Bayesian Meta-Analysis of Social Network Data via Conditional Uniform Graph Quantiles*

Carter T. Butts†

01/12/10

Abstract

Many basic questions in the social network literature center on the distribution of aggregate structural properties within and across populations of networks. Despite this, little work has been done on inference for the properties of graph populations, or on methods for comparing such populations. Here, we attempt to rectify this gap by introducing a family of techniques that combines an existing approach to the identification of structural biases in network data (the use of conditional uniform graph quantiles) with strategies drawn from Bayesian meta-analysis. The resulting methods allow for principled inference regarding the distribution of structural biases within (and comparison across) populations of networks, given data sampled at the network level. Algorithms are provided for carrying out the associated computations, and examples are provided of applications to the analysis of social structure within urban communes and radio communications among emergency personnel.

Keywords: social networks, graph comparison, conditional uniform graphs, graph-level indices, Bayesian meta-analysis, maximum entropy

1 Introduction

The social network literature is replete with concepts for the description of aggregate social structure. Familiar graph-level indices (GLIs) such as density, reciprocity, transitivity, and centralization (in all their many variants) quantify structural features of entire networks, and form an important part of the standard toolkit of classical social network analysis (see, e.g. Wasserman and Faust, 1994, for a canonical treatment). That such aggregate properties can predict the behavior of social processes has been known at least since the experiments of Bavelas (1950), with subsequent work using graph-level properties to predict everything from organizational resilience (Krackhardt and Stern, 1988) to the spread of disease (Morris and Kretzschmar, 1995). By turns, graph-level properties have also been used

*This research was supported ONR award N00014-08-1-1015. The author would like to thank David Krackhardt and Leah Reich for their helpful input.

†Department of Sociology and Institute for Mathematical Behavioral Sciences; University of California, Irvine; SSPA 2145; Irvine, CA 92697-5100; buttsc@uci.edu

as indicators of underlying social processes, as in the famous Davis and Leinhardt studies of triadic structure in small groups (motivated by balance theory; see e.g., Davis, 1970; Davis and Leinhardt, 1972), Baker and Faulker's (1993) examination of centralization in illegal price-fixing networks (motivated by questions of secrecy/efficiency trade-offs), or Bearman et al.'s (2004) work on the cycle structure of adolescent romantic networks (motivated by interest in the partnership formation process). At a deeper level, biases in the distributions of specific graph-level indices have been shown to relate directly to particular forms of relational dependence (Holland and Leinhardt, 1981; Frank and Strauss, 1986; Pattison and Robins, 2002), the modeling of which is a very active area of research (see, e.g. Robins et al., 2005; Hunter and Handcock, 2006; Snijders et al., 2006; Goodreau et al., 2008).

Common to most graph-level studies is the notion (tacit or explicit) that the network or networks under examination are drawn from some larger population of possible structures. In some cases (e.g., studies of schools, firms, or other organizations) we envision a concrete population of existing networks, to which we would like to generalize in some fashion. In other cases, the relevant "population" of networks may reflect the set of possible outcomes for some social process that has been observed in a limited range of circumstances. This poses two basic problems for analysis. First, the distributions of higher-order GLIs are known to be strongly influenced – or even constrained – by low-order GLIs such as size and density that can be expected to vary across network populations for a variety of reasons (Anderson et al., 1999; Butts, 2006; Faust, 2007). Second, to generalize to the population level, we require modeling and/or estimation strategies beyond the single-graph statistics that currently dominate the field. A long-standing answer to the first difficulty has been to replace the raw values of GLIs with the quantiles of those values in baseline distribution of networks conditioned on the low-order GLI values of the observed network. These quantiles can be interpreted as simple measures of structural *bias*, in the sense that they quantify the departure of an observed graph from what would be typical given its low-order properties (Anderson et al., 1999). Where these quantiles are found to be moderate, by turns, baseline factors (in the sense of Mayhew, 1984a;b) can be said to provide an adequate explanation for the higher-order GLI.

The problem of responding to the second difficulty – of reasoning from quantile information from a network sample to a larger population of networks – is the central concern of this paper. Specifically, we frame the problem as one of meta-analysis, in the sense of integrating information from multiple sources (possibly without access to the original data) to facilitate general conclusions. The approach taken here is Bayesian, both for foundational reasons (see, e.g. Robert, 1994) and for reasons of flexibility. We begin by providing a formal review of the notion of conditional uniform graph distributions, as well as a simple Bayesian refinement of the Monte Carlo estimators conventionally used to obtain their associated quantiles. Following this, we proceed to set forth a simple meta-analytic framework for integrating quantile information from multiple networks, using maximum entropy (exponential family) models with minimally informative priors that impose as little structure as possible on resulting inferences. Finally, we conclude with an illustrative application of the approach, first to hierarchical structure in urban communes, and secondly to centralization and cyclicity in emergency radio communication networks.

2 Conditional Uniform Graph Quantiles

As noted above, our approach centers on the use of conditional uniform graph quantiles as a basic unit of comparative measurement, filling a role somewhat like the z-score in a conventional meta-analytic context. Given this, we begin by providing a more formal definition of the quantiles themselves, as well as the distributions that give rise to them. This is followed by a brief section on the Bayesian estimation of quantiles using Monte Carlo methods, a topic which is often glossed over in standard treatments of the subject (e.g. Butts, 2008). The latter leads to a simple modification of common practice that is useful both in its own right, and for the techniques presented in subsequent sections of the paper.

2.1 Conditional Uniform Graph Distributions

Let $\mathbb{G}$ be the (infinite) 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 then defined by the pmf

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

where $\mathbb{I}$ is the standard indicator function. Allowing f to be vector-valued permits Equation 1 to be generalized to the case of multiple statistics in a straightforward manner; note, however, that the pmf may not be well-defined for certain combinations of statistics (specifically, where the induced support is empty). In the text which follows, we will use the term conditional uniform graph (CUG) to refer to distributions of this form, and denote the corresponding pmf by $\mathrm{CUG}(f, \tau)$.

CUG distributions are frequently employed as interpretable baselines against which to assess structural biases; this practice dates at least to Holland and Leinhardt (1970), and continues to be employed in current research (see, e.g. Anderson et al., 1999). Within this paradigm, structural biases are assessed by comparing the quantiles of some graph statistic, f', under a substantively motivated CUG distribution (often by conditioning on order and density, or on the dyad census) to the values observed on empirically obtained networks. This can obviously be interpreted as a null hypothesis test, although quantiles of the observed statistics may also be employed for other purposes.

To describe this process more formally, let g be an observed graph (e.g. arising from some population or social process of interest). Let f be a (possibly vector-valued) conditioning statistic, and let f' be a statistic of substantive interest. Then our interest is in the quantiles $\Pr(f'(G) < f'(g))$, $\Pr(f'(G) = f'(g))$, and $\Pr(f'(G) > f'(g))$, where $G \sim \mathrm{CUG}(f, f(g))$. It is important to note that – due to the discrete nature of $\mathbb{G}$, it is possible for $\Pr(f'(G) = f'(g))$ to be much greater than 0. Thus, all three quantiles are necessary to describe the state of $f'(g)$ with respect to the distribution arising from G.

2.2 Bayesian Monte Carlo Estimation of CUG Quantiles

With the exception of vary rare cases (e.g., reciprocity on the order-conditioned digraphs), CUG quantiles cannot be obtained analytically. Direct enumeration of the CUG support is likewise prohibitive for computational reasons. As such, CUG quantiles are typically estimated via Monte Carlo procedures, a practice which is facilitated by the fact that well-known algorithms exist for exact simulation of draws from many common CUG distributions. Letting $g^{(1)}, \ldots, g^{(n)}$ be a set of draws from $G \sim CUG(f, \tau)$, it is common to approximate the CUG quantiles by the histogram estimators $\Pr\left(f'(G) < f'(g)\right) \approx \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(f'(g^{(i)}) < f'(g)\right)$, $\Pr\left(f'(G) = f'(g)\right) \approx \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(f'(g^{(i)}) = f'(g)\right)$, and $\Pr\left(f'(G) > f'(g)\right) \approx \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(f'(g^{(i)}) > f'(g)\right)$. Although easily implemented, the histogram estimators do a poor job of estimating extreme quantiles, especially when n is small. For instance, a distribution for which the lower quantile is $\Pr\left(f'(G) < f'(g)\right) = 0.001$ requires nearly 700 draws (i.e., $n \approx 692$) in order to have a better than even chance of estimating the true value as greater than 0, and nearly 3,000 in order for this probability to exceed 95%. This lack of precision can be problematic, as conventional null hypothesis testing procedures demand the ability to make fine distinctions regarding exactly such quantities. Although increasing the size of the simulated sample is ultimately necessary to gain precision, it is nonetheless possible to use estimators which are better behaved for small n, at least in the sense of "hedging" an apparently extreme sample quantile against the uncertainty inherent in the sampling process. (This is also important for avoiding spurious 0/1 quantiles in the methods that follow.)

Although the histogram estimators seem natural enough to be inevitable in this application, there exist Bayesian alternatives that are equally simple to implement but which are less prone to pathological behavior. For instance, consider the quantile vector

$$\psi = \left(\Pr\left(f'(G) < f'(g)\right), \Pr\left(f'(G) = f'(g)\right), \Pr\left(f'(G) > f'(g)\right)\right)$$

for some observed graph g and statistic f', with $G \sim CUG(f, \tau)$ as above. Given a sample of independent draws $g^{(1)}, \ldots, g^{(n)}$ from G, let

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

be the count vector for simulated networks with statistic values below, equal to, and greater than the observed graph. Since $g^{(i)}$ are iid, it follows that $\mathbf{x} \sim \text{Multinomial}(n, \psi)$ with n known. To infer ψ from $\mathbf{x}$, we must posit some prior distribution $p(\psi)$; for a general purpose procedure, we wish for our inferences to be controlled primarily by $\mathbf{x}$ and not to depend on any particular properties of f' or g, and hence seek a minimally informative prior. Given that we do know that extreme quantiles are not improbable in a research context, a uniform prior on the three dimensional unit simplex is not appropriate – a better choice in this case is the corresponding Jeffreys prior, which is $\text{Dirichlet}(1/2, 1/2, 1/2)$ for this particular model. As the Dirichlet distribution is the conjugate prior for the multinomial, we then have the simple result that $\psi\,|\mathbf{x} \sim \text{Dirichlet}(x_1 + 1/2, x_2 + 1/2, x_3 + 1/2)$. Since the ψ in this case are probabilities, the natural estimator to employ is $\hat{\psi} = \mathbf{E}\left(\psi|\mathbf{x}\right)$, which is given by

$$\hat{\psi} = \left(\frac{x_1 + 1/2}{3/2 + \sum_{i=1}^{3} x_i}, \frac{x_2 + 1/2}{3/2 + \sum_{i=1}^{3} x_i}, \frac{x_3 + 1/2}{3/2 + \sum_{i=1}^{3} x_i}\right) \tag{2}$$

$$= \left(\frac{x_1 + 1/2}{n + 3/2}, \frac{x_2 + 1/2}{n + 3/2}, \frac{x_3 + 1/2}{n + 3/2}\right). \tag{3}$$

This estimator is no more difficult to calculate than the usual histogram estimator, and indeed the two are nearly identical for moderate ψ and large n. However, Equation 3 serves to shrink extreme observed quantiles away from the limiting values of 0 and 1, reflecting a correction for expected small-sample effects. In the important (and common) case in which some x_i is observed to be zero for a particular sample, this correction tells us how extreme we can realistically expect the associated ψ_i to be, given the number of draws from G that were employed. The fact that n is typically reported along with the observed quantile frequencies (i.e., the histogram estimators) also means that we can calculate $\hat{\psi}$ *ex post facto* from published studies. This is particularly useful in the meta-analytic context, to which we now turn.

3 Bayesian Meta-Analysis of CUG Quantiles

While CUG quantiles can be highly informative when dealing with single graphs, it is less obvious how such statistics may be combined across multiple graphs to draw conclusions about structural biases at the population level. Such inference is obviously a form of *meta-analysis*, in that it seeks to integrate results from multiple sub-populations (possibly arising from distinct studies) to provide evidence regarding aggregate behavior. In the text which follows, we will assume that we have observed a collection $g_1, \ldots, g_N$ of graphs from $\mathbb{G}$, and have obtained CUG quantiles for these graphs under some pair of conditioning and substantive statistics. Our concern is with the joint modeling of these quantiles, with an eye to drawing population-level conclusions. It is to that problem that we now turn.

3.1 Single Population Case

We begin with the case in which the quantiles for each graph are assumed to be drawn from a single population. A natural model for the observed quantiles in this case is the Dirichlet distribution, conditional on parameter vector θ. Let $\mathbf{Y} \in (0,1)^{N \times 3}$ be the matrix of observed quantiles, with columns corresponding to $\Pr\left(f'(G) < f'\left(g_i\right)\right)$, $\Pr\left(f'(G) = f'\left(g_i\right)\right)$, and $\Pr\left(f'(G) > f'\left(g_i\right)\right)$ (respectively). The likelihood of a single observation, $\mathbf{Y}_i$, is then given by

$$p(\mathbf{Y}_i|\theta) = \frac{\Gamma(\sum_{i=1}^3 \theta_i)}{\prod_{i=1}^3 \Gamma(\theta_i)} \prod_{j=1}^3 Y_{ij}^{\theta_i - 1}. \tag{4}$$

For the development which follows, it will be more useful to represent $p(\mathbf{Y}_i|\theta)$ in exponential family form. Specifically, we may re-write the above as

$$= \exp\left[(\theta - 1)^T \ln \mathbf{Y}_i + \ln \Gamma\left(\sum_{j=1}^3 \theta_j\right) - \sum_{j=1}^3 \ln \Gamma\left(\theta_j\right)\right]. \tag{5}$$

In the present context, we assume that $g_1, \ldots, g_N$ (and hence their CUG quantiles) are drawn independently from the population of interest. This implies that the joint likelihood of $\mathbf{Y}$ is the product of the row likelihoods, or

$$p(\mathbf{Y}|\theta) = \prod_{i=1}^{N} p(\mathbf{Y}_i|\theta) \tag{6}$$

$$= \exp\left[(\theta - 1)^T \left(\sum_{i=1}^{N} \ln \mathbf{Y}_i\right) + N\left(\ln\Gamma\left(\sum_{j=1}^{3}\theta_j\right) - \sum_{j=1}^{3}\ln\Gamma\left(\theta_j\right)\right)\right]. \tag{7}$$

To conduct Bayesian inference for the quantile model, it is necessary to place a prior distribution on θ. While many possible priors exist, one reasonable choice is the conjugate exponential family prior,

$$p(\theta|\phi,\gamma) \propto \exp\left[(\theta - 1)^T \phi + \gamma\left(\ln\Gamma\left(\sum_{j=1}^{3}\theta_j\right) - \sum_{j=1}^{3}\ln\Gamma\left(\theta_j\right)\right)\right], \tag{8}$$

with support on $(0,\infty)^3$. In addition to being fairly simple to work with, this prior also has the advantage of having hyperparameters which are interpretable in terms of hypothetical prior data (or "pseudo-data"). Specifically, imagine that our prior information is derived from a data set of size N^* with quantiles whose respective products are equal to $\mathbf{y}^* = (\prod_{i=1}^{N^*} Y_{i1}^*, \prod_{i=1}^{N^*} Y_{i2}^*, \prod_{i=1}^{N^*} Y_{i3}^*)$. Then $\gamma = N^*$, and $\phi = \ln \mathbf{y}^*$. While our prior information may not in fact be derived from a such a source, the above may nevertheless be useful in setting reasonable subjective priors. Alternately, we can see from the above that an uninformative prior is obtained in the limit as $\gamma \to 0$. Although this prior is improper, proper minimally informative priors can be obtained by setting $\gamma \ll 1$, $\phi_i = -\gamma \ln 3$.

To form the posterior, we now simply invoke Bayes's Theorem, substituting from Equations 7 and 8. This gives us

$$p(\theta|\mathbf{Y},\phi,\gamma) \propto \exp\left[(\theta - 1)^T \left(\phi + \sum_{i=1}^{N} \ln \mathbf{Y}_i\right) + (\gamma + N)\left(\ln\Gamma\left(\sum_{j=1}^{3}\theta_j\right) - \sum_{j=1}^{3}\ln\Gamma\left(\theta_j\right)\right)\right], \tag{9}$$

which is immediately recognizable as an exponential family of the same form as the prior (thus confirming the assertion of conjugacy). While direct simulation of θ is not possible, approximate sampling can easily be performed using a Metropolis algorithm. For this purpose, we note that the posterior ratio of parameter vectors θ' and θ' is given by

$$\frac{p(\theta'|\mathbf{Y},\phi,\gamma)}{p(\theta|\mathbf{Y},\phi,\gamma)} = \exp\left[(\theta' - \theta)^T \left(\phi + \sum_{i=1}^{N} \ln \mathbf{Y}_i\right)\right.$$
$$\left. + (\gamma + N)\left(\ln\Gamma\left(\sum_{j=1}^{3}\theta'_j\right) - \ln\Gamma\left(\sum_{j=1}^{3}\theta_j\right) + \sum_{j=1}^{3}\left(\ln\Gamma\left(\theta_j\right) - \ln\Gamma\left(\theta'_j\right)\right)\right)\right]. \tag{10}$$

Simulation proceeds by successively drawing candidate θ vectors, switching states with probability $\min\left\{1, \frac{p(\theta'|\mathbf{Y},\phi,\gamma)}{p(\theta|\mathbf{Y},\phi,\gamma)}\right\}$. One such approach is shown in Algorithm 1; note that the variance of the jumping rule, σ^2, may need to be adjusted to ensure that the chain mixes well. Experience has suggested that values in the neighborhood of 0.1 often prove effective, although this is somewhat dependent upon sample size. Likewise, the starting position of the chain (here drawn from a standard lognormal distribution) may be tuned to produce more rapid convergence.

Algorithm 1 A Metropolis Algorithm for the Homogeneous Population Model

for $j \in 1, \ldots, 3$ **do**
    Draw $\ln \theta_j^{(0)}$ from $\mathcal{N}(0, 1)$
end for
Let $i := 1$
repeat
    Let $\theta^{(i)} := \theta^{(i-1)}$
    for $j \in 1, \ldots, 3$ **do**
        Draw x from $\mathcal{N}(0, \sigma^2)$
        Let $\theta_j^{(i)} := \exp\left(\ln \theta_j^{(i)} + x\right)$
    end for
    Draw u from $U(0, 1)$
    if $u > \frac{p(\theta^{(i)}|\mathbf{Y},\phi,\gamma)}{p(\theta^{(i-1)}|\mathbf{Y},\phi,\gamma)}$ **then**
        Let $\theta^{(i)} := \theta^{(i-1)}$
    end if
    Let $i := i + 1$
until $\theta^{(\cdot)} \sim p(\theta|\mathbf{Y}, \phi, \gamma)$
return $\theta^{(\cdot)}$

3.2 Mixed Population Case

The above model is adequate to draw inferences on a reasonably homogeneous population of graphs. In some cases, however, the presence of (possibly unobserved) heterogeneity across structures makes the use of a single population model infeasible.[1] Insofar as the heterogeneity within our data arises from the presence of multiple subpopulations with distinct quantile distributions, this property can be captured by means of a discrete mixture over single population models. In particular, let us assume that each network in our sample has been drawn from one of K distinct subpopulations, with the vector $\eta \in (1, \ldots, K)^N$ indexing population membership. Then the likelihood of any single observation, $\mathbf{Y}_i$, may be written as

$$p(\mathbf{Y}_i|\mathbf{\Theta}, \eta) = \exp\left[(\mathbf{\Theta}_{\eta_i} - 1)^T \ln \mathbf{Y}_i + \ln \Gamma\left(\sum_{j=1}^{3} \mathbf{\Theta}_{\eta_i j}\right) - \sum_{j=1}^{3} \ln \Gamma\left(\mathbf{\Theta}_{\eta_i j}\right)\right], \tag{11}$$

where $\mathbf{\Theta} \in (0, \infty)^{K \times 3}$ is the K-population generalization of the parameter vector θ. Since we continue to assume that observations are independent given $\mathbf{\Theta}$ and η, the joint likelihood

[1] In point of fact, the single population model does allow for certain types of heterogeneity; certain features (such as multiple interior modes), however, cannot be captured without additional structure.

of $\mathbf{Y}$ is just the product of the individual observation likelihoods. For the analysis which follows, it is helpful to group observations by subpopulation; to that end, let the function $c_i(\eta) = \sum_{j=1}^{N} \mathbb{I}(\eta_j = i)$ provide the count of observations within each subgroup. Using this, we may write the joint likelihood as follows:

$$p(\mathbf{Y}|\mathbf{\Theta}, \eta) \propto \exp\left[\sum_{i=1}^{K}\left((\mathbf{\Theta}_i - 1)^T\left(\sum_{\{j:\eta_j=i\}} \ln \mathbf{Y}_j\right)\right) + \sum_{i=1}^{K} c_i(\eta)\left(\ln\Gamma\left(\sum_{j=1}^{3}\Theta_{ij}\right) - \sum_{j=1}^{3}\ln\Gamma(\Theta_{ij})\right)\right]. \tag{12}$$

Note that Equation 12 is equivalently the product of the likelihoods from K homogeneous population models, in keeping with the underlying substantive interpretation. Similarly, we may posit a conjugate exponential family prior for $\mathbf{\Theta}$ whose form is a product of homogeneous cases:

$$p(\mathbf{\Theta}|\mathbf{\Phi}, \gamma) \propto \exp\left[\sum_{i=1}^{K}(\mathbf{\Theta}_i - 1)^T\mathbf{\Phi}_i + \sum_{i=1}^{K}\gamma_i\left(\ln\Gamma\left(\sum_{j=1}^{3}\Theta_{ij}\right) - \sum_{j=1}^{3}\ln\Gamma(\Theta_{ij})\right)\right] \times \mathbb{I}(\Theta_{11} < \ldots < \Theta_{K1}). \tag{13}$$

The constraint that $\Theta_{11} < \ldots < \Theta_{K1}$ is introduced in order to arrive at an identifiable model, but the subpopulation distributions are otherwise held to be a priori independent. We similarly treat $p(\eta) \propto 1$, subject to the constraint that $\prod_{i=1}^{K} c_i(\eta) > 0$ (i.e., that every subpopulation has at least one member). In practice, the $\mathbf{\Theta}$ distributions for each cluster will be linked via their posterior dependency on η, and vice versa.

Given the likelihood and prior structures, we may now determine the joint posterior of $\mathbf{\Theta}$ and η given the data and hyperparameters. From Equations 12 and 13, Bayes's Theorem gives us

$$p(\mathbf{\Theta}, \eta|\mathbf{Y}, \mathbf{\Phi}, \gamma) \propto \exp\left[\sum_{i=1}^{K}\left((\mathbf{\Theta}_i - 1)^T\left(\mathbf{\Phi}_i + \sum_{\{j:\eta_j=i\}}\mathbf{Y}_j\right)\right) + \sum_{i=1}^{K}(\gamma_i + c_i(\eta))\left(\ln\Gamma\left(\sum_{j=1}^{3}\Theta_{ij}\right) - \sum_{j=1}^{3}\ln\Gamma(\Theta_{ij})\right)\right] \times \mathbb{I}\left(\Theta_{11} < \ldots < \Theta_{K1}, \prod_{i=1}^{K} c_i(\eta) > 0\right). \tag{14}$$

As foreshadowed above, the form of the posterior is greatly simplified when considering only conditional distributions. For η, for instance, we have

$$p(\eta|\mathbf{\Theta}, \mathbf{Y}) \propto \exp\left[\sum_{i=1}^{K}\left((\mathbf{\Theta}_i - 1)^T\left(\sum_{\{j:\eta_j=i\}}\mathbf{Y}_j\right)\right)\right]\mathbb{I}\left(\prod_{i=1}^{K} c_i(\eta) > 0\right). \tag{15}$$

Similarly, for the $\mathbf{\Theta}_i$ associated with any given cluster, the full conditional distribution is given by

$$p(\mathbf{\Theta}_i|\mathbf{Y}, \eta, \mathbf{\Phi}, \gamma) \propto \exp\left[(\mathbf{\Theta}_i - 1)^T \left(\mathbf{\Phi}_i + \sum_{\{j:\eta_j=i\}} \mathbf{Y}_j\right) + (\gamma_i + c_i(\eta)) \left(\ln\Gamma\left(\sum_{j=1}^3 \Theta_{ij}\right) - \sum_{j=1}^3 \ln\Gamma(\Theta_{ij})\right)\right], \mathbb{I}(\Theta_{i1} < \ldots < \Theta_{i1}) \tag{16}$$

which is quite plainly equivalent to the posterior obtained from the homogeneous population model on all $\mathbf{Y}_j$ such that $\eta_j = i$ (neglecting the support constraint). The posterior marginal for $\mathbf{\Theta}_i$ will thus be a mixture over the homogeneous posteriors induced by various cluster assignments, as is typical for a model of this kind.

As the form of Equations 15 and 16 suggest, a natural way to sample from the distribution of Equation 14 is via alternately sampling from $\mathbf{\Theta}|\eta$ and $\eta|\mathbf{\Theta}$, each of these being performed via Metropolis steps. (See Gilks et al. (1996) for a discussion of such mixed sampling procedures.) This approach is illustrated in Algorithm 2. As with Algorithm 1, tuning may be necessary in order to obtain good results; this is especially true when attempting to fit more mixture components than are supported by the data, as this will tend to produce slow convergence in some cases. By turns, this means that slow convergence may imply a poorly specified model – this is a possibility which should be investigated whenever computational difficulties arise.

4 Examples

The methods introduced in the previous section can be used to answer a wide range of questions regarding the properties of network structure within and across populations of networks. Here, we illustrate their use with two examples. First, we consider the problem of estimating the prevalence of particular structural biases in a population of organizational networks, given a representative sample of organizations. Second, we demonstrate the use of CUG meta-analysis to compare structural biases across two types of groups, based on a small sample of each type. Although these examples do not exhaust the possible applications of the meta-analytic approach, it is hoped that they will serve to illustrate its promise for addressing questions of practical scientific interest.

4.1 Hierarchy in Urban Commune Networks

Although power relations have been widely studied in formal organizations (see, e.g. Pfeffer and Salancik, 1978; Brass, 1984; Krackhardt, 1990), their role in informal organizations is less well-understood. This is in no small part because of the relative dearth of systematic information on informal organizations, which tend by nature to be heterogeneous and difficult to study. One partial exception to this rule is the urban commune, a form of organization which has been subjected to at least one attempt at systematic population sampling. In

Algorithm 2 An MCMC Algorithm for the Mixed Population Model

for $i \in 1, \ldots, K$ **do**
    for $j \in 1, \ldots, 3$ **do**
        Draw x from $\mathcal{N}(0, 1)$
        if $i > 1, j = 1$ **then**
            Let $\boldsymbol{\Theta}_{ij}^{(0)} := \boldsymbol{\Theta}_{(i-1)j}^{(0)} + x$
        else
            Let $\boldsymbol{\Theta}_{ij}^{(0)} := \exp(x)$
        end if
    end for
end for
repeat
    for $i \in 1, \ldots, N$ **do**
        Draw $\eta_i^{(0)}$ from Multinom$(K, (1/K, \ldots, 1/K))$
    end for
until $\prod_{j=1}^{K} c_j \left(\eta^{(0)}\right) > 0$
Let $i := 1$
repeat
    for $j \in 1, \ldots, K$ **do**
        Let $\boldsymbol{\Theta}_j^{(i)} := \boldsymbol{\Theta}_j^{(i-1)}$
        for $k \in 1, \ldots, 3$ **do**
            Draw x from $\mathcal{N}(0, \sigma^2)$
            Let $\Theta_{jk}^{(i)} := \exp\left(\ln \Theta_{jk}^{(i)} + x\right)$
        end for
        Draw u from $U(0, 1)$
        if $u > \frac{p(\boldsymbol{\Theta}_j^{(i)} | \mathbf{Y}, \eta, \boldsymbol{\Phi}, \gamma)}{p(\boldsymbol{\Theta}_j^{(i-1)} | \mathbf{Y}, \eta, \boldsymbol{\Phi}, \gamma)}$ **then**
            Let $\boldsymbol{\Theta}_j^{(i)} := \boldsymbol{\Theta}_j^{(i-1)}$
        end if
    end for
    for $j \in 1, \ldots, N$ **do**
        Draw u from $U(0, 1)$
        if $u < p_c$ **then**
            Draw η_j from Multinom$(K, (1/K, \ldots, 1/K))$
        end if
    end for
    Draw u from $U(0, 1)$
    if $u > \frac{p(\eta^{(i)} | \mathbf{Y}, \boldsymbol{\Theta})}{p(\eta^{(i-1)} | \mathbf{Y}, \boldsymbol{\Theta})}$ **then**
        Let $\eta^{(i)} := \eta^{(i-1)}$
    end if
    Let $i := i + 1$
until $\boldsymbol{\Theta}^{(\cdot)}, \eta^{(\cdot)} \sim p(\boldsymbol{\Theta}, \eta | \mathbf{Y}, \boldsymbol{\Phi}, \gamma)$
return $\boldsymbol{\Theta}^{(\cdot)}, \eta^{(\cdot)}$

particular, the Urban Communes Data Set (Zablocki, 1980) consists of a multi-wave study of urban communes in the United States. The data employed here are drawn from the network component of the first wave (conducted in 1974), consisting of information regarding 61 communes ranging in size from 4 to 26 members. Using this base data, Reich and Butts (2006) construct ten relations which they identify as being potentially indicative of the exercise of informal power within each commune. These relations are referred to respectively as "power" (claims that ego exercises power over alter), "authfictkin" (claims that ego occupies a superordinate fictive kin role vis a vis alter), "depend" (claims that alter depends on ego more than vice versa), "exploit" (claims that ego exploits alter), "attrinf" (ego is regarded by alter as being influential), "attmag" (ego is regarded by alter as "sexy" and/or "charismatic"), "attrdom" (ego is regarded by alter as dominant), "attrholy" (ego is regarded by alter as being "holy"), "attrweak" (alter is regarded by ego as being "passive" or "dependent"), "signif" (alter regards ego as a "significant person" in his/her life). Following Krackhardt (1994), Reich and Butts argue that the presence of a systematic tendency towards hierarchical organization within these networks – net of baseline factors – is *prima facie* evidence for their use as conduits for the exercise of informal power.

Hierarchy can be measured in a number of ways, two of which are employed here. First, we may think of hierarchy in terms of local asymmetry, i.e., the additive inverse of the edgewise reciprocity score (equivalently, the probability that v does not send a tie to v', given a $v' \to v$ edge). Second, we may think of hierarchy in terms of path asymmetry, as with Krackhardt's (1994) measure (which corresponds to the local hierarchy of the reachability network). Intuitively, we may expect that if v exercises power over v', then v' is unlikely to exercise power over v via the same relation. Thus, if a given relation widely serves as an indicator for the informal exercise of power, we expect it to be less reciprocal (more locally hierarchical) than would be typical for networks of the same size and density. By the same token, the *indirect* use of power via longer paths (e.g., $v \to v' \to v''$) should also be asymmetrically organized, leading to networks with higher level of Krackhardt hierarchy than would be expected given their size, density, and degree of local asymmetry (all of which are implicitly given by the dyad census).

While direct inspection of CUG quantiles for the above statistics can tell us much about the behavior of our sampled networks, our interest lies in what we can conclude about the larger population whence the sample was drawn (namely, the population of all US urban communes in operation during the data collection period). To infer the CUG quantile distributions for the larger population, we use the method of Algorithm 1 to draw from the posterior distribution of the single-population model. Our data inputs in this case are the estimated CUG quantiles for the sampled networks (estimated using the method of Equation 3 with $n = 100,000$), and the prior parameters ϕ and γ (chosen to give a prior weight of one observation centered at $(1/3, 1/3, 1/3)$). A total of 5,000 draws were taken using the MCMC algorithm in each case (thinned from a sample of 2,500,000 with 7,500 burn-in iterations), with convergence verified using a combination of Geweke's z_G and the Raftery-Lewis diagnostics (Gelman et al., 1995). These draws, in turn, were used to generate a sample from the corresponding posterior predictive distribution of CUG quantiles, which reflects our estimate of the distribution of CUG quantiles at the population level. The posterior predictive reciprocity and hierarchy quantile distributions are respectively provided in Figures 1 and 2. Each ternary diagram within the figure contains a set of contour lines indicating the posterior

Relation	Reciprocity | Density		Hierarchy | Dyad Census	
	Pr(sig.low)	Pr(sig.high)	Pr(sig.low)	Pr(sig.high)
power	0.271	0.006	0.051	0.254
authfictkin	0.094	0.061	0.046	0.144
depend	0.111	0.068	0.042	0.058
exploit	0.019	0.010	0.029	0.049
attrinf	0.041	0.018	0.067	0.173
attrmag	0.033	0.026	0.073	0.119
attrdom	0.033	0.008	0.019	0.096
attrholy	0.021	0.009	0.086	0.204
attrweak	0.049	0.010	0.067	0.160
signif	0.056	0.111	0.094	0.127

Table 1: Posterior predictive probabilities for one-tailed CUG significance at $p < 0.05$ by UCDS relation, reciprocity given density and hierarchy given the dyad census

predictive pdf, along with the original data (red points). (Note that the data points are not jittered; high-density regions containing apparently few data points typically have multiple observations approximately superimposed.) As the high mass near the top of each triangle suggests, reciprocity scores in these networks tended to be highly typical of the baseline distributions, in large part due to size and density constraints. Somewhat similar results are observed for the hierarchy measure, with a somewhat higher degree of variability apparent.

In addition to simple inspection of the posterior predictive distributions, the flexibility of the Bayesian framework allows us to easily answer more complex questions (even seemingly frequentist ones). For instance, it is natural to ask in the present context for the probability that an arbitrarily selected member of the population of communes will have a significantly high (or low) score on a particular structural index, where "significance" is here interpreted as usual for a one-tailed CUG test with a threshold of $p < 0.05$. Although significance levels are not a Bayesian concept, we can nevertheless answer this question using out approach, by determining the probability that a randomly chosen population member will have $\Pr(f(G) \leq (g)) < 0.05$ (respectively $\Pr(f(G) \leq (g)) < 0.05$). These probabilities are given for each relation in Table 1. The table serves to clarify the intuition of figures: the majority of communes are predicted to have unremarkable levels of asymmetry at both the local and the global level. On the other hand, we do see some noteworthy departures from the baseline significance rate. For instance, over 25% of communes are predicted to have perceived power relations that are both locally and globally hierarchical (5 times the base rate). Rates of significantly high global hierarchical organization are likewise predicted to be well above baseline for all relations except for dependence and exploitation (the former of which tends to be locally asymmetric at above-chance rates). While we do not find strong evidence of power exercise in all networks (or all relations), we do find evidence that the associated structural pattern appears in a consequential minority of organizations. Going forward, this suggests that there may be value in looking for covariate information that can help predict which communes fall into this latter category.

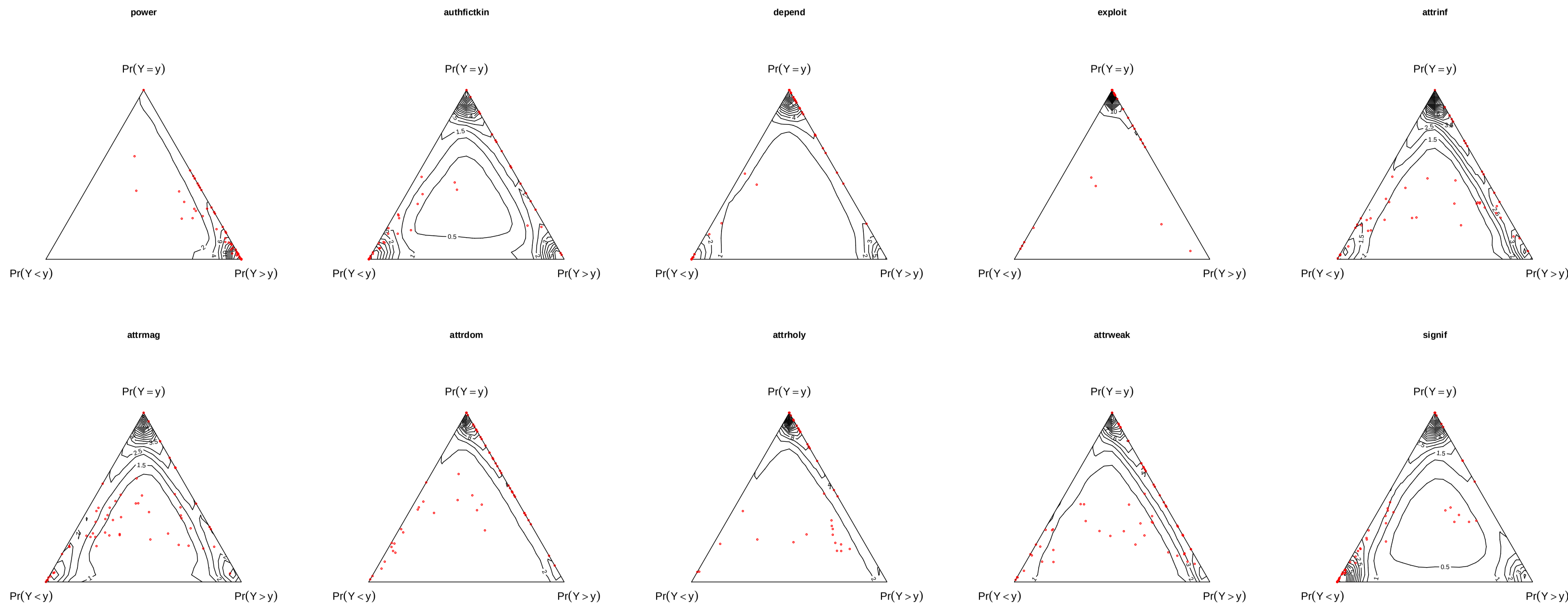

Figure 1: Posterior predictive distributions for reciprocity CUG quantiles (versus density) by relation, UCDS; observed quantiles are shown in red

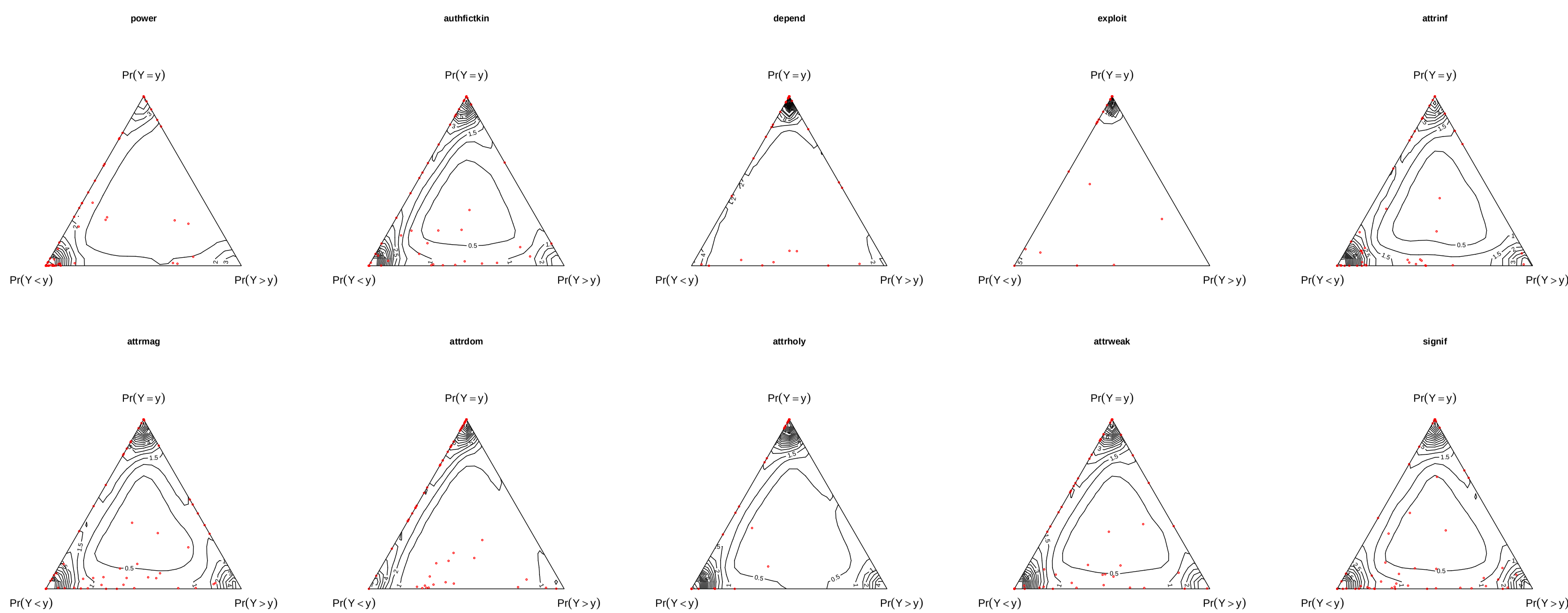

Figure 2: Posterior predictive distributions for hierarchy CUG quantiles (versus dyad census) by relation, UCDS; observed quantiles are shown in red

4.2 Centralization and Cyclicity in Responder Radio Networks

In recent work Butts et al. (2007); Petrescu-Prahova and Butts (2008) examine the properties of 17 interpersonal radio communication networks arising from responders to the 2001 World Trade Center Disaster. Each network corresponds to a single organizational unit, whose members employed hand-held radios sharing a single common frequency. Although all units were affiliated at the time of the disaster with the Port Authority of New York and New Jersey (from whom the information was collected), they differ substantially in other respects. In particular, 9 of the 17 units are identified by Butts et al. (2007) as "specialist" response organizations – i.e., organizations for which responding to emergencies is a core part of their mission – while 8 are "non-specialist" response organizations (organizations for whom emergency response is not an expected task). Drawing on arguments from organizational theory, Butts et al. (2007) put forward a number of competing hypotheses regarding the structure of the communication networks for the 17 units, including the relationship between specialization and conflicting demands for coordination and efficiency. If the emergency context dominates prior preparation, then all groups are expected to show similar structural biases. These biases should be in the direction of increased centralization and reduced clustering (e.g., cyclicity) if efficiency pressures predominate, or in the direction of decreasing centralization and increased clustering if structure is dominated by the need to coordinate the performance of complex tasks. If, on the other hand, existing organizational arrangements prove more important than the immediate context, we expect to see biases that differ based on specialization (with specialists being more prone to centralization and showing less clustering, and non-specialists displaying the opposite pattern).

To investigate this question using the present framework, we start by drawing CUG quantiles (MC sample size 2,000) for each network on total degree centralization, betweenness centralization, and cyclicity. We then take 5,000 posterior draws for each GLI/specialization combination using the same procedure as for the commune analysis above. The posterior predictive distributions associated with the posterior samples are shown in Figure 3. Differences between these distributions and those of the commune networks are immediately apparent: most notably, the equality probability (mostly a function of size and density) is very small in the majority of cases. In general, most mass is concentrated in the lower left corner of each triangle, indicating that the observed graphs tend to be higher than average on all three measures. The qualitative patterns appear similar for specialist and non-specialist networks, suggesting a context-dominant explanation. We need not rely on a visual inspection, however, but can (as before) use the posterior predictive distribution to obtain direct answers to our questions of interest.

Table 2 shows median and central 50% interval for the posterior predictives of $Pr(f(G) \leq f(g))$ and $Pr(f(G) \geq f(g))$ (typically interpreted as one-tailed CUG p-values) for each GLI/specialization status combination. As the table indicates, sizable minorities (25%-50%) of both specialist and non-specialist networks produced under similar conditions (i.e., drawn from an equivalent population) would be expected to have significantly high levels of degree centralization, betweenness centralization, and cyclicity as conventionally assessed. At the same time, we see little systematic difference between specialist and non-specialist predictions (in the sense that a randomly chosen specialist network has nearly even odds of being higher (or lower) on a given quantile than a a randomly chosen non-specialist network).

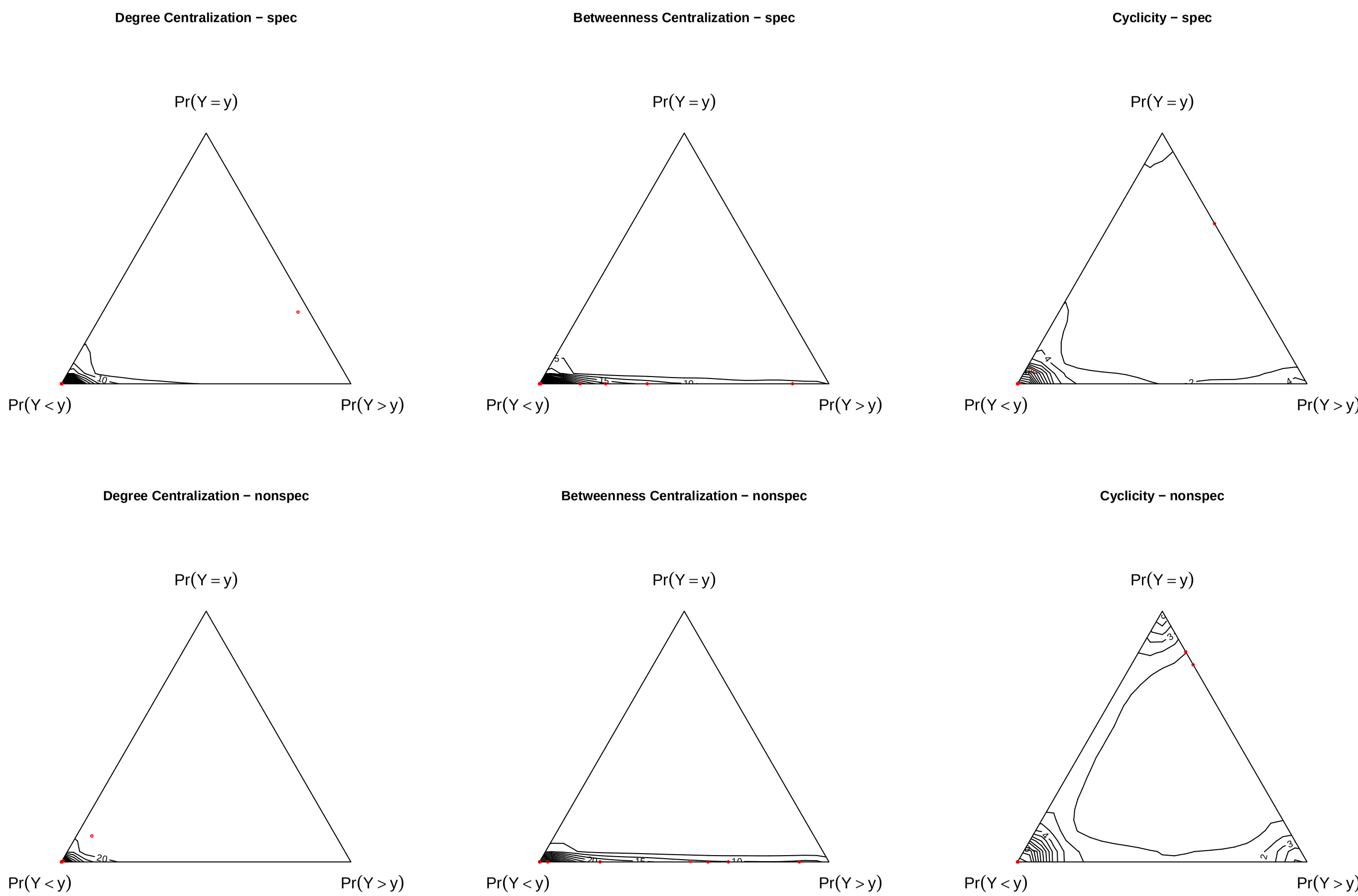

Figure 3: Posterior predictive distributions for degree centralization, betweenness centralization, and cyclicity CUG quantiles (versus the dyad census) by specialist status, PAPD networks; observed quantiles are shown in red

The lack of difference between specialist and non-specialist groups is consistent with the "context predominates" hypothesis, but the combination of centralization and cyclicity is not in keeping with standard expectations. The answer to the puzzle in this case (as noted in the above-referenced papers) is that the cycle structure results disproportionately from dense interactions among a relatively small group of "hubs." The radio networks thus show a hybrid structure which is both highly centralized and which also contains regions of dense interaction, reflecting a possible compromise between the need to minimize communication overall and the need for multi-way conversation to resolve complex problems.

5 Conclusion

In this paper, we have outlined a strategy for population-level inference on social network data, via Bayesian meta-analysis of CUG quantiles. Like standard meta-analytic approaches, the methods discussed here can be applied to published accounts from multiple sources, where complete network data is unavailable (so long as the relevant quantile information – often used for significance testing – is provided). Given a sample of networks (or quantiles from such networks), it is then possible to infer the properties of the larger population from which those networks were drawn, to predict properties of future observations from said population, etc. The use of a Bayesian strategy facilitates the development of a flexible and extensible family of models that can incorporate heterogeneity, prior knowledge, or even hierarchical structure as needed for the problem at hand. The utility of the specific scheme used here was illustrated via two simple applications, both of which showed how information from a network sample could be used to ask more general questions about organizations and the social processes that underlie them. It is hoped that the development of these and other methods will encourage increased use of comparative, multi-network study designs, thereby allowing us to draw stronger conclusions regarding the aggregate properties of social structure.

6 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.

Baker, W. E. and Faulkner, R. R. (1993). The social organization of conspiracy: Illegal networks in the heavy equipment industry. *American Sociological Review*, 58:837–860.

Bavelas, A. (1950). Communication patterns in task oriented groups. *Journal of the Acoustical Society of America*, 22:271–282.

Bearman, P. S., Moody, J., and Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology*, 110:44–91.

Brass, D. J. (1984). Being in the right place: A structural analysis of individual influence in an organization. *Administrative Science Quarterly*, 29:519–529.

GLI		Specialist			Non-Specialist			Pr(Spec>Non-Spec)
		25%	50%	75%	25%	50%	75%	
Degree	$Pr(f(G) \leq f(g))$	0.863	0.992	1.000	0.941	0.995	1.000	0.467
Centralization	$Pr(f(G) \geq f(g))$	0.009	0.100	0.425	0.007	0.049	0.179	0.586
Betweenness	$Pr(f(G) \leq f(g))$	0.762	0.961	0.998	0.537	0.850	0.978	0.631
Centralization	$Pr(f(G) \geq f(g))$	0.020	0.141	0.430	0.078	0.293	0.648	0.383
Cyclicity	$Pr(f(G) \leq f(g))$	0.739	0.984	1.000	0.545	0.975	1.000	0.529
	$Pr(f(G) \geq f(g))$	0.024	0.237	0.735	0.048	0.502	0.954	0.407

Table 2: Posterior predictives (median and IQR) for specialist versus non-specialist PAPD responder networks on three GLIs

Butts, C. T. (2006). Exact bounds for degree centralization. *Social Networks*, 28(4):283–296.

Butts, C. T. (2008). Social networks: A methodological introduction. *Asian Journal of Social Psychology*, 11(1):13–41.

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.

Davis, J. A. and Leinhardt, S. (1972). The structure of positive interpersonal relations in small groups. In *Sociological Theories in Progress*, volume 2, pages 218–251. Houghton Mifflin, Boston.

Faust, K. (2007). Very local structure in social networks. *Sociological Methodology*, 37(1):209–256.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81:832–842.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall, London.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Goodreau, S. M., Kitts, J. A., and Morris, M. (2008). Birds of a feather, or friend of a friend?: Using exponential random graph models to investigate adolescent social networks. *Demography*, 46(1):103–125.

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

Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15:565–583.

Krackhardt, D. (1990). Assessing the political landscape: Structure, cognition, and power in organizations. *Administrative Science Quarterly*, 35:242–260.

Krackhardt, D. (1994). Graph theoretical dimensions of informal organizations. In Carley, K. M. and Prietula, M. J., editors, *Computational Organizational Theory*, pages 88–111. Lawrence Erlbaum Associates, Hillsdale, NJ.

Krackhardt, D. and Stern, R. N. (1988). Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly*, 51:123–140.

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

Mayhew, B. H. (1984b). Chance and necessity in sociological theory. *Journal of Mathematical Sociology*, 9:305–339.

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

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

Petrescu-Prahova, M. and Butts, C. T. (2008). Emergent coordinators in the world trade center disaster. *International Journal of Mass Emergencies and Disasters*, 28(3):133–168.

Pfeffer, J. and Salancik, G. R. (1978). *External Control of Organizations.* Harper and Row, New York.

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

Robert, C. P. (1994). *The Bayesian Choice: a Decision-theoretic Motivation.* Springer, New York.

Robins, G., Pattison, P., and Woolcock, J. (2005). Small and other worlds: Network structures from local processes. *American Journal of Sociology*, 110(4):894–936.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications.* Cambridge University Press, Cambridge.

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