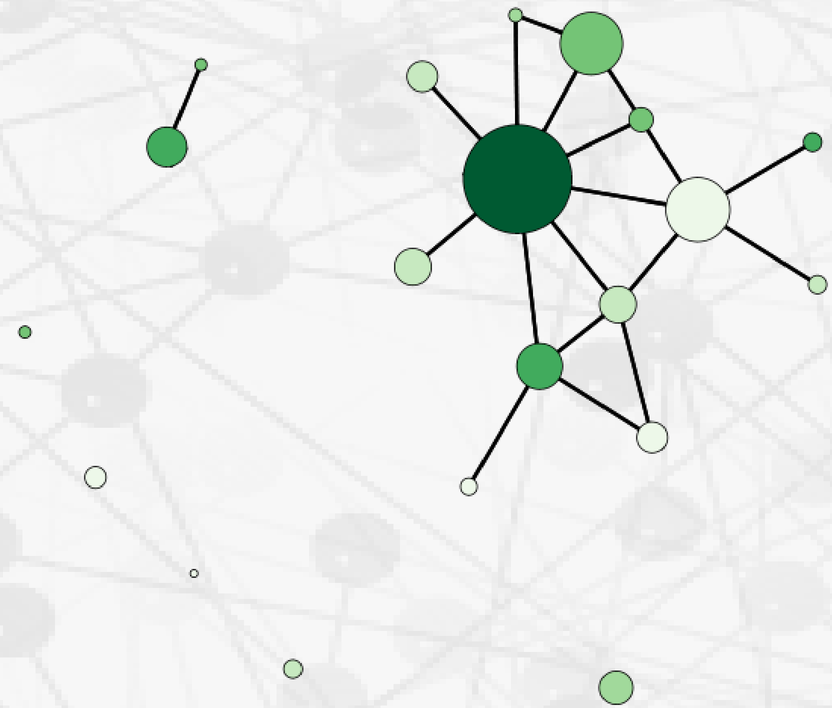
Supplemental Slides for the *Network Analysis with* `statnet` Workshop

2010 Political Networks Conference

Carter T. Butts
Department of Sociology and
Institute for Mathematical Behavioral Sciences
University of California, Irvine

Relating NLIs to Vertex Covariates

- **Common question: are NLIs related to non-structural covariates?**
  - Centrality to power or influence
  - Constraint to advancement
  - Diversity to development



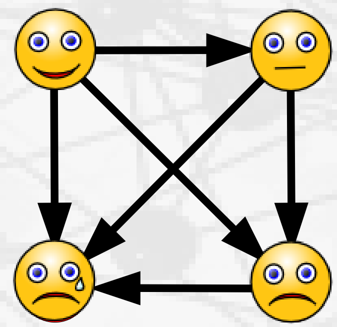
(Texas SAR EMON Decision Rank Score (scale) vs. Degree (color), Mutually Reported "Continuous" Communication" - ρ=0.86)

"Linear" Permutation Tests

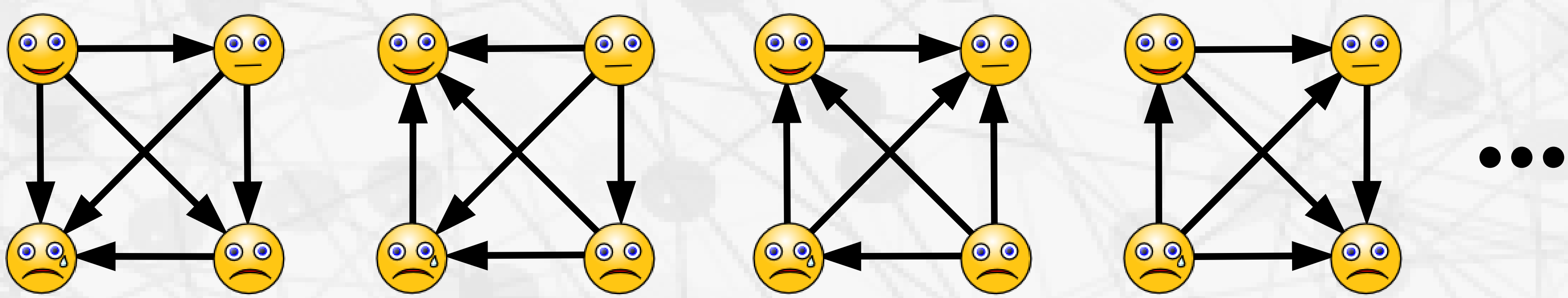
- **Simple, nonparametric test of association between vectors**
  - Sometimes called "linear" or "vector" permutation test
  - Tests marginal association against exchangeability null (independence conditional on marginal distributions)
- **Null interpretation: "musical chairs" model**
  - If we randomly switched the *positions* of people in the network (leaving *structure* as-is), what is the chance of observing a similar degree of association?
- **Monte Carlo procedure:**
  - Let $\mathbf{x}_{obs}=(f(v_1,G),\ldots,f(v_N,G))$ be the observed NLI vector, w/covariate vector $\mathbf{y}$
  - Let $t_{obs}=s(\mathbf{x}_{obs},\mathbf{y})$
  - For i in $1,\ldots,n$
    - Let $\mathbf{x}^{(i)}$ be a random permutation of $\mathbf{x}_{obs}$
    - Let $t^{(i)}=s(\mathbf{x}^{(i)},\mathbf{y})$
- **Estimated *p*-values:**
  - One-sided:
    - $\Pr(\mathbf{t}^{(i)}\leq t_{obs})\approx\sum_i I(\mathbf{t}^{(i)}\leq t_{obs})/n$ (Lower)
    - $\Pr(\mathbf{t}^{(i)}\geq t_{obs})\approx\sum_i I(\mathbf{t}^{(i)}\geq t_{obs})/n$ (Upper)
  - Two-sided:
    - $\Pr(|\mathbf{t}^{(i)}|\geq|t_{obs}|)\approx\sum_i I(|\mathbf{t}^{(i)}|\geq|t_{obs}|)/n$

Understanding the Null Model

We Observed:



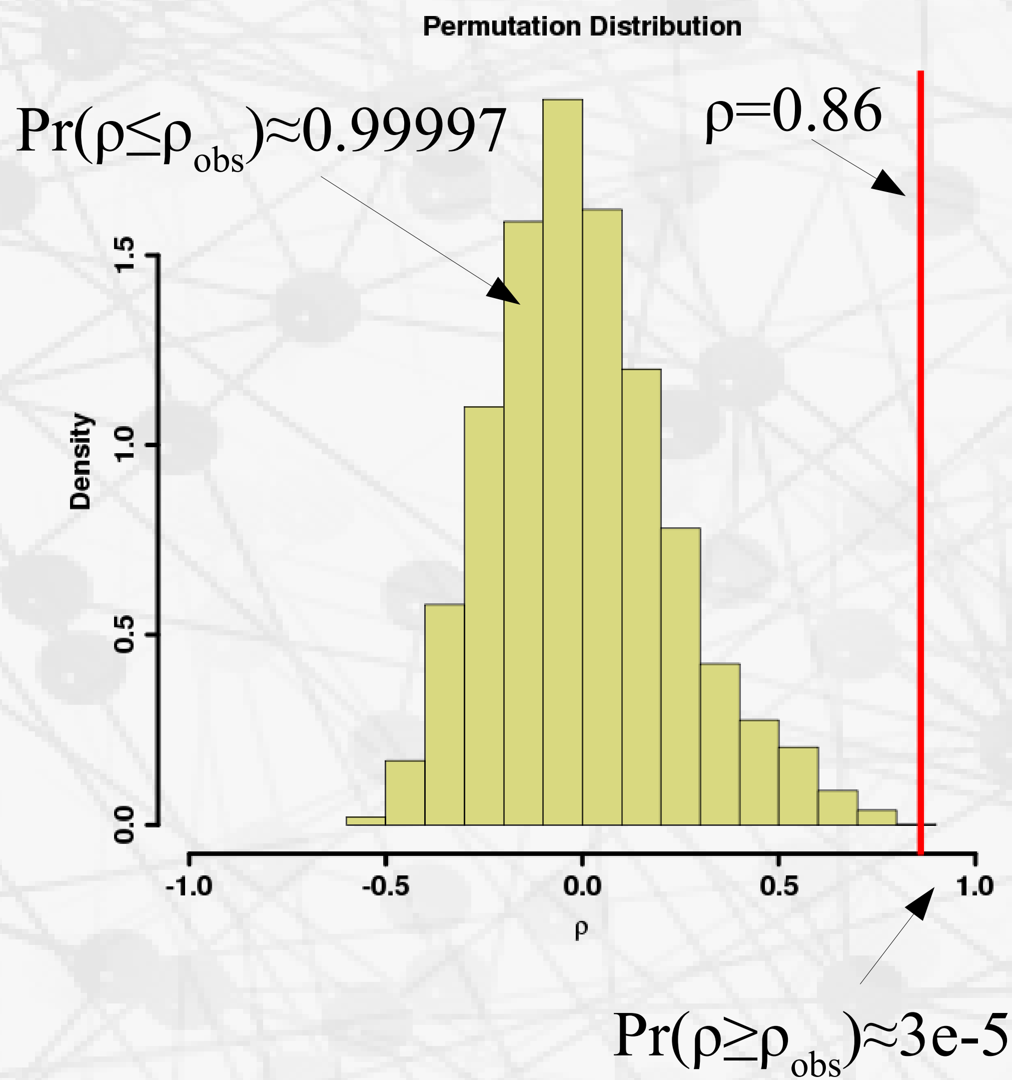
We *Could Have* Observed:



We Ask: *"Is the observed relationship extreme compared to what we would expect to see, if assignment to positions were independent of the covariate?"*

Texas SAR EMON Example

- **Question: do organizations in constant communication w/many alters end up more/less prominent in the decision-making process?**
  - Measure (s): correlation of decision rank score ($\mathbf{y}$) with degree in confirmed "continuous communication" network ($\mathbf{x}_{obs}$)
  - Null: no relationship between degree and decision making
  - Alternative: decision making has linear marginal relationship w/degree
- **Results:**
  - t_{obs}=0.86; $\Pr(|\mathbf{t}^{(i)}|\geq|t_{obs}|)\approx$3e-5
  - Correlation this large very unlikely under null hypothesis
  - Upper tail test similar (see figure)



NLIs as Covariates

- **NLIs can also be used as covariates (e.g., in regression analyses)**
  - Modeling assumption: *position* properties predict properties *of those who hold them*
  - Conditioning on NLI values, so dependence doesn't matter (if no error in G)
  - NLIs as *dependent* variables are much more problematic; often need to deal with autocorrelation (e.g., via ERGMs)

- **Things to keep in mind....**
  - Make sure that your theory *really* posits a direct relationship w/the NLI
  - NLI distributions could be quite skewed or irregular; be sure this makes sense (e.g., via analysis of residuals)
  - Multiple NLIs may be strongly correlated; may not be able to distinguish among related measures in practice

Graph Correlation

- **Simple way of comparing graphs on same vertex set: *graph correlation***
  - Start with *graph mean* – grand mean of adjacency matrix
  - *Graph covariance*: elementwise covariance of adjacency matrices
    - *Graph variance*: covariance of graph with itself
  - *Graph correlation*: elementwise correlation of adjacency matrices
- **Easily interpretable, works with valued data, etc.**

$$\mathbf{X}=\begin{bmatrix} - & 1 & 1 \\ 0 & - & 1 \\ 0 & 0 & - \end{bmatrix}, \mathbf{Y}=\begin{bmatrix} - & 1 & 0 \\ 0 & - & 1 \\ 0 & 0 & - \end{bmatrix}$$

$$\overline{\mathbf{X}}=\frac{\sum_{(i,j)} \mathbf{X}_{\mathbf{ij}}}{N(N-1)}=\frac{1}{2}, \overline{\mathbf{Y}}=\frac{\sum_{(i,j)} \mathbf{Y}_{\mathbf{ij}}}{N(N-1)}=\frac{1}{3}$$

$$\operatorname{Cov}(\mathbf{X}, \mathbf{Y})=\frac{\sum_{(i,j)} \left(\mathbf{X}_{\mathbf{ij}}-\overline{\mathbf{X}}\right)\left(\mathbf{Y}_{\mathbf{ij}}-\overline{\mathbf{Y}}\right)}{N(N-1)-1}$$

$$=0.2$$

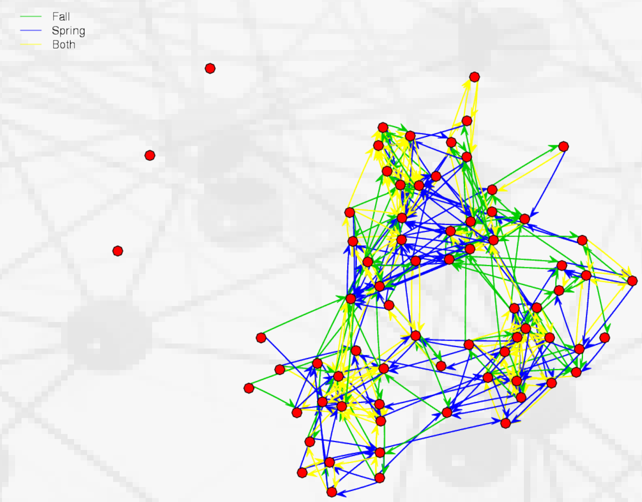
$$\operatorname{Var}(\mathbf{X})=\operatorname{Cov}(\mathbf{X}, \mathbf{X})=0.3$$

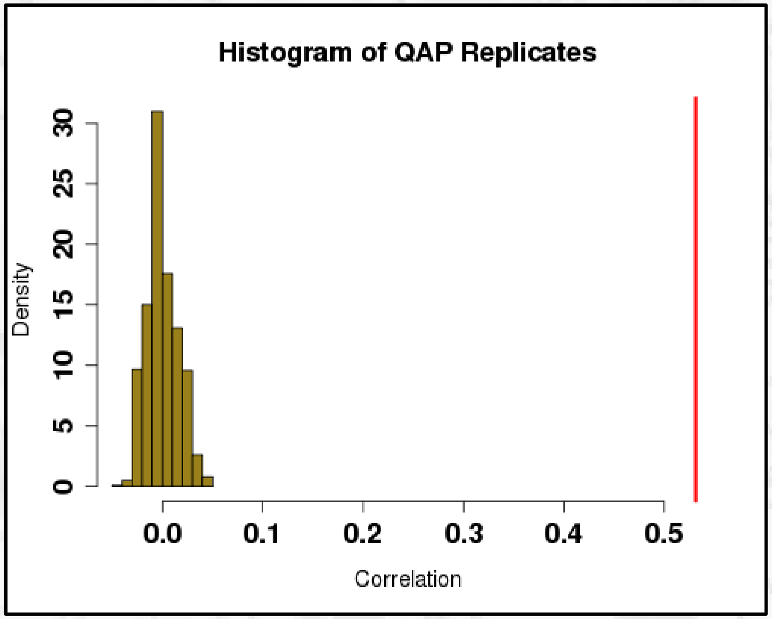
$$\operatorname{Var}(\mathbf{Y})=\operatorname{Cov}(\mathbf{Y}, \mathbf{Y})=0.27$$

$$\rho(\mathbf{X}, \mathbf{Y})=\frac{\operatorname{Cov}(\mathbf{X}, \mathbf{Y})}{\sqrt{\operatorname{Var}(\mathbf{X}) \operatorname{Var}(\mathbf{Y})}}=0.71$$

Hubert's QAP

- **How to tell if our observed correlation is "large"?**
  - Due to autocorrelation, large excursions possible
- **Hubert's QAP**
  - Fix one matrix, repeatedly permute the other
  - Compare observed correlation w/permutation distribution
  - As usual, look to the quantiles of the observed correlation to determine *p*-values
  - Interpretation: CUG test w/<u>all</u> unlabeled properties fixed





Network Regression

- **Simple family of models for predicting social ties**
  - Special case of standard OLS regression
  - Dependent variable is a network adjacency matrix
- **Model form:**

$$\mathbf{E}Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ijk} + \ldots + \beta_p X_{pij}$$

  - where $\mathbf{E}$ is the expectation operator (analogous to "mean" or "average"), Y_{ij} is the value of the edge from i to j on the dependent relation with adjacency matrix $\mathbf{Y}$, X_{kij} is the value of the kth predictor for the (i,j) ordered pair, and $\beta_0,\ldots,\beta_p$ are coefficients

Dependent Variable

- **From previous, dependent variable is an adjacency matrix**
  - Standard case: dichotomous data
    - Interpretation: model predicts tie probability (maybe not well)
  - Valued case
    - Interpretation: model predicts tie strength
- **To prepare data, just code network into adjacency matrix form**
  - No special tactics required for one-mode data
  - For two-mode data, either treat as one-mode or use projection matrix

Independent Variable(s)

- **For independent variables (X), may need to prepare data**
  - Always take matrix form, but may be based on vector data
- **Several examples:**
  - Simple adjacency matrices
  - Sender/receiver effects
  - Attribute differences
  - Elements held in common

Initial Concept: Baseline Models

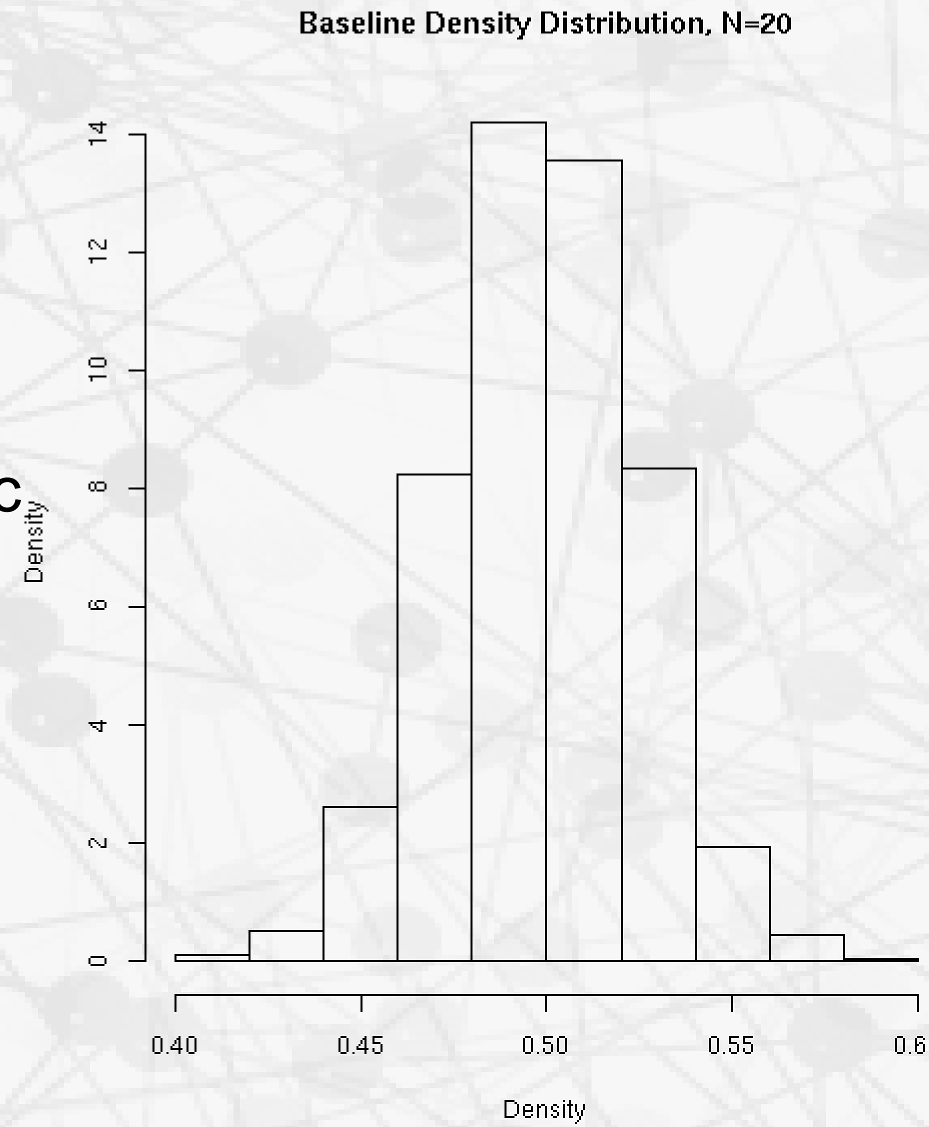
- **<u>Baseline model</u>: model which treats social structure as maximally random, given some fixed constraints**
  - Constraints could include size, density, etc.
- **Method of baseline models (from Mayhew)**
  - Identify potentially constraining factors
  - Compare observed properties to baseline model
    - Interpret deviations from baseline
    - May repeat with additional constraints
  - Note similarity to classical null hypothesis testing
    - Baseline model acts as null hypothesis
  - Useful even when baseline model is not "realistic"
    - Emphasis on triangulation to identify nature of biases; multiple baselines may be used to "pin down" complex behavior

A Few Baseline Models

- **Uniform conditional on size**
  - Given number of individuals, all structures taken to be equally likely
- **Uniform conditional on number of edges**
  - Given number of individuals and interactions, structure is random (fixes density)
  - Valued version: condition on edge values (randomize over who gets edges)
- **Uniform conditional on dyad census**
  - Given number of individuals, mutual, asymmetric, and null relationships, structure is random (fixes density and reciprocity)
  - Valued version: condition on dyad pair values (randomize over who belongs to each dyad)
- **Uniform conditional on all unlabeled properties**
  - This is the permutation distribution we saw earlier!

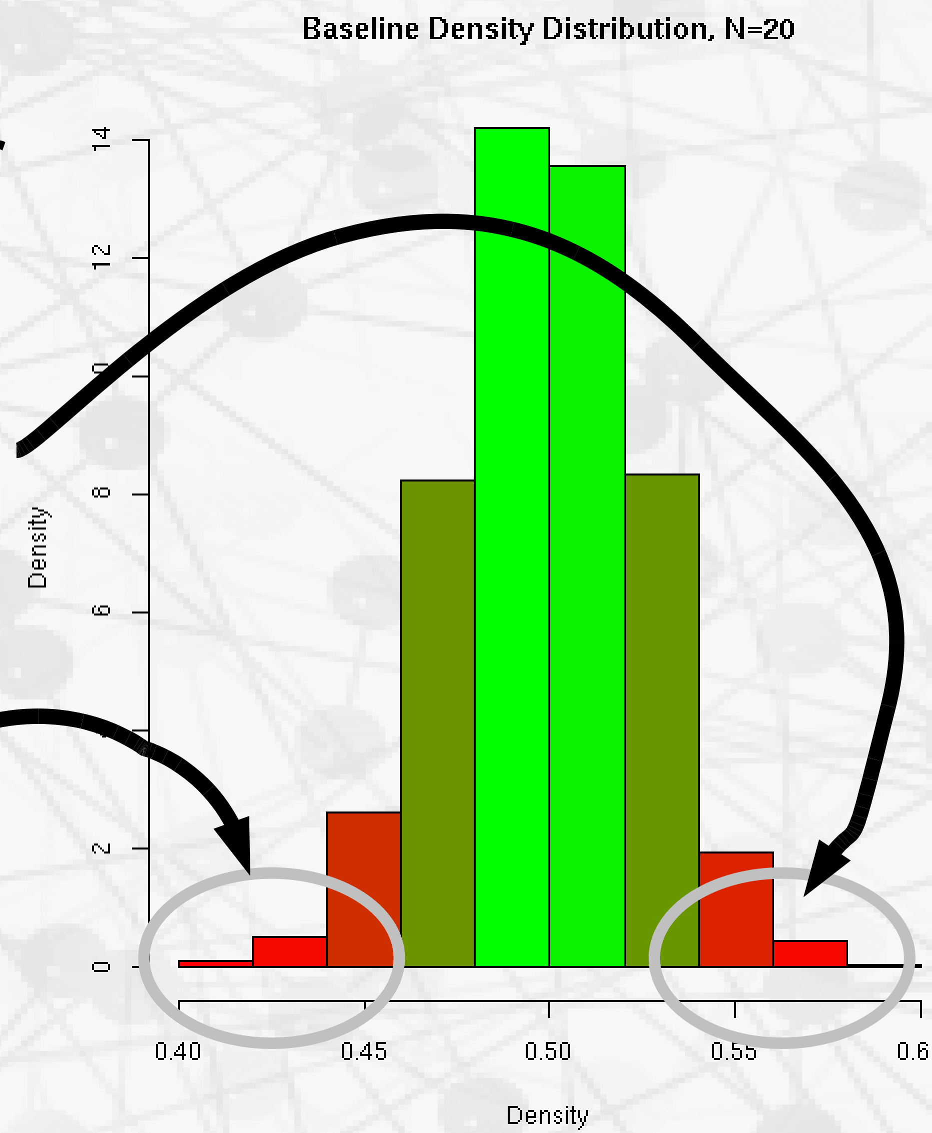
Comparing Observations to Baselines

- **To compare observations w/baseline behavior, must choose a statistic to evaluate**
  - Should choose a statistic which reflects the type of property being examined
  - Obviously, cannot use a statistic on which one is conditioning
- **Next, generate distribution of statistic under baseline model**
  - Simulate networks from baseline model, then calculate statistic
- **Finally, compare observed statistic to baseline distribution**



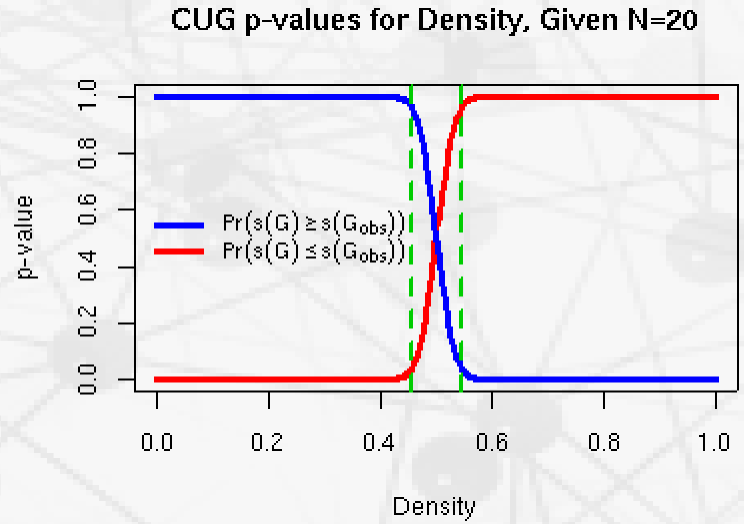
Looking High and Low

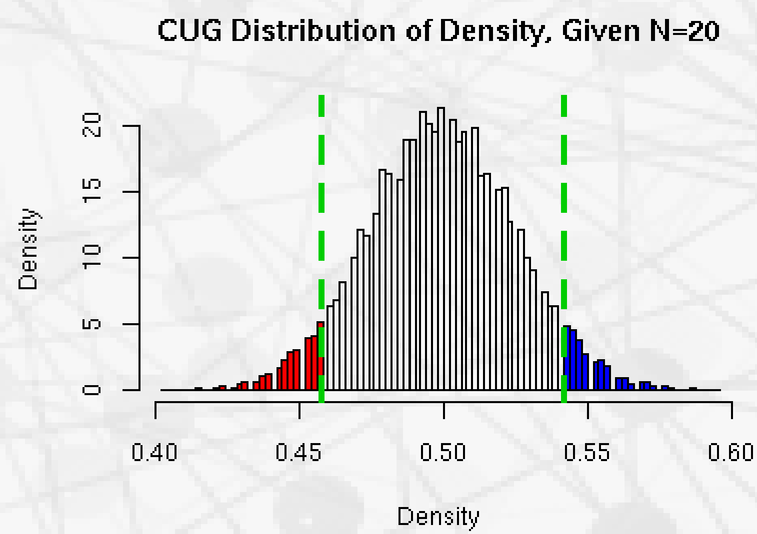
- **Primary quantities of interest: probabilities of obtaining values greater than/equal to or less than/equal to observed value under baseline model**
  - $\Pr(s(G) \geq s(G_{obs})) \approx 0$ implies that $s(G_{obs})$ is <u>large</u> compared to baseline
    - Small chance of observing a value that large under baseline
  - $\Pr(s(G) \leq s(G_{obs})) \approx 0$ implies that $s(G_{obs})$ is <u>small</u> compared to baseline
    - Small chance of observing a value that small under baseline



Conditional Uniform Graph Tests

- **CUG test: use of conditional uniform baseline as a null hypothesis test**
  - Propose that $s(G_{obs})$ drawn from a baseline model
  - Reject at significance level p if $\Pr(s(G)\geq s(G_{obs}))<p$ (upper tail) or $\Pr(s(G)\leq s(G_{obs}))<p$ (lower tail)
  - Conventional significance levels 0.05, 0.01, 0.001
- **Interpretation**
  - Rejection: Data shows noteworthy departure from model
  - Non-rejection: Data consistent with baseline model
  - Direction indicates nature of deviation

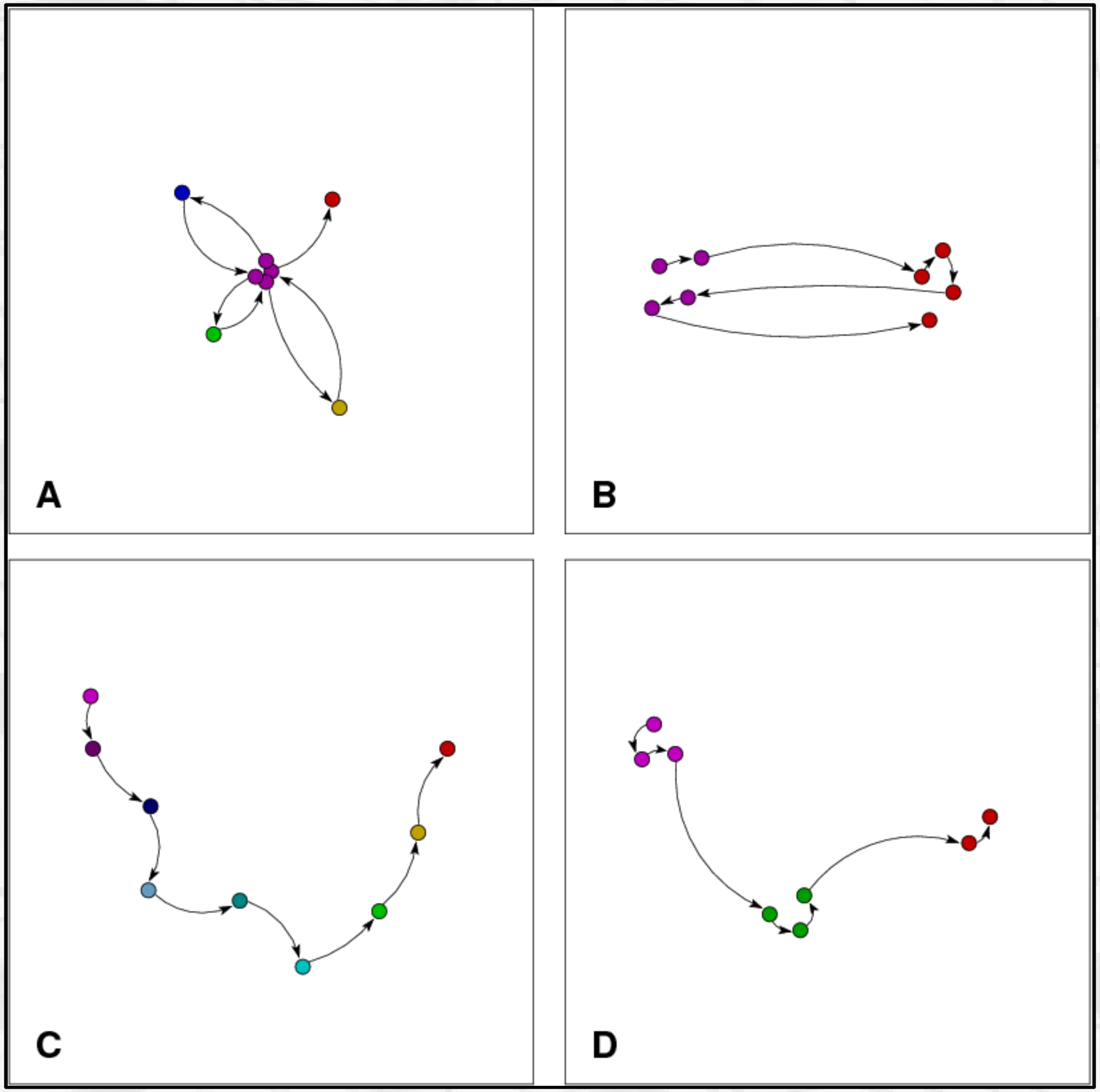




Carter T. Butts, 2010 PolNet `statnet` Workshop

Studying Qualitative Dynamics w/Network MDS

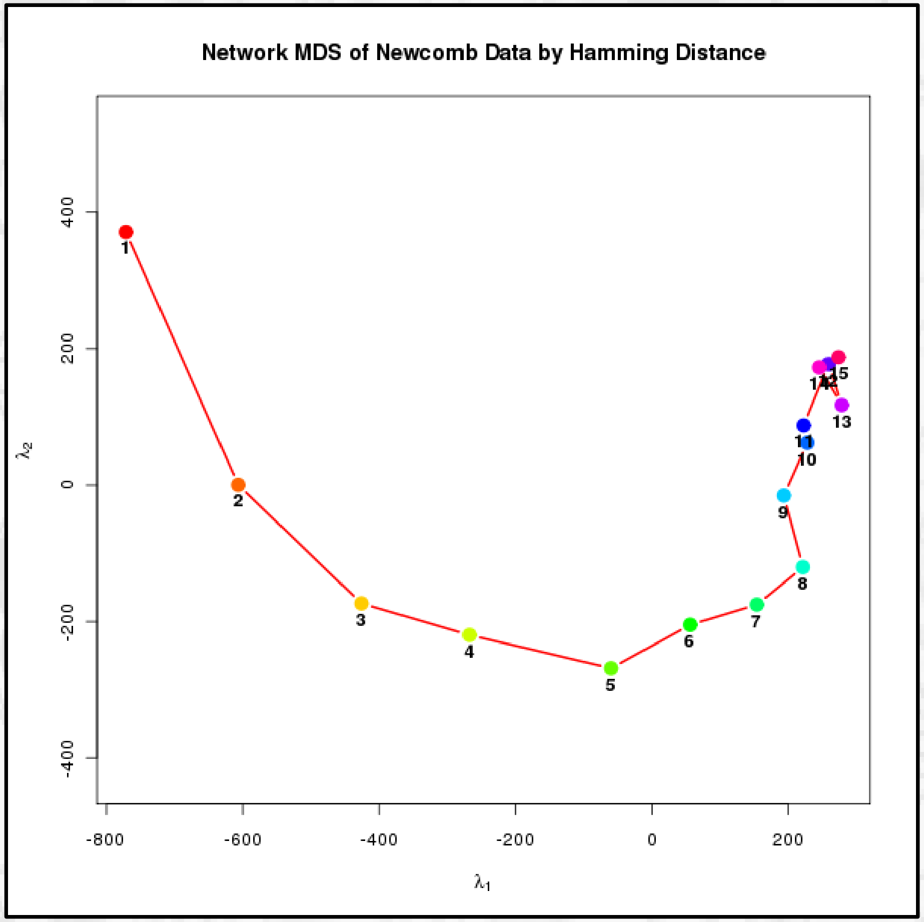
- **MDS solution can also be used to study dynamics**
  - Motion in MDS space reveals qualitative aspects of network dynamics
  - Can also get general information on pace of change
- **Good for "holistic" assessments**
  - Downside: doesn't tell you about <u>what</u> is changing



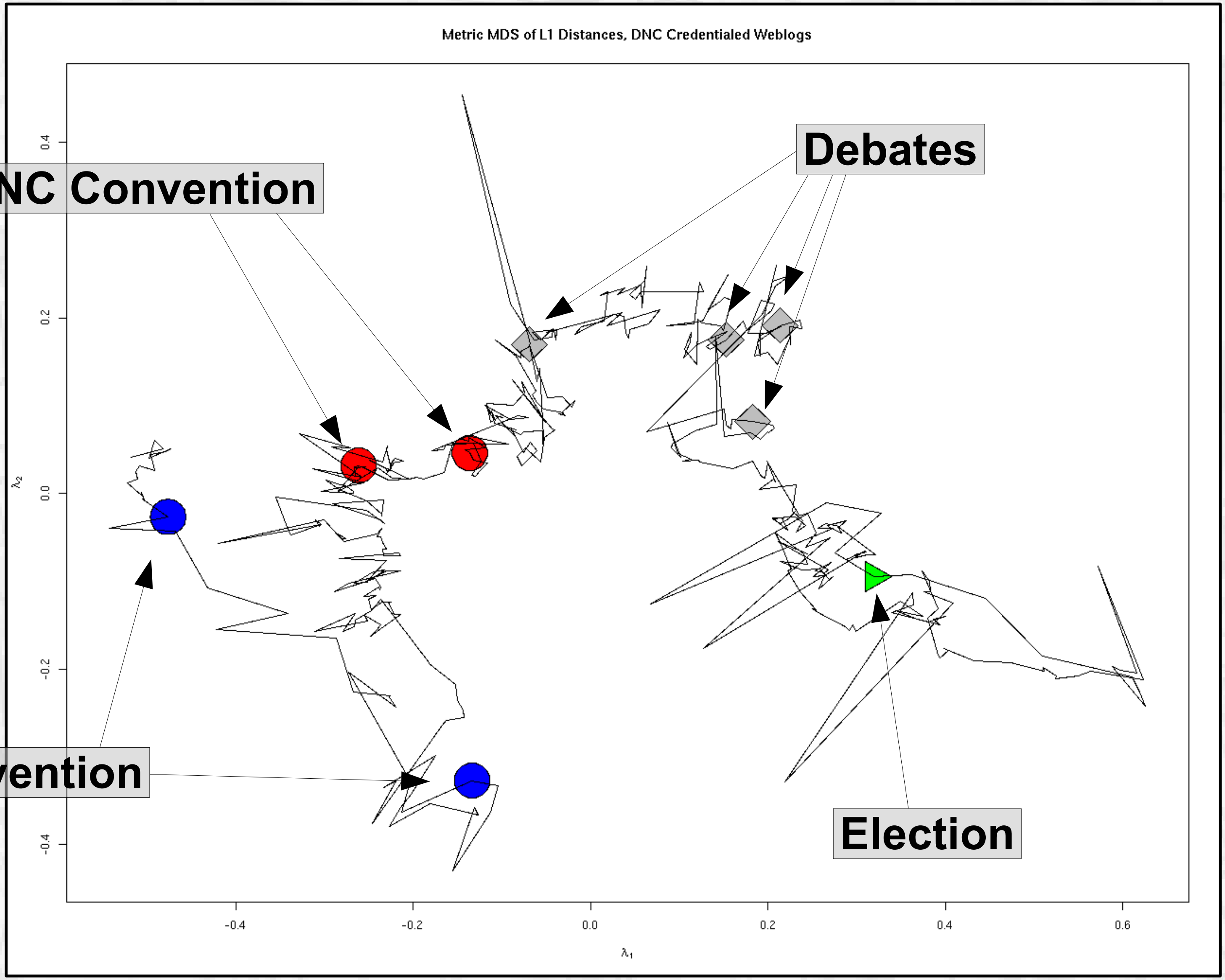
From Butts and Cross (2009)

Seriation Curves

- **Common pattern when change accumulates over time**
  - Essentially a 1-dimensional pattern
    - Don't over-interpret curvature
  - Proximity of points suggests pace of change
    - Ex: equilibration in Newcomb frat after about 11 weeks
- **As w/other examples, works well with valued data**
  - Can detect tie strength changes, even if edges fixed



Trajectories in Context



From Butts and Cross (2009)

Some Additional References

- **The following recent papers provide introductory reviews to these topics, and/or to `statnet`:**

  - Butts, Carter T. (2008). ``Social Networks: A Methodological Introduction.'' *Asian Journal of Social Psychology*, 11(1), 13-41.
  - Butts, Carter T. (2008). ``Social Network Analysis with `sna`.'' *Journal of Statistical Software*, 24(6).
  - Handcock, Mark S.; Hunter, David R.; Butts, Carter T.; Goodreau, Steven M.; and Morris, Martina. (2008). ``` ``statnet ```: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data.'' *Journal of Statistical Software*, 24(1).

- **We hope these tools aid you in your research!**