Bayesian Meta-Analysis of Network Data via Reference Quantiles

Carter T. Butts

Department of Sociology and

Institute for Mathematical Behavioral Sciences

University of California, Irvine
`buttsc@uci.edu`

UCI MURI AHM, 6/3/11

This work was supported in part by NSF award CMS-0624257 and ONR award N00014-08-1-1015.

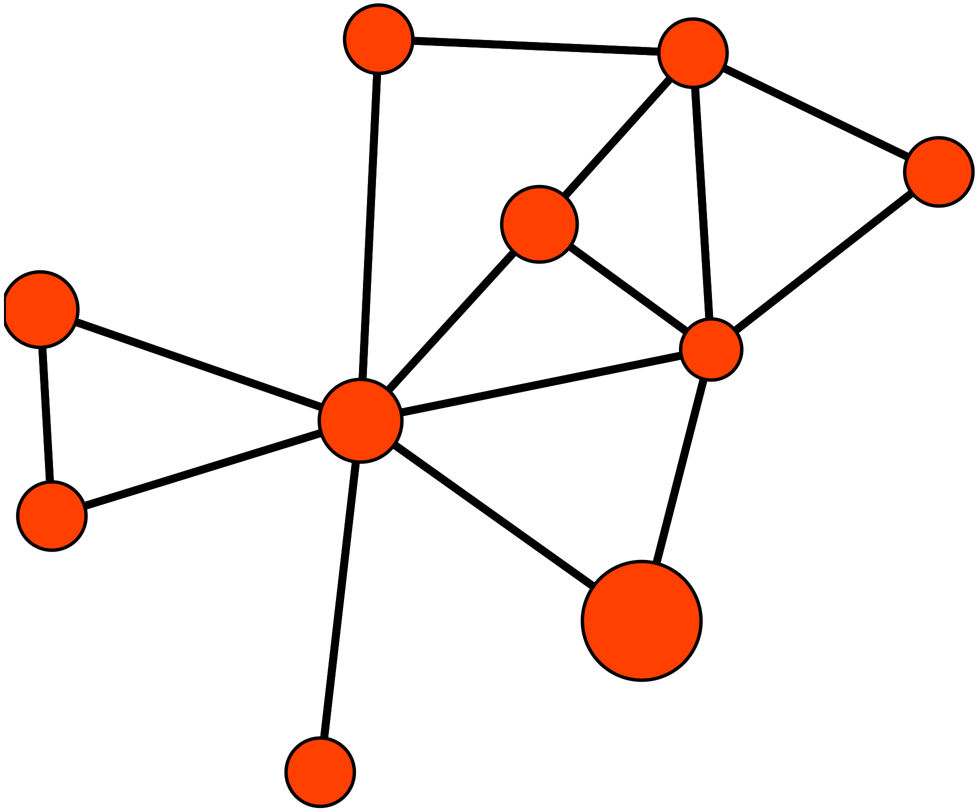
Studying Structural Populations

- Fundamental concern of social network theory: the distribution of structural properties within and across networks
  - Indicators of underlying processes (e.g., structural balance (Harary, 1959), structural dependence (Pattison and Robins, 2002))
  - Determinants of social outcomes (e.g., diffusion (Morris and Kretzschmar, 1995), resource access (Burt, 1992))
- Most existing empirical literature built on case studies, but our interest is often on a broader population/superpopulation of structures
- Increasingly, we have access to *samples* of networks (e.g., Add Health, UCDS, GSS/ISSP, online network studies like Gjoka et al. (2010))
- Data thus exists to make progress on population questions, but we need methods that can support this objective
  - Today, one approach that builds on classic reference quantile methods to allow inference for graph populations

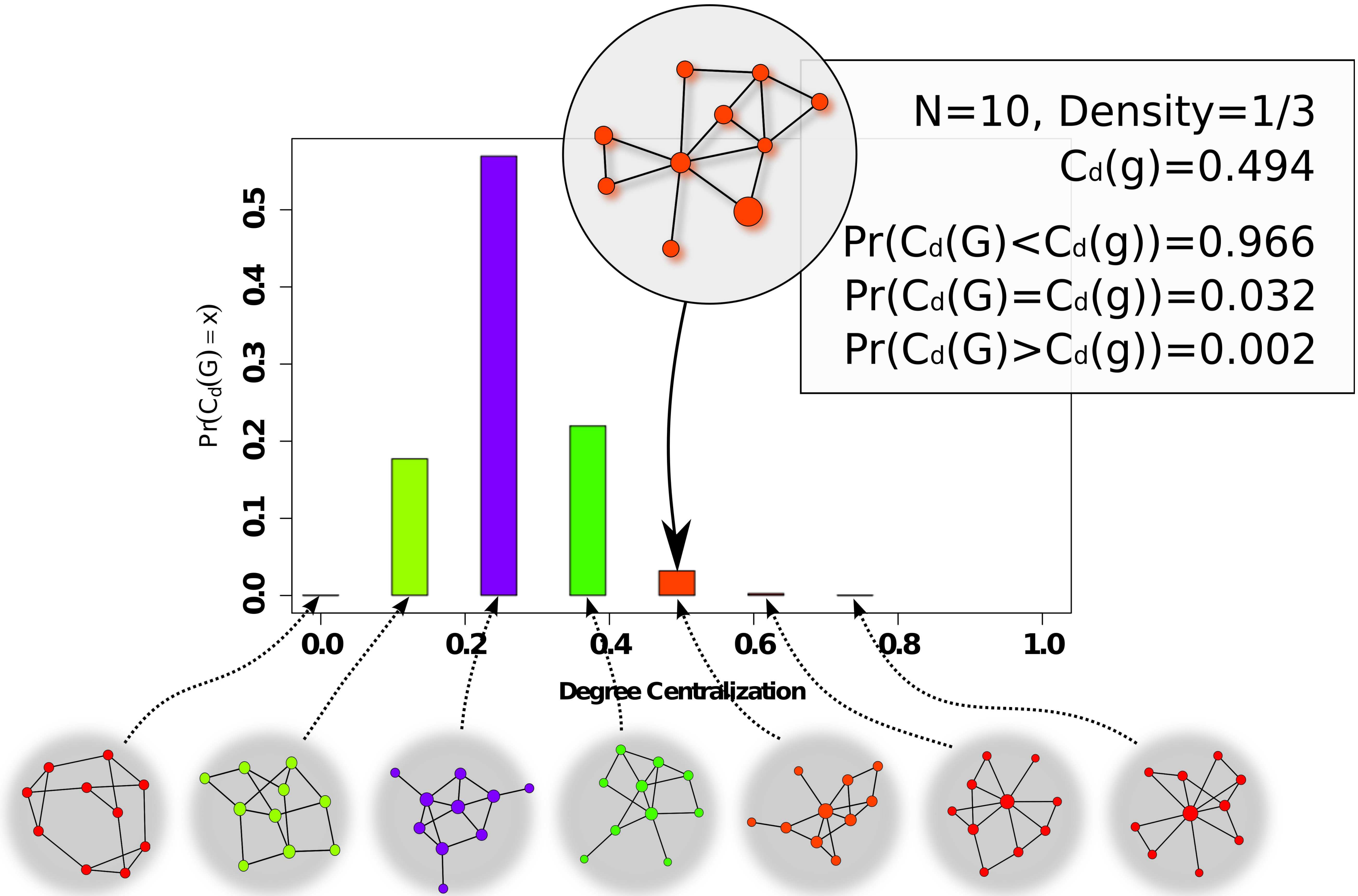
Roadmap

- Introduction
- Background: Detecting Structural Bias w/Reference Quantiles
- Extension to Multiple Networks
- Examples
- Conclusion

Question: Is This a Centralized Network?



Assessing Structure w/Reference Quantiles



Side Comment: ERGs vs RQs

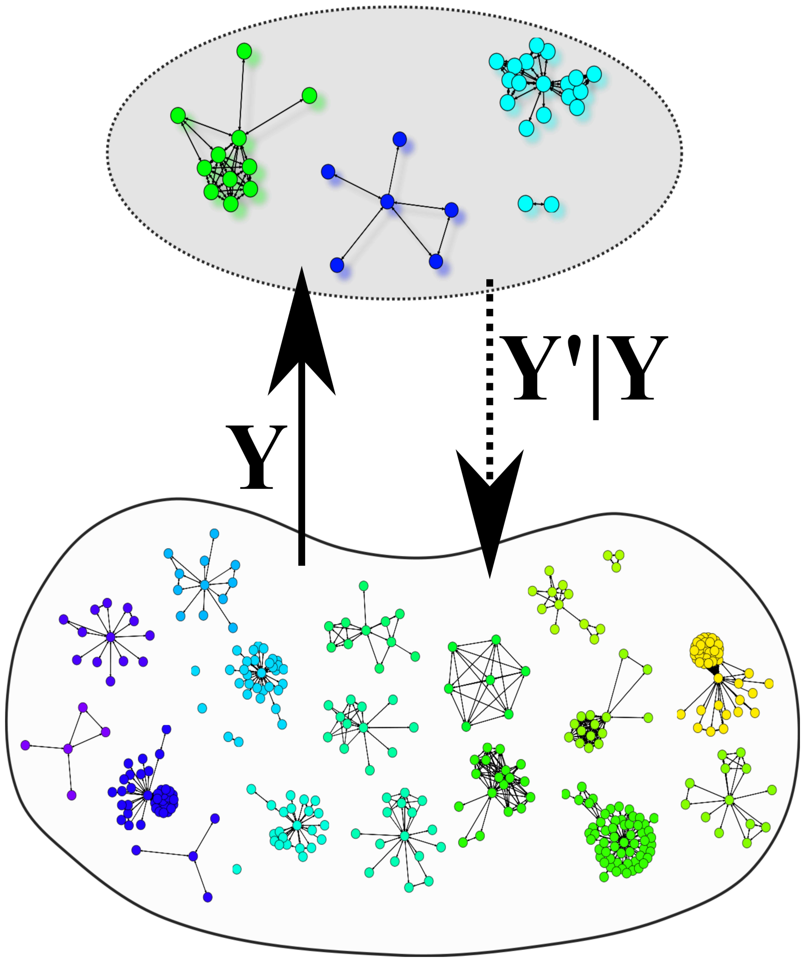
- Why not specify a parametric model for the network, and measure structural biases that way?
- We can (and do) using exponential family models (ERGs), but this is very hard:
  - Good specifications often hard to find - unlike the CUG case, bad models can lead to inaccurate results for purely computational reasons (other problems notwithstanding)
  - Often requires worrying about large numbers of nuisance parameters, when we want only a single relationship
  - Difficult to scale – large systems computationally hard, small ones have convex hull problems
  - Natural parameters usually lack scale-independent interpretation; makes size-heterogeneous comparison difficult
  - Most viable models require detailed structural information; with rare exceptions, effectively need the whole network
- Bottom line: there is still something to be said for RQs when you have a straightforward question
  - And note that you can use ERGs as reference models – this was one of the reasons they were invented! (See Holland and Leinhardt, 1981)

Roadmap

- Introduction
- Background: Detecting Structural Bias w/Reference Quantiles
- Extension to Multiple Networks
- Examples
- Conclusion

From Single Networks to Populations

- Imagine that we have a sample of networks drawn from some population of interest
- We'd like to assess the prevalance of some structural bias within the population (recognizing that there may be heterogeneity)
- Natural approach: model the reference quantiles, employ predictive distributions
  - Allows arbitrary questions of the form "if I drew another network from this population/process, what would its quantiles be?"
  - Can compare populations via their distributions
  - Questions still posed in terms of the quantiles, which we understand



Modeling the Population

- Assume we have a set of N quantile estimates, $\mathbf{Y}$, from graphs drawn at random from population of interest
  - Can include sampling mechanisms, but am not going there today.
  - Will treat estimated quantiles as exact – can extend, but this has few benefits and many costs
- As we know little about the generating process, we would like to treat distribution of quantiles as maximally dispersed, given the observed central tendency, data size, and support constraints
  - Maximum entropy interpretation of above implies an exponential family; in this case the Dirichlet distribution is a natural choice:

$$p(\mathbf{Y}|\theta) = \prod_{i=1}^{N} \frac{\Gamma(\sum_{j=1}^{3} \theta_j)}{\prod_{j=1}^{3} \Gamma(\theta_j)} \prod_{j=1}^{3} Y_{ij}^{\theta_j - 1}$$

  - Sufficient stats are the log products of quantile values, and N (compare to Fisher's method)

Prior Structure for the Population Model

- As before, want to maximize entropy. That again suggests exponential families (Jaynes, 1983); since our likelihood is also an exponential family, this can be satisfied by the conjugate prior:
  - Let $\theta, \phi \in (\mathbb{R}^{+})^{3}$, $\gamma \in \mathbb{R}^{+}$. Then

$$p(\theta|\phi,\gamma) \propto \exp\left[(\theta-1)^{T}\phi + \gamma\left(\ln\Gamma\left(\sum_{j=1}^{3}\theta_j\right) - \sum_{j=1}^{3}\ln\Gamma\left(\theta_j\right)\right)\right]$$

  - Prior parameters are interpretable as pseudo-data: equivalent of log quantile product (ϕ) and prior data size (γ). Can be helpful to think of $\exp(\phi/\gamma)$ as prior pseudo-geometric mean.
  - No analytical normalization – have to resort to computation. Sufficient condition for propriety is $\max\phi_i < -\gamma\ln 3$.
  - Uninformative prior obtained as limit $\gamma, \phi_i \to 0$ with $\phi_1 = \phi_2 = \phi_3$
  - Reasonable weakly informative prior: $\gamma = 1, \phi_i \approx -\ln 4.5$; corresponds to 1 observation w/geometric mean matching uniform Dirichlet

Posterior Analysis

- Since this is a conjugate prior, posterior is straightforward:
  - Given $\theta, \psi \in (\mathbb{R}^+)^3, \gamma \in \mathbb{R}^+$

$$p(\theta|\mathbf{Y}, \phi, \gamma) \propto \exp\left[(\theta - 1)^T \left(\phi + \sum_{i=1}^{N} \ln \mathbf{Y}_i\right) + (\gamma + N)\left(\ln \Gamma\left(\sum_{j=1}^{3} \theta_j\right) - \sum_{j=1}^{3} \ln \Gamma\left(\theta_j\right)\right)\right]$$

  - Normalizing factor again must be computed numerically
- Need to be able to simulate; two approaches so far
  - MCMC: easy to implement, but slow and approximate
  - Rejection sampling: harder to perform, but faster and exact

Posterior Predictive Analysis

- We rarely want the distribution of θ, but rather the associated distribution of quantiles (the posterior predictive)
  - Draw $\theta^{(i)}$ from $\theta|\mathbf{Y}, \psi, \gamma$, then draw $\mathbf{y}'^{(i)} \sim \mathrm{Dirichlet}(\theta^{(i)})$
  - Interpretable as what one would be predicted to observe, if one drew another case from the same population (in the same way)
- Useful for asking many sorts of questions
  - "What's the probability that G will have $p < 0.05$ on this statistic?"
  - "What's the probability that an arbitrary member of population A will be more extreme than a member of population B on a given statistic?"

Complicating Matters: Additional Heterogeneity

- Our structure is as diffuse as possible, but only given the central tendency/data size
- What if the population contains a mix of types with multiple internal modes?
- One option: assume a finite mixture model
  - Can do, but choosing the number of mixture components is slow and difficult; want to keep analysis as automated as possible
- Alternative: infinite mixtures using nonparametric priors
  - Start w/K-class mixture of homogeneous models:

$$\begin{aligned} \lambda &\sim \mathrm{Dirichlet}(\alpha/K, \ldots, \alpha/K) \\ \eta_i &\sim \mathrm{Multinom}(\lambda) \\ \mathbf{\Theta}_{\eta_i \cdot} &\sim \theta|\phi, \gamma \\ \mathbf{Y}_{\mathbf{i}\cdot} &\sim \mathrm{Dirichlet}(\mathbf{\Theta}_{\eta_i \cdot}) \end{aligned}$$

  - Taking the limit of this mixture as $K \to \infty$ leads to a *Dirichlet process* prior with base measure equivalent to the homogeneous prior and concentration parameter α
  - Intuition: total population composed of a large group of subpopulations (potentially unlimited, but finite in practice), having unknown sizes/composition
    - Prior expected Herfindahl index of the size distribution is $1/(1+\alpha)$ (follows from Anderson (1990))

Posterior Simulation

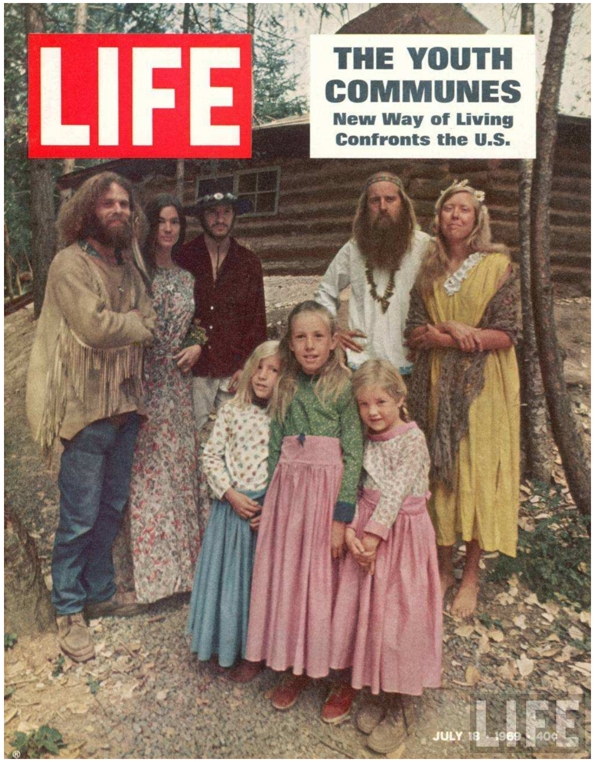
- Here use the Gibbs sampling algorithm of Neal (2000, Alg. 2)
  - Requires ability to draw from posterior and to calculate Bayes factors
  - Use rejection sampling for former, numerical quadrature for latter
- Idea: alternate between "grouping" data points by cluster co-membership and drawing parameters per cluster (based on allocated data points); this eventually converges to target distribution
  - Conditional probability of assigning $\mathbf{Y}_{i\cdot}$ to existing cluster j w/prob $\propto \frac{\kappa_j^{-i}}{N-1+\alpha}\mathrm{Dirichlet}(\mathbf{Y}_{i\cdot}|\mathbf{\Theta}_{j\cdot})$, w/$\kappa_j^{-i}$ the number of current members; assign to new cluster w/prob $\propto \frac{\alpha}{N-1+\alpha}\iiint_{(\mathbb{R}+)^3}\mathrm{Dirichlet}(\mathbf{Y}_{i\cdot}|\theta)p(\theta|\phi,\gamma)d\theta$
  - Per-cluster parameters $\mathbf{\Theta}_{j\cdot}$ drawn from homogeneous model posterior, given assigned members
  - At each round, select existing cluster j's parameter for $\theta^{(i)}$ w/prob $\kappa_j/(N+\alpha)$, or draw from prior with prob $\propto \alpha/(N+\alpha)$
- Given posterior draws $\theta^{(i)}$, can take posterior predictives as usual

Roadmap

- Introduction
- Background: Detecting Structural Bias w/Reference Quantiles
- Extension to Multiple Networks
- Examples
- Conclusion

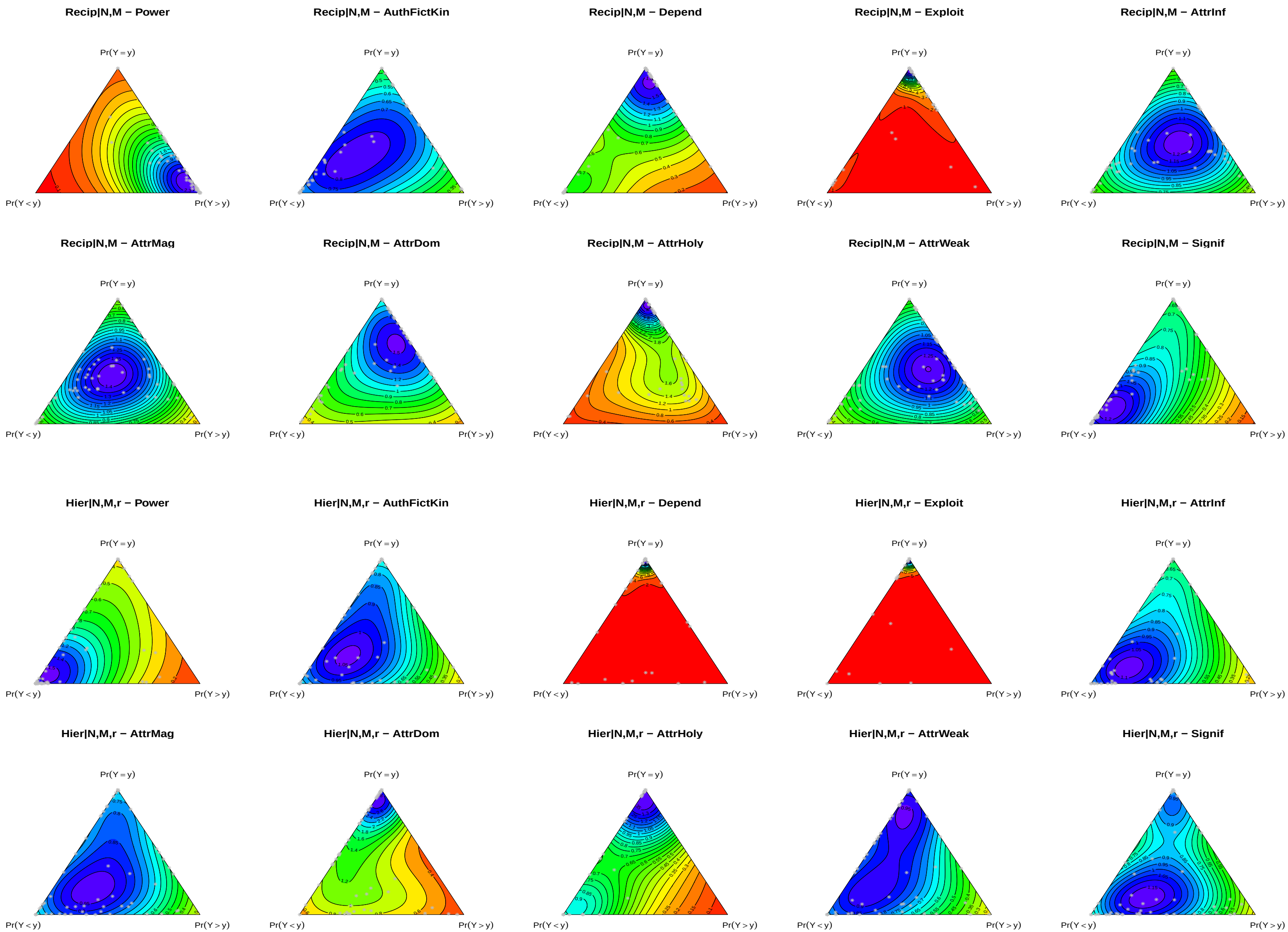
Example: Hierarchy and Reciprocity in Urban Communes

- Motivation: structure of "power-active" relations in informal organizations (Reich and Butts, 2006)
  - To what extent are potential power "conduits" biased in ways that would allow/inhibit local vs. extended exercise of power?
- Data: Urban Communes Data Set (Zablocki, 1980)
  - 61 urban communes, ranging from 4 to 26 members
  - 10 relations
    - **Power:** ego claimed to exercise power over alter
    - **AuthFictKin:** ego occupies superordinate fictive kin role vs. alter
    - **Depend:** alter depends on ego more than vice versa
    - **Exploits:** ego exploits alter
    - **AttrInf:** ego regarded by alter as influential
    - **AttrMag:** ego regarded by alter as "sexy"/"charismatic"
    - **AttrDom:** ego regarded by alter as dominant
    - **AttrHoly:** ego regarded by alter as "holy"
    - **AttrWeak:** alter regarded by ego as"passive"/"dependent"
    - **Signif:** alter regards ego as "significant person" in own life



- Analysis using heterogeneous population model
  - Quantiles: edgewise reciprocity given size, density; Krackhardt hierarchy given size, dyad census (1e5 MC draws)
  - Weak "neutral" priors ($\phi_i = -\ln(4.5), \gamma = 1, \alpha = 1$) employed; pulls towards the null hypothesis
  - Posterior predictive simulation using MCMC (5 chains, 150 burn-in draws/chain, 2000 draws)

Posterior Predictives, UCDS



Posterior Predictives for Extreme Quantiles, UCDS

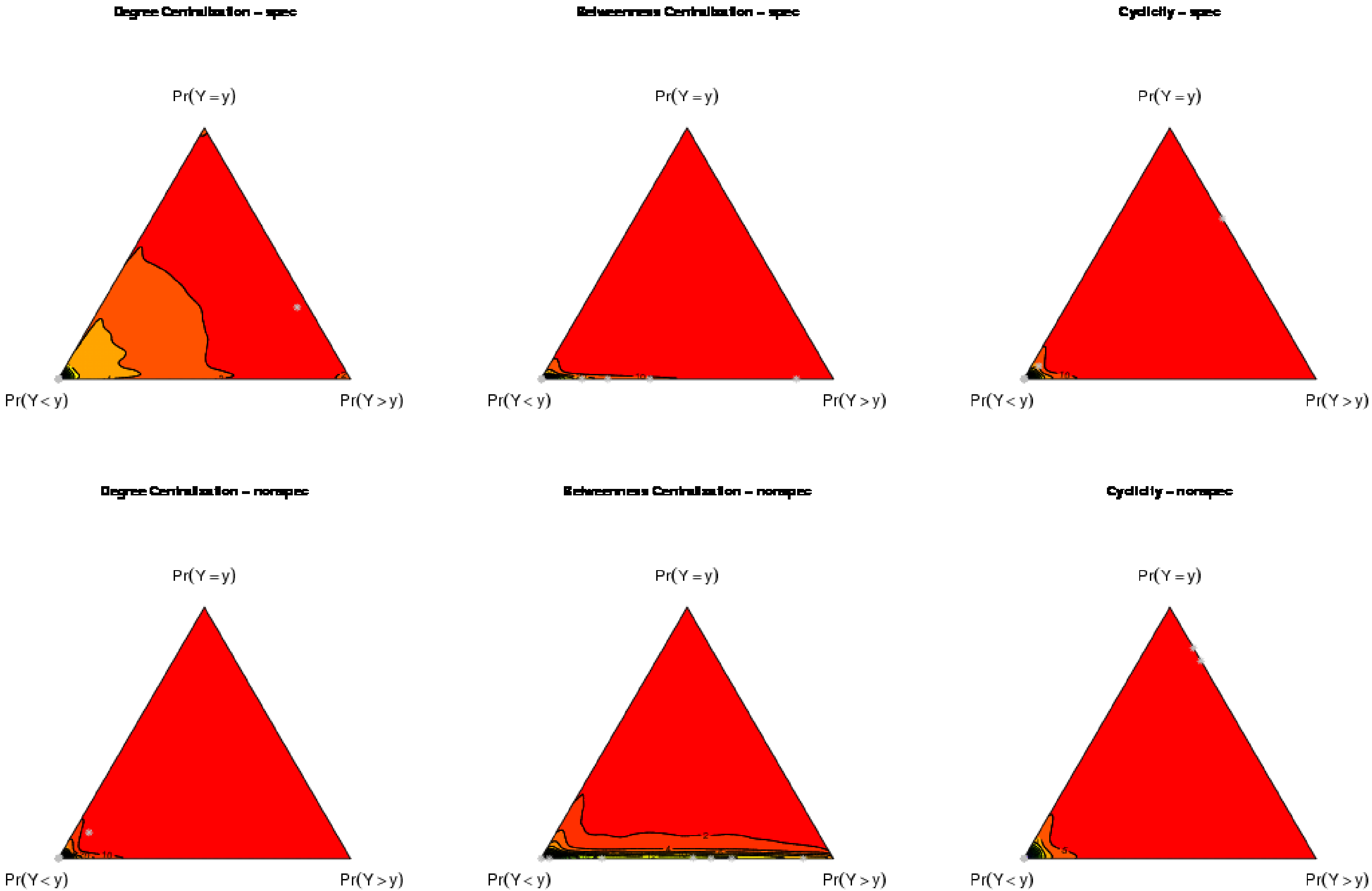
Relation	Reciprocity | Density		Hierarchy | Dyad Census	
	$\ln\left(\frac{\Pr(\mathrm{sig.low})}{0.05}\right)$	$\ln\left(\frac{\Pr(\mathrm{sig.high})}{0.05}\right)$	$\ln\left(\frac{\Pr(\mathrm{sig.low})}{0.05}\right)$	$\ln\left(\frac{\Pr(\mathrm{sig.high})}{0.05}\right)$
Power	1.82	-3.91	-3.91	1.89
AuthFictKin	-1.27	0.87	-1.83	0.69
Depend	-3.22	0.96	-2.81	-2.30
Exploit	-2.81	-1.43	-3.91	-1.27
AttrInf	-1.71	-0.69	-2.12	0.91
AttrMag	-2.30	-0.20	-1.83	0.54
AttrDom	-2.30	-1.61	-2.53	-1.27
AttrHoly	-3.22	-2.53	-3.91	1.34
AttrWeak	-2.12	-0.87	-1.71	0.85
Signif	-3.91	1.41	-2.12	0.25

(Blue indicates odds ratios ≥ 2; red indicates odds ratios $\leq 1/2$)

Example: Centralization and Cyclicity in WTC Radio Networks

- Motivation: structure of emergent communication networks in emergencies (Butts et al., 2007)
  - How do organizations balance the trade-off between efficiency gains from centralization and the need for multilateral negotiation to resolve complex dependencies?
  - Does the approach taken vary by organization type?
- Data: WTC emergency-phase radio networks (Butts et al., 2007)
  - 17 networks of radio communications among responders at WTC, PA-NYNJ, Newark Airport sites
  - 9 networks from "specialist" organizations, 8 from "non-specialists"
  - Sizes range from 24 to 256 persons (median 118)
- Analysis via homogeneous population models for specialist, non-specialist organizations
  - Quantiles: degree/betweenness centralization, cyclicity given size, dyad census (2000 MC draws)
  - Weak "neutral" priors ($\phi_i = -\ln(4.5), \gamma = 1, \alpha = 1$)
  - Posterior predictive simulation using rejection sampler (5000 draws per model)

Posterior Predictives, WTC



Posterior Predictives for Extreme Quantiles, WTC

GLI	Responder Type	$\ln \frac{\Pr(\text{Sig. Low})}{0.05}$	$\ln \frac{\Pr(\text{Sig. High})}{0.05}$	$\ln \frac{\Pr(\text{Sig. High})}{\Pr(\text{Sig. Low})}$
Degree	Specialist	-1.83	2.04	3.91
Centralization	Non-Specialist	-3.91	2.24	6.75
Betweenness	Specialist	-2.12	1.84	3.89
Centralization	Non-Specialist	-0.97	1.27	2.24
Cyclicity	Specialist	-0.65	1.77	2.42
	Non-Specialist	0.28	1.53	1.26

(Blue indicates odds ratios ≥ 2; red indicates odds ratios $\leq 1/2$)

Roadmap

- Introduction
- Background: Detecting Structural Bias w/Reference Quantiles
- Extension to Multiple Networks
- Examples
- Conclusion

Conclusion

- Studying the properties of graph populations is important, and we need practical solutions
- Reference quantile method a good starting point: simple, scalable, few data requirements, easily understood
- Can build Bayesian superstructure on reference quantile scheme to study populations in a principled way
  - Rely on maximum entropy distributions as a weakly informative default
  - Can accomodate additional heterogeneity using nonparametric priors
  - Simulation practical using MCMC/rejection; note that difficulty does *not* increase with graph size!
- Of course, challenges remain....
  - Faster computation for rejection sampler
  - Assessing robustness to prior selection (particularly α)
  - Predictive accuracy evaluation
  - Methods for integrating data quality measures

The End

Thanks for your attention!

(For more details, see the forthcoming paper in *Sociological Methodology,* vol 41 (2011).)

Some Desiderata

- A few properties we'd like our methods to have:
  - Principled – should be backed up by some reasonable theory of inference
  - Scalable – should be practical for very small and fairly large networks
  - Relatively unsupervised – should minimize the degree of analyst intervention/expertise needed to make things work
- Also important: suitability for "meta-analytic" use (loosely defined)
  - Basis for reasoning from sample of networks to some more general case (population or super-population)
  - Should have minimal per-case data requirements (e.g., extractable from published reports or other limited disclosure)
  - Should be reasonably robust to poorly understood heterogeneity
  - Should allow for use of weak prior information (aka "non-informative" analysis), and/or conservative estimation for extreme outcomes

A Starting Point: Baseline Models

- *Baseline models* (Mayhew, 1984): a way of accounting for opportunity and constraint
  - Intuition: fix specified properties, assume system state is uniform given those properties
    - Compare with maximum entropy, thermodynamic models (e.g., Jaynes (1983))
  - Of particular interest are the *conditional uniform graphs*:
    - Let $\mathbb{G}$ be the set of all graphs, and let $f : \mathbb{G} \mapsto \mathbb{R}$ be a *graph statistic*. The *uniform graph distribution conditional on* f, τ is defined by the pmf

$$\Pr(G = g|f, \tau) = \left|\left\{g' \in \mathbb{G} : f\left(g'\right) = \tau\right\}\right|^{-1} \mathbb{I}\left(f\left(g\right) = \tau\right), \tag{1}$$

where $\mathbb{I}$ is the standard indicator function.

- Sometimes used directly as an approximation, more often as a point of reference
  - Incorporate factors that should be taken into account before reaching for more complex explanations
  - Easy to work with, understand

Baseline Models and Reference Quantiles

- *Method of baseline modeling*: assess system by reference to its baseline behavior
  - Classic approach (see, e.g. Davis, 1970; Holland and Leinhardt, 1970; 1972; Mayhew, 1984; Snijders and Stokman, 1987; Anderson et al., 1999)
  - Key tool: *reference quantiles*
    - Let g be an observed graph, M a reference model, and f a statistic of substantive interest. Then the reference quantiles for $f(g)$ with respect to M are defined as $\Pr\left(f(G) < f\left(g\right)\right)$, $\Pr\left(f(G) = f\left(g\right)\right)$, and $\Pr\left(f(G) > f\left(g\right)\right)$, where $G \sim M$
    - Intuition: extremity of reference quantiles for $f(g)$ tell us how the properties of g relate to what one would expect to see from its baseline characteristics
      - Can be viewed as a index of the level of f versus the baseline distribution
      - Compare similar use in classical null hypothesis testing, model adequacy checks
- Primitive, but has some important virtues:
  - Well-understood approach; relatively easy for non-specialists to grasp
  - Scalable; baselines (e.g., CUGs) often easy to simulate for large systems
  - Minimal data requirements; given conditioning statistics (e.g., size, mean degree) and $f(g)$, one can reconstruct reference quantiles. (Useful in meta-analytic settings!)

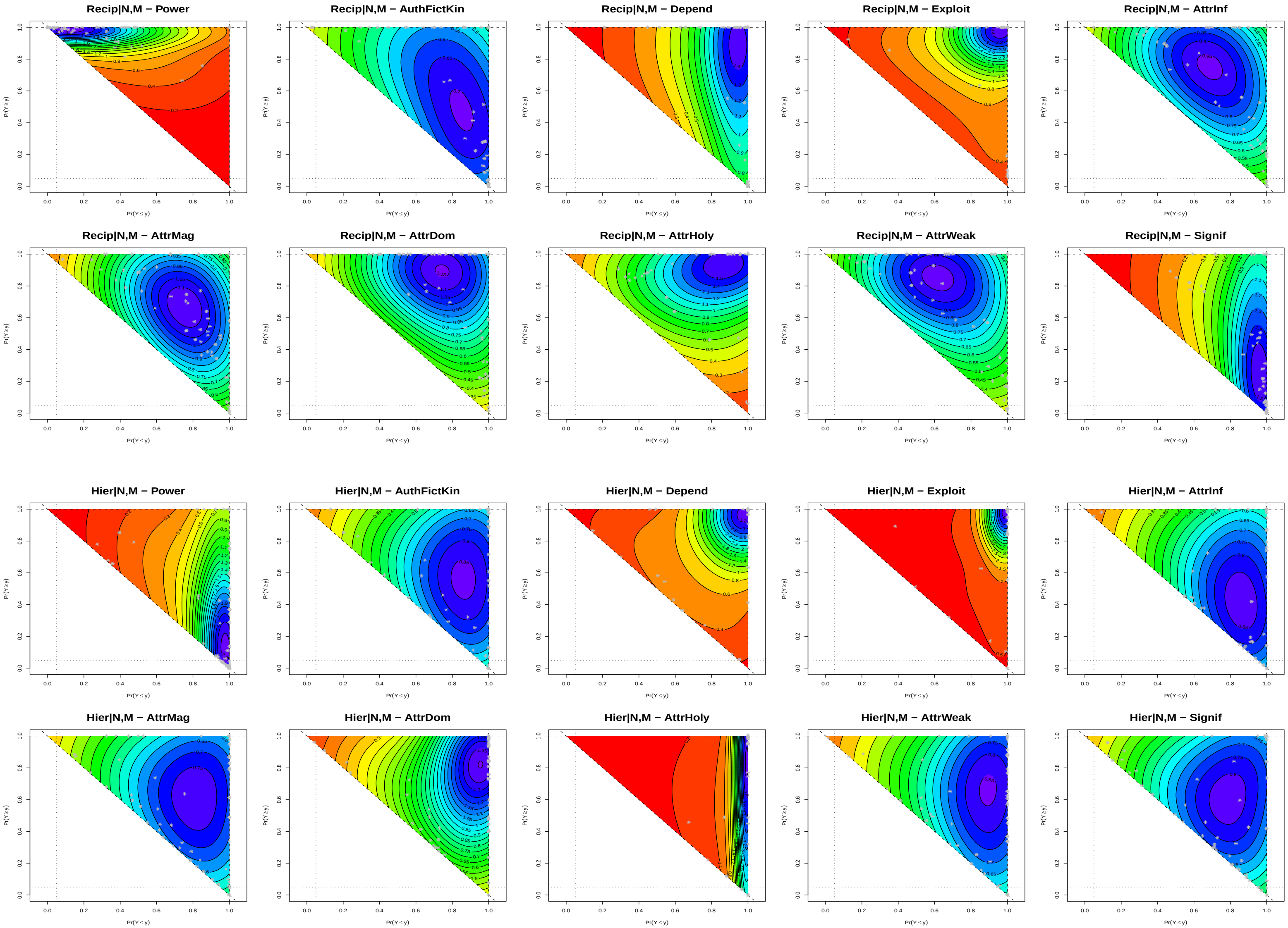
First Things: Estimating the Quantiles

- In general, this is taken for granted:
  - Simulate n draws $g^{(1)}, \ldots, g^{(n)}$ from the reference distribution and compute

$$\mathbf{x} = \left(\sum_{i=1}^{n} \mathbb{I}\left(f(g^{(i)}) < f(g) \right), \sum_{i=1}^{n} \mathbb{I}\left(f(g^{(i)}) = f(g) \right), \sum_{i=1}^{n} \mathbb{I}\left(f(g^{(i)}) >' (g) \right) \right)$$

  - Then use $\mathbf{x}/n$ as estimator of true quantile vector, ψ (this is the trivial MLE)
- In fact, not so simple: easy to lose precision at extremes (especially w/large graphs and small simulation runs)
- Simple Bayesian alternative
  - Likelihood of $\mathbf{x}|\psi$ multinomial, so choose uninformative prior for ψ and employ $\hat{\psi} = \mathbf{E}\psi|\mathbf{x}$
    - Several choices; here suggest Jeffreys prior, which is Dirichlet$(1/2, 1/2, 1/2)$; posterior given $\mathbf{x}$ is Dirichlet$(x_1 + 1/2, x_2 + 1/2, x_3 + 1/2)$
  - Using posterior mean gives us $\hat{\psi} = (\mathbf{x} + 1/2)/(3/2 + n)$
    - Non-extreme quantiles don't change much; "shrinks" extremes, but effect falls with n. Total effect is on par w/1.5 data points.
- BTW, you can use this on any similar problem – take it home w/you!

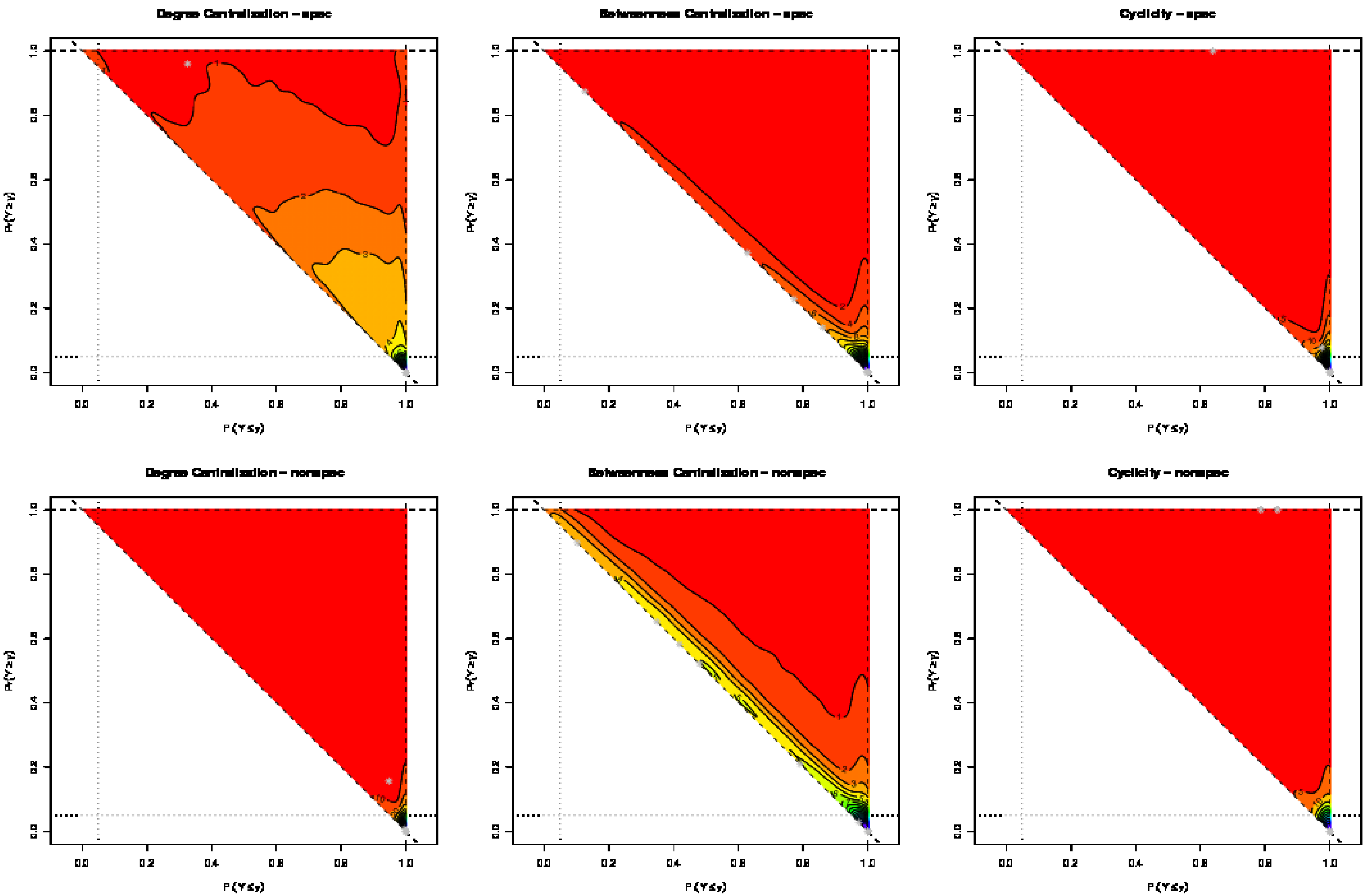
Posterior Predictive p-values, UCDS



Raw Posterior Predictives for Extreme Quantiles, UCDS

Relation	Reciprocity | Density		Hierarchy | Dyad Census	
	$\Pr(\text{sig.low})$	$\Pr(\text{sig.high})$	$\Pr(\text{sig.low})$	$\Pr(\text{sig.high})$
Power	0.310	0.001	0.001	0.331
AuthFictKin	0.014	0.119	0.008	0.100
Depend	0.002	0.130	0.003	0.005
Exploit	0.003	0.012	0.001	0.014
AttrInf	0.009	0.025	0.006	0.124
AttrMag	0.005	0.041	0.008	0.086
AttrDom	0.005	0.010	0.004	0.014
AttrHoly	0.002	0.004	0.001	0.198
AttrWeak	0.006	0.021	0.009	0.117
Signif	0.001	0.204	0.006	0.064

Posterior Predictive p-values, WTC



Raw Posterior Predictives for Extreme Quantiles, WTC

GLI	Responder Type	$\Pr($Sig. Low$)$	$\Pr($Sig. High$)$	$\frac{\Pr(\text{Sig. High})}{\Pr(\text{Sig. Low})}$
Degree	Specialist	0.008	0.385	50.039
Centralization	Non-Specialist	0.001	0.470	857.555
Betweenness	Specialist	0.006	0.316	48.688
Centralization	Non-Specialist	0.019	0.178	9.368
Cyclicity	Specialist	0.026	0.295	11.191
	Non-Specialist	0.066	0.232	3.531

1 References

Anderson, B. S., Butts, C. T., and Carley, K. M. (1999). The interaction of size and density with graph-level indices. *Social Networks*, 21(3):239–267.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Erlbaum, Hillsdale, NJ.

Burt, R. S. (1992). *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, MA.

Butts, C. T., Petrescu-Prahova, M., and Cross, B. R. (2007). Responder communication networks in the World Trade Center Disaster: Implications for modeling of communication within emergency settings. *Journal of Mathematical Sociology*, 31(2):121–147.

Davis, J. A. (1970). Clustering and hierarchy in interpersonal relations: Testing two theoretical models on 742 sociograms. *American Sociological Review*, 35:843–852.

Gjoka, M., Kurant, M., Butts, C. T., and Markopoulou, A. (2010). Walking in Facebook: A case study of unbiased sampling of OSNs. *Proceedings of the 29th IEEE Conference on Computer Communications (IEEE INFOCOM)*.

Harary, F. (1959). On the measurement of structural balance. *Behavioral Science*, 4:316–323.

Holland, P. W. and Leinhardt, S. (1970). A method for detecting structure in sociometric data. *American Journal of Sociology*, 70:492–513.

Holland, P. W. and Leinhardt, S. (1972). Some evidence on the transitivity of positive interpersonal sentiment. *American Journal of Sociology*, 72:492–513.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association*, 76(373):33–50.

Jaynes, E. T. (1983). *Papers on Probability, Statistics, and Statistical Physics*. Dordrecht, Reidel. Rosencrantz, R. D. (Ed.).

Mayhew, B. H. (1984). Baseline models of sociological phenomena. *Journal of Mathematical Sociology*, 9:259–281.

Morris, M. and Kretzschmar, M. (1995). Concurrent partnerships and transmission dynamics in networks. *Social Networks*, 17(3):299–318.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265.

Pattison, P. E. and Robins, G. L. (2002). Neighborhood-based models for social networks. *Sociological Methodology*, 32:301–337.

Reich, L. and Butts, C. T. (2006). Structural properties of power-relevant relations. Presentation to the 26th Sunbelt Network Conference (INSNA).

Snijders, T. A. B. and Stokman, F. N. (1987). Extensions of triad counts to networks with different subsets of points and testing underlying random graph distributions. *Social Networks*, 9:249–275.

Zablocki, B. (1980). *Alienation and Charisma: A Study of Contemporary American Communes*. The Free Press, New York.