	0.12	0.32	Girls prefer more friends
alter	-0.17	0.13	0.19	Girls less attractive as friend

Co-evolution

Model		Parameter estimate	SE	p-value	Interpretation
Smoking homophily		0.42	0.34	0.22	Prefer same smoke friends
	Ego	**0.28**	**0.13**	**0.03**	**Smokers name more friends**
	Alter	**-0.25**	**0.13**	**0.05**	**Smokers less attractive**
Cannabis homophily		0.18	0.51	0.72	Prefer same hash friends
	Ego	-0.15	0.09	0.1	Hashers name less friends
	Alter	0.09	0.1	0.37	Hashers more attractive
Alcohol homophily		**0.96**	**0.38**	**0.01**	**Prefer same drink friends**
	Ego	-0.04	0.04	0.32	Drinkers name less friends
	alter	0.06	0.05	0.23	Drinkers more attractive

Co-evolution - Smoking

Model	Parameter estimate	SE	p-value	Interpretation
tendency	**-3.36**	**1.19**	**0.004**	**Low smoking tendency**
assimilation	0.39	0.37	0.29	Some smoking influence
gender	0.91	0.47	0.05	Girls smoke more
cannabis	**1.09**	**0.38**	**<0.01**	**Hashers smoke more**
alcohol	0.33	0.25	0.19	Drinkers smoke more
Rate 1	2.01	0.87		Rate of smoking change
Rate 2	1.31	0.35		Rate of smoking change

Co-evolution - Cannabis

Model	Parameter estimate	SE	p-value	Interpretation
tendency	-1.02	0.93	0.27	Low hashing tendency
assimilation	**3.54**	**1.43**	**0.01**	**High hash influence**
gender	-0.99	0.64	0.12	Girls hash less
smoking	0.6	0.47	0.2	Smokers hash more
alcohol	0.42	0.41	0.31	Drinkers hash more
Rate 1	0.61	0.17		Rate of hash
Rate 2	1.53	0.4		Rate of hash

Co-evolution - Alcohol

Model	Parameter estimate	SE	p-value	Interpretation
tendency	0.25	0.3	0.41	High drinking tendency
assimilation	**1.63**	**0.43**	**<0.001**	**High drinking influence**
gender	0.23	0.22	0.3	Girls drink more
smoking	-0.5	0.44	0.26	Smokers drink less
alcohol	0.51	0.37	0.17	Hashers drink more
Rate 1	1.67	0.31		Rate of drinking
Rate 2	2.33	0.47		Rate of drinking

Backup slides

NORTHWESTERN UNIVERSITY

Statistical analysis

- Estimate parameters of the process
  - Joint estimation of multiple, possibly correlated, effects
- Inference from sampled data to population
  - Uncertainty in parameter estimates
- Goodness of fit
  - Traditional diagnostics
    - Model fit (BIC, AIC)
    - Estimation diagnostics (MCMC performance)
  - Network-specific GOF
    - Network statistics in the model as covariates
    - Network properties not in the model

What is a statistical model?

- A *statistical* model is a…
  - …formal representation of a…
  - …stochastic process…
  - …specified at one level (e.g., person, dyad) that…
  - …aggregates to a higher level (e.g., population, network)

Dependence Graph I

- Any observed single relational network may be regarded as a realization **x** = $[x_{ij}]$ of a random two-way binary array **X** = $[X_{ij}]$.
- In general, the entries of the array **X** cannot be assumed to be independent;
- Consequently, it is helpful to specify a dependence structure for the random variables $\{X_{ij}\}$ as originally suggested by Frank and Strauss (1986).

Dependence Graph II

- The first step for any probabilistic model of a network is to consider the statistical dependencies among the elements of this set.
- To do this, we construct a *dependence graph*. Such a device allows us to distinguish among the many possible graph probability distributions, which can often be characterized by considering which relational ties are assumed to be statistically independent

Dependence Graph III

- The *dependence graph D* of the random array **X** is itself a graph whose:

- *Nodes* are elements of the index set {(*i, j*); *i, j* ∈ *N, i* ≠ *j* } for the random variables in **X**, and

- *Edges* signify pairs of the random variables that are assumed to be conditionally dependent (given the values of all other random variables).

Dependence Graphs for Various Classes of Random Graph Distributions

- As Frank and Strauss (1986) observed for univariate graphs and associated two–way binary arrays, several well–known classes of distributions for random graphs may be specified in terms of the structure of the dependence graph.
- In social networks, Wasserman and Pattison (2000, 2002) note that there are three major classes –
  - *Bernoulli and conditional uniform graph distributions,*
  - *Dyadic independence distributions, and*
  - *Markov graph and other p* random graph distributions*

Dependence Graphs for Dyadic Dependence Random Graph Distributions

- The assumption of conditional dependence of X_{ij} and X_{kl} if and only if $\{k, l\} = \{j, i\}$ leads to the class of *dyad dependence* models (see Wasserman, 1987; Wasserman and Pattison, 2002), the second family of graph distributions mentioned earlier.

- These "multinomial dyad" distributions assume all dyads are statistically independent, and postulate substantively interesting parameterizations for the probabilities of the various dyad states.

- The dependence graph for such distributions has an edge set with edges connecting only the two random variables within each dyad: $E_D = \{((i, j), (j, i)), \text{ for all } i \neq j\}$. This class of models was termed *p*1 by Holland and Leinhardt (1977, 1981).

Dependence Graphs for *p** Random Graph Distributions

- Consider now a general dependence graph, with an arbitrary edge set. Such a dependence graph yields a very general probability distribution for a (di)graph, which we term *p**.

- One very general dependence graph, for which this distribution was first developed, assumes conditional independence of X_{ij} and X_{kl} if and only if $\{i, j\} \cap \{k, l\} = \emptyset$.

- This dependence graph links any two relational ties involving the same actor(s); thus, any two relational ties are associated if they involve one or more of the same actor(s).

- This type of dependency resembles a Markov spatial process, so these dependencies were defined as a Markov graph by Frank and Strauss (1986).

Dependence Graphs for Generalized *p** Random Graph Distributions

- This *p** family of distributions has been extended in many ways, and estimates of its parameters scrutinized.
- If the dependence graph is fully connected, then a general class of random graphs is obtained.
- However, any model deriving from a fully connected dependence graph is not identifiable.
- Recent efforts introduce more complex dependence structures that permit models more general than Markov random graphs, but avoid fully connected dependence graphs (see Carrington, Scott, & Wasserman, in press).

Modeling *p** Random Graph Distributions

- For an observed network, which we consider to be a realization **x** of a random array **X**, we assume the existence of a dependence graph *D* for the random array **X**.
- The edges of *D* are crucial here; consider the set of edges, and determine if there are any *complete subgraphs*, or cliques found in the dependence graph.
- For a general dependence graph, a subset *A* of the set of relational ties N_D is *complete* if every pair of nodes in *A* (that is, every pair of relational ties) is linked by an edge of *D*. A subset comprising a single node is also regarded as complete.
- These cliques specify which subsets of relational ties are all pair wise, conditionally dependent on each other.

Using Dependence Graphs to Model *p** Random Graph Distributions

- The Hammersley-Clifford theorem (Besag, 1974) provides the important link between the dependence graph and the structure of the model that encapsulates its dependence assumptions.
- The theorem establishes that the probability model for the random multigraph, **X**, depends on the complete subgraphs of the dependence graph, *D*.
- A *complete subgraph*, or *clique*, is a subset of nodes in the dependence graph every pair of which is linked by an edge. A subset consisting of a single node is also regarded as complete.
- Each complete subgraph corresponds to a configuration of possible ties in the network.
- There is a model parameter corresponding to each complete subgraph in the dependence structure (and so to each corresponding configuration of possible ties).
- The parameter for a particular configuration reflects the effect of observing that configuration on the likelihood of the network.

Using Dependence Graphs to Model *p** Random Graph Distributions

- The random graph model is of the following exponential form:

$$\Pr(X = x) = p^*(x) = \frac{1}{k}\exp\left(\sum_{A \subset N_D} \lambda_A z_A(x)\right)$$

where:

x is a realization of the random graph, **X**;

$\kappa = \sum_x \exp\left(\sum_{A \subset N_D} \lambda_A z_A(x)\right)$ is a normalizing quantity; the summation is over all subsets *A* of nodes of *D*;

z_A(**x**) is the empirically observed network statistic in **x** corresponding to the subgraph *A* of *D* and is given by $z_A(\mathbf{x}) = \prod_{X_{ij} \in A} x_{ij}$;

λ_A is the parameter corresponding to the subgraph *A* of D; and $\lambda_A = 0$ whenever the subgraph induced by the nodes in *A* is not a complete subgraph of *D*.

Interpreting Parameters in the Model *p** Random Graph Distributions

- The random graph model is of the following exponential form:

$$P(Y = y) = \frac{1}{k(\theta)} \exp(\theta^T g(y))$$

- The quantities $z_A(\boldsymbol{x})$ are calculated from the observed network and correspond to the hypothesized structural tendencies expressed in the dependence graph.
- λ_A are parameters corresponding to the cliques A of **D**. These parameters express the importance of the associated structural tendency for the probability of the graph.

Number of Parameters to Model *p** Random Graph Distributions

$$P(Y = y) = \frac{1}{k(\theta)} \exp(\theta^T g(y))$$

- The set of non–zero parameters in this probability distribution for *Pr*(**X** = **x**) depends on the *maximal* cliques of the dependence graph.
- A maximal clique is a complete subgraph that is not contained in any other complete subgraph).
- Any subgraph of a complete subgraph is also complete (but not maximal), so that if *A* is a maximal clique of *D*, then the probability distribution for the (di)graph will contain non–zero parameters for *A* and all of its subgraphs.
- Clearly, the number of parameters can be overwhelming.

Reducing Parameters to Model *p** Random Graph Distributions

- $Pr(\boldsymbol{X} = \boldsymbol{x}) = p^*(\boldsymbol{x}) = \kappa^{-1} \exp\{\sum_{A \subseteq N_D} \lambda_A z_A(\mathbf{x})\}$

- Thus, it is wise to limit these numbers by either postulating a simple dependence graph, or by making assumptions about the parameters.

- One usual assumption is *homogeneity*, in which parameters for isomorphic *configurations* of nodes are equated.

- Two subgraphs, *A* and *A'* are isomorphic if there is a one-to-one mapping, φ, from the nodes of *A* to the nodes of *A'* that preserves the adjacency of nodes. Formally, *A* and *A'* are isomorphic if there exists a mapping φ on the node set *N* such that $(i, j) \in A$ if and only if $(\varphi(i), \varphi(j)) \in A'$ for $i, j \in N$.

- Other assumptions about homogeneity of parameters can be based on actor attributes, social positions or geography

Estimating Parameters to Model *p** Random Graph Distributions

- The probability distribution arising from the Hammersley-Clifford Theorem, equation (1), can be written in the general form:

$$Pr(\mathbf{X} = \mathbf{x}) = \exp\{\lambda \mathbf{z}(\mathbf{x})\}/\kappa(\lambda)$$

where:

λ is a vector of model parameters
$\mathbf{z}(\mathbf{x})$ is a vector of network statistics
$\kappa(\lambda)$ is a normalizing constant, which guarantees that the distribution is proper.

Likelihood function to Model *p** Random Graph Distributions

- The likelihood function for the distribution is quite simple:

$$L(\boldsymbol{\theta}) = \exp\{\lambda\, \mathbf{z}(\mathbf{x})\}/\kappa(\lambda)$$

Even though it has a simple expression, the function is not easy to work with, due to the dependence of $\kappa(\lambda)$ on the unknown parameters.

Likelihood function to Model *p** Random Graph Distributions

- Two approaches have been used to date:
  - Maximum pseudo-likelihood estimation technique, pioneered by Besag (1975, 1977) and refined and applied by Strauss (1988) and Strauss and Ikeda (1991); and
  - Markov chain Monte Carlo maximum likelihood estimation, being applied to *p** by Crouch and Wasserman (1998), Snijders (2001), and others.

MCMC Estimation to Model *p** Random Graph Distributions

- For the particular case of *p** models, the central idea is to simulate a distribution of random graphs from a starting set of parameter values, and then to refine these estimated parameter values by comparing the distribution of graphs with the observed graph.
- The process is repeated until the parameter estimates stabilize.
- Simulation procedures establish a Markov chain of graphs that, under suitable conditions, will converge to a stationary graph distribution.
- Two of the most popular algorithms that can produce such a Markov chain are
  - the Gibbs sampler (Geman and Geman, 1983) and
  - the Metropolis–Hastings algorithm (which includes the Gibbs sampler as a special case; see Chib and Greenberg, 1995).

General framework

1. Regard each possible tie of each possible type as a random variable. All possible X_{ij} are random variables.
2. Formulate hypotheses about the interdependencies of the ties of different types. Which pairs of random variables are conditionally dependent given the other random variables? Are we interested in hypotheses about role-sets? Role-reciprocity? Role interlocking?
3. Construct a dependence graph, where the nodes are the random variables (X_{ij}) and the edges link pairs of random variables hypothesized to be conditionally dependent in Step 2.
4. Invoke the Hammersley-Clifford theorem to obtain a joint probability model for the set of random variables. Each model parameter corresponds to a complete subgraph in the dependence graph.
5. Consider homogeneity constraints. Should some parameters be expected to be equal (e.g. based on isomorphism classes, actor attributes, social positions or geography)?
6. Estimate model parameters (e.g. using Maximum Pseudo-likelihood Estimation or MCMC)