# UCLA
## Research Reports

**Title**
Bayesian modeling and uncertainty quantificationfor descriptive social networks

**Permalink**
https://escholarship.org/uc/item/4s84j9g5

**Authors**
Nemmers, Thomas
Narayan, Anjana
Banerjee, Sudipto

**Publication Date**
2019

**Supplemental Material**
https://escholarship.org/uc/item/4s84j9g5#supplemental

**Data Availability**
The data associated with this publication are in the supplemental files.

Peer reviewed

# Bayesian modeling and uncertainty quantification for descriptive social networks*

Thomas Nemmers, Anjana Narayan, and Sudipto Banerjee†

This article presents a simple and easily implementable Bayesian approach to model and quantify uncertainty in small descriptive social networks. While statistical methods for analyzing networks have seen burgeoning activity over the last decade or so, ranging from social sciences to genetics, such methods usually involve sophisticated stochastic models whose estimation requires substantial structure and information in the networks. At the other end of the analytic spectrum, there are purely descriptive methods based upon quantities and axioms in computational graph theory. In social networks, popular descriptive measures include, but are not limited to, the so called Krackhardt's axioms. Another approach, recently gaining attention, is the use of PageRank algorithms. While these descriptive approaches provide insight into networks with limited information, including small networks, there is, as yet, little research detailing a statistical approach for small networks. This article aims to contribute at the interface of Bayesian statistical inference and social network analysis by offering practicing social scientists a relatively straightforward Bayesian approach to account for uncertainty while conducting descriptive social network analysis. The emphasis is on computational feasibility and easy implementation using existing R packages, such as `sna` and `rjags`, that are available from the Comprehensive R Archive Network (https://cran.r-project.org/). We analyze a network comprising 18 websites from the US and UK to discern transnational identities, previously analyzed using descriptive graph theory with no uncertainty quantification, using fully Bayesian model-based inference.

Keywords and phrases: Bayesian modeling, Directed graphs, Krackhardt's axioms, PageRank algorithms, Social network analysis, Uncertainty quantification.

## 1. INTRODUCTION

Social network analysis (SNA), sometimes referred to as "structural analysis", constitutes a key methodology in modern sociology. SNA models social phenomenon using networks and studies them using graph theory. The vertices of the graph are referred to as nodes or actors and the edges between these nodes represent links or connections defined using an underlying relation. SNA has generated much interest in diverse disciplines well beyond sociology; they include but need not be limited to biology, economics, geography, organizational studies, political science, and computer science due to their ability to model potentially complex interactions in the underlying network. Examples of social networks include, but are not limited to, media networks, friendship and acquaintance networks, disease transmission, and biological gene networks. Network analysis techniques can broadly be described as either descriptive or statistical. Descriptive methods exploit measures based upon the structure of the network to quantify properties such as how connected or centralized the network is and quantifies the importance of each node in the network. Such methods abound in the literature and a comprehensive review is beyond the scope of this article [see, e.g., 28, 3, 13, 20, 25, 26, for excellent reviews]. Stochastic methods formulate statistical models for the network and include random graph models [24], latent space approaches [15], and several other probability models using matrix-variate distributions [8].

We outline some model-based strategies for quantifying uncertainty in descriptive SNA. Our particular focus is on small networks, where information is usually too limited to fit richer stochastic models such as the latent class or exponential random graph models. We pursue Bayesian inference for small networks. This is appealing because it offers exact sampling-based inference as opposed to asymptotics that may be difficult to envisage for small networks. We also offer direct and easy interpretation of uncertainty in estimation. We build a Bayesian hierarchical model for the variables associated with the network and compute the posterior distribution of some descriptive measures from the underlying graph using posterior predictive data replicates [see, e.g., 12]. There are a variety of such descriptive measures [see, e.g., 27, 28, for a detailed outline of such measures]; we opt for some common measures described in [28] and [18], often referred to as Krackhardt's axiomatic measures. Our choice is governed primarily by the availability of these measures in statistical software packages. We will also outline simple Bayesian inference on "PageRanks".

Our intended contribution lies at the interface between descriptive SNA and parametric Bayesian inference. The

Table 1. The adjacency matrix

|  | Site Names | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Cambridge | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | Glasgow | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 3 | UK-National | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 4 | Nottingham | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | STG | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 6 | UCL | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 7 | UMU | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 8 | Warwick | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 9 | Austin | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 10 | Berkeley | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 11 | CMU | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | Cornell | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| 13 | George Mason | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
| 14 | US-National | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| 15 | MIT | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 16 | NCSU | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| 17 | PSU | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 18 | Stanford | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

proposed approach will likely assist social scientists in gleaning information from small networks with limited information regarding the underlying processes generating structures within the network. We also emphasize easy implementation. The Bayesian models deliver clear interpretation and are implemented entirely within the `R` statistical computing environment using packages such as `sna` (for descriptive social networks) and `rjags` (for Bayesian computations).

Our application concerns a small network studied by [21] to better understand recent intercultural and social phenomenon involving migrant students and how they develop their ethnic identities. Narayan et al. [21] have conducted some descriptive structural analysis on a network constructed from websites of Hindu student groups within the United States and the United Kingdom to examine whether websites can help foster a transnational identity.

The article proceeds as follows. Section 2 gives a basic description of the network we will be revisiting. Section 3 begins with a brief review of familiar concepts in graph theory used in SNA. This is followed by a discussion of the quantities and measures in descriptive SNA that will be analyzed subsequently using Bayesian modeling, including uncertainty quantification with posterior predictive data replicates. Section 4 presents a Bayesian analysis of the network. We conclude the paper with a brief discussion in Section 5.

## 2. DATA

For the analysis, the dataset from Narayan, et al. [21] was used. The network analyzed there was a collection of 18 websites for Hindu student groups. Again, the focus of the paper is on small networks such as these. Websites 3 and 14 were the national student group websites for the United Kingdom and the United States, respectively. Based upon links from
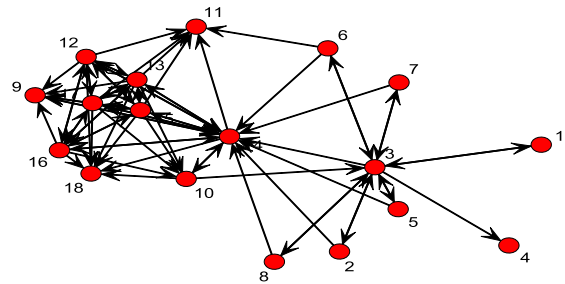
Figure 1. The network of websites represented as a directed graph corresponding to the adjacency table. The labels correspond to sites in the table.

those two websites, other websites were found that were local, functioning, and currently maintained. Seven additional websites were found from the UK, and nine other from the US. Pictorially, the network can be seen in Figure 1 and the corresponding adjacency matrix is shown in Table 1. A fully deterministic network analysis for this dataset can be found in Narayan, et al. [21].

## 3. METHODS

### 3.1 Descriptive social network analysis

Descriptive SNA refers to descriptive measures derived from the structure of the network. These measures are easy to compute and interpret, but do not account for the uncertainties inherent in modeling the relationships in the network. These measures can be regarded as local or global. Local measures determine the relative importance of a vertex (for example, how important or "central" a website is within a network), while global measures apply to the network as a

whole, such as how connected (or not) the network is. After some basic definitions in Section 3.1.1, we will discuss local measures using vertex analysis in Section 3.1.2 and PageRank algorithms in Section 3.1.3. We describe global measures using Krackhardt's axiomatic analysis in Section 3.1.4.

### 3.1.1 Definitions and measures

We will model the social network as a *directed graph*, or *digraph*, $G = \{V, E\}$, where $V$ is the set of vertices representing websites and $E = \{(i, j) : i \in V, j \in V\}$ consists of pairs of websites that are connected by a link from website $i$ to website $j$. The set of edges $E$ describes a relation on the set of vertices $V$. The *adjacency matrix* or relational table has its $(i, j)$-th entry equal to 1 if $(i, j) \in E$ and 0 if $i$ and $j$ are not related. Each edge can be thought of as having a direction. For any given vertex, its *indegree* is the number of edges coming into it, and its *outdegree* is the number of edges leaving from it. A *path* in $G$ is a sequence of vertices such that from each of its vertices there is an edge to the subsequent vertex. Also of interest is the distance between two vertices in a graph: this is the number of directed edges in the shortest path connecting them. Note that the shortest path is not always unique, but the distance is. The shortest paths are called *geodesics* and the distance is the *geodesic distance*. There are a number of other measurements defined in terms of distance: the eccentricity of a vertex $v$ is the greatest distance between $v$ and any other vertex. Two vertices $u$ and $v$ are called *connected* if $G$ contains a directed path from $u$ to $v$. Otherwise, the two vertices are called disconnected. A graph is called *connected* if every pair of vertices in the graph is connected. A *maximal connected subgraph* $S = (V_s, E_s)$ of $G$ is such that $E_s \subseteq E$ and $V_s \subseteq V$, if $i, j \in V$, $i \in V_s$, but $j \notin V_s$, neither $(i, j) \in E$ nor $(j, i) \in E$. A *connected component* of a graph, $G$, is the same as a maximal connected subgraph of $G$. Each vertex belongs to exactly one connected component, as does each edge. The graph is *weakly connected* if replacing all of its directed edges with undirected edges produces a connected (undirected) graph. It is *strongly connected*, or *strong*, if it contains a *directed* path from $u$ to $v$ for every pair of vertices $u, v \in V$. A *weak component*, $W$, is a subgraph of $G$ such that $W$ only becomes a maximal connected subgraph when its directed edges are replaced with undirected edges. The *strong components* are the maximal connected subgraphs (no need to replace the directed edges with undirected edges). A graph structure can be extended by assigning a weight to each edge of the graph. A digraph with weighted edges is called a network. Networks have many practical uses.

### 3.1.2 Vertex analysis using digraphs

Within graph theory and network analysis, there are various measures for the centrality of a vertex within a graph that determine the relative importance of that vertex (for example, how important a website is within a network). These measures attempt to quantify the prominence of an individual node embedded in a network. The measures investigated here include, among others, (i) *degree*, (ii) *betweenness*, (iii) *closeness*, (iv) *eigenvector centrality*, and (v) *structure statistics*.

The calculation of overall network centralization starts by determining the centrality of each individual in the network. There are many different definitions of centrality in the literature [28], and the choice between measures is based on the nature of the relationships. The most basic definition is based on degree: nodes that "receive" or "send" more connections are more central than those that do not. *Indegree* and *outdegree* can be used; in addition, the *total degree* can be used, which is the sum of the indegree and outdegree. This total degree is also called the *Freeman measure of connectedness or centrality* [10].

The *betweenness*, $CB(v)$, for vertex $v$ is

$$CB(v) = \sum_{i \neq v} \sum_{j \neq i, j \neq v} \frac{g_{ivj}}{g_{ij}}$$

where $g_{ij}$ is the number of geodesics from $i$ to $j$ and $g_{ivj}$ is the number of those geodesics that pass through $v$. Conceptually, vertices with high betweenness lie on a large number of non-redundant shortest paths between other vertices [9].

The next measure of centrality is *closeness*. The closeness of a vertex $v$ is defined as

$$CC(v) = \frac{|V| - 1}{\sum_{i:i \neq v} d(v, i)} \ ,$$

where $|V|$ is the cardinality of $V$ and $d(i, j)$ is the geodesic distance between $i$ and $j$ (where defined). Intuitively, closeness provides an index of the extent to which a given vertex has short paths to all other vertices in the graph.

Two other measures of centrality are *Harary centrality* and *eigenvalue centrality* of a vertex $v$. The Harary centrality of $v$ is

$$CH(v) = \frac{1}{\max_u d(v, u)}$$

where $d(v, u)$ is the geodesic distance from $v$ to $u$. *Eigenvector centrality* scores correspond to the values of the first eigenvector of the graph's adjacency matrix; the first eigenvector being the eigenvector corresponding to the largest eigenvalue. These scores may, in turn, be interpreted as arising from a reciprocal process in which the centrality of each actor is proportional to the sum of the centralities of those actors to whom it is connected [14].

Also useful is the *structure statistics*, which is computed as follows: let $d(i, j)$ be the geodesic distance from vertex $i$ to vertex $j$ in $G$. The structure statistics of $G$ are then given by the sequence $\{s_0, \ldots, s_{n-1}\}$, where $n = |V|$ and

$$s_i = \frac{1}{n^2} \sum_{j \in V} \sum_{k \in V} I(d(j, k) \leq i)$$

where $I(d(j,k) \leq i) = 1$ if $d(j,k) \leq i$ and 0 otherwise. Intuitively, $s_i$ is the fraction of $G$ which lies within distance $i$ of a randomly chosen vertex. Structure statistics have been of particular importance to biased net theorists, because of the link with Rapoport's original tracing model [28]. They may also be used along with component distributions or connectedness scores as descriptive indices of connectivity at the graph-level.

The centrality measures can also be aggregated to obtain a global centrality index for the entire network. For example, centralization refers to the extent to which the network is concentrated on one vertex or a group of vertices. Numerically, a centralized network is one which has a few or one vertex with considerably higher centrality scores than others in the network. The *centralization* or *global centrality index* of a graph $G$ for centrality measure $C(v)$ is defined as

$$C^*(G) = \sum_{i \in V} |C(i) - \max_{v \in V} C(v)| \ ,$$

or, the absolute deviation from the maximum of the centrality measure $C(v)$ on $G$ [10].

Tables showing all of the aforementioned measures have been presented in detail in [21], hence are not reproduced here.

### 3.1.3 PageRank algorithm

The PageRank algorithm [6] can be added to the other, more usual vertex analysis measures. It can be regarded as a centrality measure of a certain vertex within a network. This algorithm starts by creating a transition probability matrix, $\boldsymbol{P}$, whose rows and columns correspond to each vertex in the network, composed of $n$ vertices in total. For the $i$-th row, $r_i$ refers to the count of all vertices to which vertex $i$ has an edge, or hyperlink. So, $r_i$ is equal to the outdegree of $i$. Now,

$$(\boldsymbol{P})_{ij} = \begin{cases} 1/r_i & \text{if } i \text{ has a link to } j \\ 0 & \text{otherwise} \end{cases} ,$$

$\forall i, j$. If vertex $i$ has an outdegree of 0, then each entry in the $i^{\text{th}}$ row of $\boldsymbol{P}$ is simply $1/n$. It is easy to see that, for the $i^{th}$ row in $\boldsymbol{P}$, all of the entries sum to 1. Or, denoting $\mathbf{1}$ as the $n \times 1$ dimensional vector with all entries equal to 1 ($n$ being the number of rows or columns of $\boldsymbol{P}$),

$$\boldsymbol{P}\mathbf{1} = \mathbf{1} \ .$$

Thus, $\mathbf{1}$ is the eigenvector of $\boldsymbol{P}$ corresponding to the eigenvalue of 1. Matrices having this property are said to be row-stochastic and this transition probability matrix corresponds to a discrete Markov chain with $n$ states. It is well-known that if $\boldsymbol{P}$ is a *regular* transition matrix, i.e., some power of $\boldsymbol{P}$ has all its entries positive, then

$$\lim_{m \to \infty} \boldsymbol{P}^m = \boldsymbol{W} \ ,$$

where $\boldsymbol{W}$ is a matrix all of whose row vectors are the same [see, e.g. 19, 17, 1]. Denoting any one of these rows as $\boldsymbol{\alpha}^\top$, $\boldsymbol{\alpha}$ is the stationary distribution of $\boldsymbol{P}$. Should the matrix $\boldsymbol{P}$ converge, in this sense, to a stationary distribution $\boldsymbol{\alpha}$, then the PageRank algorithm returns the rank of the $i$-th vertex as the $i$-th entry in $\boldsymbol{\alpha}$.

As far as finding $\boldsymbol{\alpha}$ for a row-stochastic transition probability matrix $\boldsymbol{P}$, the following results provide a solution. First, if $\boldsymbol{W}$ and, therefore, $\boldsymbol{\alpha}$ exist, then the vector $\boldsymbol{\pi}$ such that

$$\boldsymbol{P}^\top \boldsymbol{\pi} = \boldsymbol{\pi}$$

is equal to $k\boldsymbol{\alpha}$, where $k$ is a scalar constant. Since $\boldsymbol{\alpha}^\top$ is a row of a row-stochastic matrix,

$$\boldsymbol{\alpha}^\top \mathbf{1} = 1 \ .$$

Thus, we can find $k$ from the relation

$$\left(\frac{1}{k}\right) \boldsymbol{\pi}^\top \mathbf{1} = 1$$

and $(1/k)\boldsymbol{\pi}$ is the stationary distribution. Notice that $\boldsymbol{\pi}$ is simply the eigenvector of $\boldsymbol{P}^\top$ associated with the eigenvalue of 1.

The second result is that all irreducible transition probability matrices converge to a unique stationary distribution. A transition probability matrix is irreducible if there exists no permutation matrix $\boldsymbol{M}$ such that

$$\boldsymbol{M}^\top \boldsymbol{P} \boldsymbol{M} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \\ \boldsymbol{O} & \boldsymbol{Z} \end{bmatrix} \ ,$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are both square matrices and $\boldsymbol{O}$ has all entries equal to zero ($\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{O}$, and $\boldsymbol{Z}$ are all dimensionally compatible). Thus, if the transition matrix can be shown to be irreducible, the eigenvector of its transpose that is associated with the eigenvalue of 1 is its stationary distribution, after appropriate scaling. The entries of the stationary distribution are the ranks for each vertex. The vertex with the highest rank is deemed the most important vertex in the network as it has the highest probability of getting a hit were there a web search over this network.

While the original transition probability matrix $\boldsymbol{P}$ may not always be irreducible, the algorithm actually proceeds to work with a similar transition probability matrix, $\widetilde{\boldsymbol{P}}$:

$$\widetilde{\boldsymbol{P}} = d\boldsymbol{P} + (1 - d)\left(\frac{1}{n}\right)\boldsymbol{J} \ ,$$

where $d \in [0, 1]$ and $\boldsymbol{J}$ is an $n \times n$ matrix with all entries equal to one. Conventionally, $d$ is set to 0.85 [6]. Notice that when $d > 0$, all of the entries in $\widetilde{\boldsymbol{P}}$ are strictly positive, meaning $\widetilde{\boldsymbol{P}}$ is irreducible and will always converge to a stationary distribution. The stationary distribution of $\widetilde{\boldsymbol{P}}$ returns the PageRanks for each vertex [19, 17, 1].

#### Table 2. PageRanks for the network

| Vertex | PageRank |
|--------|----------|
| 14 | 0.17 |
| 3 | 0.11 |
| 18 | 0.07 |
| 9 | 0.06 |
| 11 | 0.06 |
| 10 | 0.05 |
| 12 | 0.05 |
| 13 | 0.05 |
| 15 | 0.05 |
| 16 | 0.05 |
| 17 | 0.05 |
| 1 | 0.03 |
| 2 | 0.03 |
| 4 | 0.03 |
| 5 | 0.03 |
| 6 | 0.03 |
| 7 | 0.03 |
| 8 | 0.03 |

The above deterministic algorithm was implemented for the adjacency matrix corresponding to Table 1 and the PageRanks for the network are preseted in Table 2. The vertices are arranged from the highest to the lowest rank (up to 2 significant digits).

### 3.1.4 Krackhardt's axiomatic analysis

The preceding section devoted itself to analyzing the centrality properties in the structure of a network by locating and ranking nodes in the network according to their importance. We now turn to global measures describing the level of connectedness and the hierarchy of a network's structure. These are evaluated using Krackhardt's axiomatic approach [18]. Rather than assign an importance measure to each vertex, we now provide a score to the entire network. A basic method for assessing connectedness is constructing the reachability table or matrix for a network. Two vertices reach each other if there exists a path connecting them. A reachability graph is constructed from a given graph or network by joining or linking vertices that can be reached from one another in the original graph. For the reachability matrix, the $(i, j)$-th entry is 1 if vertex $j$ can be reached from vertex $i$ in the original network.

In addition to reachability, Krackhardt's measures can be used. There are four such measures developed for studying the underlying structure: connectedness, hierarchy, efficiency, and least upper bound (LUB), or "LUBness". These four measures quantify the four conditions Krackhardt considered necessary for a graph to be considered a hierarchy or an out-tree. First, *connectedness* is defined as:

$$C_K = 1 - \frac{D}{n(n-1)/2}$$

where $D$ is the number of pairs of points that cannot reach one another in a "weak" sense, meaning that all directed paths are converted to undirected paths before assessing reaching. In other words, Krackhardt's connectedness for a digraph $G$ is equal to the fraction of all pairs of nodes, $(i, j)$, such that there exists an undirected path from $i$ to $j$ in $G$. The connectedness score ranges from 0, for a null graph to 1, for a weakly or strongly connected graph.

The Krackhardt *efficiency* of a graph $G$ is computed as follows: suppose all of $G$'s weak components are $G_1, G_2, \ldots, G_m$. Denote the cardinalities of these graphs' vertex sets by $|V(G)| = N$ and $|V(G_i)| = N_i$, for $i = 1, \ldots, m$. Then the Krackhardt efficiency of $G$ is given by

$$E_K = 1 - \frac{|E| - \sum_{i=1}^{m}(N_i - 1)}{[N(N-1)/2] - \sum_{i=1}^{m}(N_i - 1)} \ ,$$

where $|E|$ is the number of edges in $G$. A high value, close to 1, implies that the network has the minimal number of edges in order for its weak components to be connected. Efficiency can also be interpreted as 1 minus the proportion of possible "extra" edges above those needed to connect the weak components. A graph with an efficiency of 1 has precisely as many edges as are needed to connect its components; as additional edges are added after that, the efficiency gradually falls towards 0.

Hierarchy measures quantify the extent of asymmetry in a structure. To understand symmetric directed paths, first consider that, for each directed path, there is a "sender", the first vertex in the sequence, and a "receiver", the final vertex in the sequence. A path is said to be *symmetric* if there exists a directed path from its receiver back to its sender; this new path is also trivially symmetric. All other such paths are said to be asymmetric. The Krackhardt *hierarchy* is defined as the fraction of paths which are asymmetric. The closer this value is to 1, the more hierarchical, in a conventional sense, are the relationships in the communications network. That is, a link from website A does not also have some existing path of links from website B back to A.

The *Least Upper Boundedness*, or "*LUBness*", is defined as follows. A node $k$ is said to be an "upper bound" for two nodes $i$ and $j$ if the directed paths $k \to i$ and $k \to j$ belong to $G$. If such a node does not exist, then $i$ and $j$ do not have an upper bound. An upper bound $l$ is known as a *least upper bound*, or LUB, for $i$ and $j$ if, for all upper bounds, $k$, of $i$ and $j$, $l$ belongs to at least one of the $k \to i$ paths and at least one of the $k \to j$ paths. Where all vertex pairs possess a least upper bound, Krackhardt's LUBness is equal to 1; in general, it approaches 0 as this condition is broached. Krackhardt offers that the LUB is the common "boss" for two nodes, either directly or up though the network.

Krackhardt's axiomatic analysis has also been presented in detail in [21], hence are not reproduced here. Instead, we will now turn to the statistical models for uncertainty quantification.

## 3.2 Bayesian modeling

### 3.2.1 Bayesian hierarchical modeling

To statistically model the various network analysis measures, the adjacency matrix must first be modeled. Let $Y_{ij}$ be the $(i,j)$-th entry of an $n \times n$ adjacency matrix $\boldsymbol{Y}$ associated with the network. We will assume a binary adjacency matrix, i.e.,

$$Y_{ij} = \begin{cases} 1 & \text{if there is an edge going from node } i \text{ to node } j \\ 0 & \text{if not} \end{cases} .$$

Note that the edges are directed, so the matrix $\boldsymbol{Y}$ need not be symmetric; a directed edge from node $i$ to node $j$ implies $Y_{ij} = 1$, but does not imply $Y_{ji} = 1$. Each entry in the adjacency matrix is modeled using logistic regression,

$$Y_{ij} \,|\, \pi_{ij} \sim \text{Bern}(\pi_{ij})$$
$$(1) \qquad \text{logit}(\pi_{ij}) = \boldsymbol{x}_{ij}^{\top} \boldsymbol{\beta} \,,$$

for $i,j = 1,2,\ldots,n$, where $\boldsymbol{x}_{ij}$ is a $p \times 1$ vector of covariates and $\boldsymbol{\beta}$ is the vector of slopes. The posterior distribution for $\boldsymbol{\beta}$ is

$$\text{p}(\boldsymbol{\beta} \,|\, \boldsymbol{Y}) \propto \text{p}(\boldsymbol{\beta}) \times \prod_{i,j=1}^{n} \text{Bern}(Y_{ij} \,|\, \text{logit}^{-1}(\boldsymbol{x}_{ij}^{\top} \boldsymbol{\beta})) \,,$$

where $p(\boldsymbol{\beta})$ is the prior on $\boldsymbol{\beta}$.

Bayesian logistic regression and the effect of priors (weakly informative, non-informative and informative) have been explored in depth by Gelman et al. in [11] and also by other authors in the context of noninformative (including flat or improper) priors for generalized linear models [see, e.g., 7, 2, 16, 30, among others]. A full exploration for different choices of priors is not attempted here. Instead, we will focus on some typical choices for the $p \times 1$ vector $\boldsymbol{\beta}$.

One specification assumes a weakly informative Gaussian prior, $\boldsymbol{\beta} \sim N(\boldsymbol{0}, \tau^2 \boldsymbol{I}_p)$, where $\tau^2$ is a positive scalar. A very large value of $\tau^2$ yields a vague proper prior for $\boldsymbol{\beta}$ and $\lim \tau^2 \to \infty$ results in the improper uniform prior. Necessary and sufficient conditions for the propriety of the posteriors under improper priors have been established rigorously in [7]. Here, we do not use improper priors, so our posterior distributions are proper.

Another proper specification assumes a multivariate normal-Wishart prior

$$\boldsymbol{\beta} \,|\, \boldsymbol{\Sigma} \sim \text{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma}^{-1} \sim \text{Wish}(\boldsymbol{I}_p, p+1) \,,$$

where $\boldsymbol{\Sigma}$ is an unknown $p \times p$ positive definite matrix, and $\boldsymbol{I}_p$ is the $p \times p$ identity matrix. Markov chain Monte Carlo (MCMC) methods are used to sample from $\text{p}(\boldsymbol{\beta} \,|\, \boldsymbol{Y})$ or $\text{p}(\boldsymbol{\beta}, \boldsymbol{\Sigma} \,|\, \boldsymbol{Y})$, as the case may be, for each of the above models [see, e.g., 12]. These can be implemented in a number of available R packages; for the specific analysis in this paper we used `rjags`.

### 3.2.2 Predictive distribution for network analysis measures

One possible approach to quantifying uncertainty in descriptive SNA is to evaluate the proposed descriptive measures using replicated datasets generated from the posterior predictive distribution [12]. A *replicated* data matrix is defined as the random $n \times n$ adjacency matrix $\boldsymbol{Y}^{\text{rep}}$ with entries $Y_{ij}^{\text{rep}}$ such that $\text{p}(\boldsymbol{Y}^{\text{rep}} \,|\, \boldsymbol{Y}, \boldsymbol{\theta}) = \text{p}(\boldsymbol{Y}^{\text{rep}} \,|\, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ represents the set of model parameters to be estimated. Thus, $\boldsymbol{Y}^{\text{rep}}$ is assumed to be generated from the same model as the observed data and is assumed to be conditionally independent of $\boldsymbol{Y}$ given $\pi_{ij}$. One can also regard $\boldsymbol{Y}^{\text{rep}}$ as a *future* or *alternate* adjacency matrix that could have been observed from the assumed underlying probability model for the realized $\boldsymbol{Y}$.

Since uncertainty quantification needs to account for the uncertainty in estimating $\boldsymbol{\theta}$, we compute the posterior predictive distribution of the replicated adjacency matrix,

$$(2) \qquad \text{p}(\boldsymbol{Y}^{\text{rep}} \,|\, \boldsymbol{Y}) = \int \text{p}(\boldsymbol{Y}^{\text{rep}} \,|\, \boldsymbol{\theta}) \text{p}(\boldsymbol{\theta} \,|\, \boldsymbol{Y}) d\boldsymbol{\theta} \,,$$

where $Y_{ij}^{\text{rep}} \,|\, \boldsymbol{\pi} \overset{ind}{\sim} \text{Bern}(\pi_{ij})$ and $\pi_{ij}$ is determined by $\boldsymbol{\beta}$. We sample from (2) using samples from $p(\boldsymbol{\beta} \,|\, \boldsymbol{Y})$. From each sampled $\boldsymbol{\beta}^{(s)} \sim p(\boldsymbol{\beta} \,|\, \boldsymbol{Y})$ for $s = 1,2,\ldots,S$, we compute the corresponding $\boldsymbol{\pi}^{(s)}$ and draw $Y_{ij}^{\text{rep},(s)} \sim \text{Bern}(\pi_{ij}^{(s)})$ for $i,j = 1,2,\ldots,n$. The collection $\{\boldsymbol{Y}^{\text{rep},s}\}$ for $s = 1,2,\ldots,S$ are samples of the replicated data from the distribution in (2). Each of the descriptive measures in Section 3 are fully determined from the graph $G$, hence the adjacency matrix $\boldsymbol{Y}$. Therefore, we can express them as a function $T(\boldsymbol{Y})$. Computing $T(\boldsymbol{Y}^{\text{rep},s})$ for each $s$ yields samples from $\text{p}(T(\boldsymbol{Y}^{\text{rep}}) \,|\, \boldsymbol{Y})$.

### 3.2.3 Bayesian PageRank model

Posterior predictive distributions for replicated adjacency matrices can also offer uncertainty quantification in PageRanks by applying the PageRank algorithm to the replicated datasets. However, the $\pi_{ij}$'s from (1) are not modeled jointly and, hence, do not necessarily correspond to a transition matrix; i.e., they need not be row-stochastic. An alternative is to develop a Bayesian model for the row-stochastic matrix involved in the PageRank calculation. The following model is an option.

$$(3) \qquad \begin{aligned} Y_{ij} \,|\, \pi_{ij} &\overset{ind}{\sim} \text{Bern}(\pi_{ij}) \,, \quad i,j = 1,2,\ldots,n \,, \\ \boldsymbol{\pi}_i \,|\, \boldsymbol{\alpha} &\overset{ind}{\sim} \text{Dirichlet}(\boldsymbol{\alpha}) \,, \quad i = 1,2,\ldots,n \,, \\ \alpha_1 &= \frac{1}{1 + \sum_{m=2}^{n} \mu_m} \,, \\ \alpha_j &= \frac{\mu_j}{1 + \sum_{m=2}^{n} \mu_m} \,, \quad j = 2,3,\ldots,n \,, \\ \mu_j &= \exp\left(\boldsymbol{x}_j^{\top} \boldsymbol{\beta}\right) \,, \quad j = 2,\ldots,n \\ \boldsymbol{\beta} &\sim \text{N}(\boldsymbol{0}, \tau^2 \boldsymbol{I}_p) \,, \end{aligned}$$

where $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{in})^\top$, $\alpha_i$ is the $i$-th element of the $n \times 1$ vector $\boldsymbol{\alpha}$, where the $\alpha_i$'s sum to one, and each $\boldsymbol{x}_j$ is a $p \times 1$ vector of covariates.

Since the support of the Dirichlet distribution is a simplex, the matrix with $\boldsymbol{\pi}_i^\top$ is row-stochastic. So, the posterior distribution for the $\pi_{ij}$'s define the posterior distribution of the PageRank transition matrix. Computing the stationary distribution for each posterior sample of this row-stochastic matrix will return a sample of the posterior distribution for the PageRank vector. The posterior samples for the $i$-th entry of these vectors gives the posterior distribution of the PageRank for the $i$-th node. Ranks assigned according to this algorithm will be referred to as Dirichlet-1-Ranks for the rest of this paper.

A generalization allows the prior on $\boldsymbol{\pi}_i$ to vary across $i$:

$$Y_{ij} \,|\, \pi_{ij} \stackrel{ind}{\sim} \text{Bern}(\pi_{ij}), \quad i, j = 1, 2, \ldots, n \,,$$

$$\boldsymbol{\pi}_i \,|\, \boldsymbol{\alpha}_i \stackrel{ind}{\sim} \text{Dirichlet}(\boldsymbol{\alpha}_i) \,, \quad i = 1, 2, \ldots, n \,,$$

$$(4) \qquad \alpha_{i1} = \frac{1}{1 + \sum_{m=2}^n \mu_{im}} \,, \quad i = 1, 2, \ldots, n \,,$$

$$\alpha_{ij} = \frac{\mu_{ij}}{1 + \sum_{m=2}^n \mu_{im}} \,, \quad j = 2, \ldots, n \,,$$

$$\mu_{ij} = \exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta}) \,, \quad i = 1, 2 \ldots, n \,;\; j = 2, \ldots, n \,.$$

## 4. RESULTS

For our data we will consider the following covariates,

$$X_{1i} = \begin{cases} 1 & \text{if vertex } i, \text{ the "sender", is from the UK} \\ 0 & \text{if } i \text{ is from the US} \end{cases}$$

$$X_{2j} = \begin{cases} 1 & \text{if vertex } j, \text{ the "recipient", is from the UK} \\ 0 & \text{if } j \text{ is from the US} \end{cases} .$$

We fit five specific models. The first three models use the likelihood in (1). Model 1 specifies $\boldsymbol{x}_{ij}^\top = (1, X_{1i}, X_{2j})$, so $p = 3$, and specifies a vague Gaussian prior $\boldsymbol{\beta} \sim N(\boldsymbol{0}, \tau^2 \boldsymbol{I}_3)$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$. Model 2 uses the same $\boldsymbol{x}_{ij}$ as Model 1, but uses the Normal-Wishart specification for $\{\boldsymbol{\beta}, \boldsymbol{\Sigma}\}$ as described in Section 3.2.1. Model 3 uses the same Normal-Wishart prior for $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\Sigma}\}$ as in Model 2, but adds an interaction term to the regressors, so $\boldsymbol{x}_{ij}^\top = (1, X_{1i}, X_{2j}, X_{1i}X_{2j})$ and $p = 4$ with $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$. Models 4 and 5 correspond to (3) and (4), respectively. In Model 4, we take $\boldsymbol{x}_j^\top = (1, X_{3j})$, where

$$X_{3j} = \begin{cases} 1 & \text{if site } k \text{ is from the U.K.} \\ 0 & \text{if site } k \text{ is from the U.S.} \end{cases} ,$$

and in Model 5 we take $\boldsymbol{x}_{ij}$ to be the same as for Model 3.

The results from models 1, 2, and 3 are summarized in Tables 3, 4, and 5, respectively. Since the main slope coefficients have 95% Bayesian credible intervals with 2.5th and 97.5th quantiles of the posterior distributions as endpoints

*Table 3.* Posterior summaries for Model 1

| Coefficient | Mean | SD | 95% CI |
|---|---|---|---|
| $\beta_0$ | -0.20 | 0.20 | [-0.59, 0.20] |
| $\beta_1$ | -1.14 | 0.31 | [-1.75, -0.55] |
| $\beta_2$ | -1.68 | 0.33 | [-2.36, -1.04] |

*Table 4.* Posterior summaries for Model 2

| Coefficient | Mean | SD | 95% CI |
|---|---|---|---|
| $\beta_0$ | -0.27 | 0.19 | [-0.65, 0.19] |
| $\beta_1$ | -1.04 | 0.29 | [-1.61, -0.48] |
| $\beta_2$ | -1.50 | 0.31 | [-2.15, -0.95] |
| $\sigma_0^2$ | 1.03 | 3.09 | [0.11, 5.31] |
| $\sigma_1^2$ | 1.93 | 4.87 | [0.19, 9.95] |
| $\sigma_2^2$ | 3.15 | 14.45 | [0.30, 15.64] |
| $\rho_{01}$ | 0.14 | 0.50 | [-0.83, 0.92] |
| $\rho_{02}$ | 0.18 | 0.50 | [-0.82, 0.92] |
| $\rho_{12}$ | 0.52 | 0.41 | [-0.54, 0.97] |

*Table 5.* Posterior summaries for Model 3

| Coefficient | Mean | SD | 95% CI |
|---|---|---|---|
| $\beta_0$ | 0.14 | 0.20 | [-0.24, 0.53] |
| $\beta_1$ | -2.41 | 0.43 | [-3.30, -1.61] |
| $\beta_2$ | -3.97 | 0.88 | [-6.02, -2.54] |
| $\beta_3$ | 4.76 | 1.07 | [3.00, 7.22] |
| $\sigma_0^2$ | 0.98 | 3.13 | [0.11, 4.77] |
| $\sigma_1^2$ | 6.84 | 37.80 | [0.64, 32.06] |
| $\sigma_2^2$ | 16.48 | 60.53 | [1.45, 82.03] |
| $\sigma_3^2$ | 23.48 | 85.51 | [1.99, 120.05] |
| $\rho_{01}$ | -0.12 | 0.51 | [-0.92, 0.84] |
| $\rho_{02}$ | -0.12 | 0.51 | [-0.92, 0.84] |
| $\rho_{03}$ | 0.12 | 0.51 | [-0.84, 0.92] |
| $\rho_{12}$ | 0.85 | 0.19 | [0.29, 0.99] |
| $\rho_{13}$ | -0.86 | 0.19 | [-1.00, -0.33] |
| $\rho_{23}$ | -0.92 | 0.12 | [-1.00, -0.62] |

(subsequently referred to as CI) that are wholly below zero in all three models, the network seems to imply that the UK websites were less likely than US sites to both "send" and "receive" a hyperlink to and from another site. In fact, re-examining the adjacency matrix (Table 1), the UK sites are generally fairly isolated. The UK-National site has links to all the other UK sites, and to the US-National Site. Both of the national sites receive links from most of the UK sites. But, outside of the national websites, the only hyperlink is from UCL to CMU, which *is* a link from a UK site to a US site, or a transnational link.

While none of the correlations between the slope parameters in Model 2 were deemed significantly different from zero (the credible intervals all included 0), there were statistically significant correlations between all pairs of the slope coefficients ($\beta_1$, $\beta_2$, and $\beta_3$) in Model 3. For the latter model, the interaction slope coefficient, $\beta_3$, had a 95% CI that was wholly above zero, which indicates significantly higher probabilities of connections between nodes within a country than

between countries (described as lack of transnational connections in [21]). We also note that while modest posterior shrinkage is observed, most of the 95% CI's for the correlations in Table 4 are somewhat wide and there is no significant correlation between the global intercept and either of the two main effect slope parameters. We see greater shrinkage in Table 5, clearly brought about by the introduction of the interaction parameter. Here, the correlations between the interaction effects and the two main effect slopes are significant, but not with the global intercept; the two main effect slopes continue to be essentially uncorrelated with each other and the global intercept.

A model's fit was assessed using the Watanabe-Akaike Information Criterion, or Widely Applicable Information Criterion (WAIC) [see, e.g., 12, 29]. We used the definition

$$(5) \qquad \text{WAIC} = -2 \times \left( \text{lppd} - p_{\text{waic}} \right) \, ,$$

where lppd is the *log pointwise predictive density*, which is the log of the pointwise likelihood averaged over the posterior distribution of the model parameters, and summed over all the data points. This yields a measure of predictive accuracy and is computed as

$$(6)$$
$$\sum_{i,j=1}^{n} \log \text{E}_{\boldsymbol{\theta}\,|\,\boldsymbol{Y}} \left[ \text{p}(Y_{i,j}\,|\,\boldsymbol{\theta}) \right] \approx \sum_{i,j=1}^{n} \log \left( \frac{1}{S} \sum_{s=1}^{S} \text{p}(Y_{ij}\,|\,\boldsymbol{\theta}^{(s)}) \right) \, ,$$

where $\boldsymbol{\theta}^{(s)}$ for $s = 1, 2, \ldots, S$ denote posterior samples of model parameters, and $p_{\text{waic}}$ is a measure of model complexity given by the sum of the posterior variance of the log predictive density for each observed data point $Y_{ij}$. This is computed as

$$(7)$$
$$p_{\text{waic}} = \sum_{i,j=1}^{n} \text{var}_{\text{post}} \left( \log \text{p}(Y_{ij}\,|\,\boldsymbol{\theta}) \right)$$
$$\approx \sum_{i,j=1}^{n} \text{var}_{\text{sample}} \left( \left\{ \log \text{p}(Y_{ij}\,|\,\boldsymbol{\theta}^{(s)}) : s = 1, 2 \ldots, S \right\} \right)$$

where $\text{var}_{\text{sample}} \left( \left\{ \log \text{p}(Y_{ij}\,|\,\boldsymbol{\theta}^{(s)}) : s = 1, 2 \ldots, S \right\} \right)$ computes the sample variance of the quantities $\log \text{p}(Y_{ij}\,|\,\boldsymbol{\theta}^{(s)})$ for $s = 1, 2, \ldots, S$. Lower values for WAIC indicate a better-fit model. The WAIC values for models 1, 2, and 3 are 355.0, 315.1, and 269.5, respectively. So, model 3 was used for generating draws from the posterior predictive distribution for the subsequent analysis.

For plots indicating the medians and the 95% CI's for the odds ratios comparing transnational connections against intra-national connections, see Figure 2. Note that these intervals are based only on Model 3 and that the reference groups are both types of intra-national connections (US-to-US and UK-to-UK).
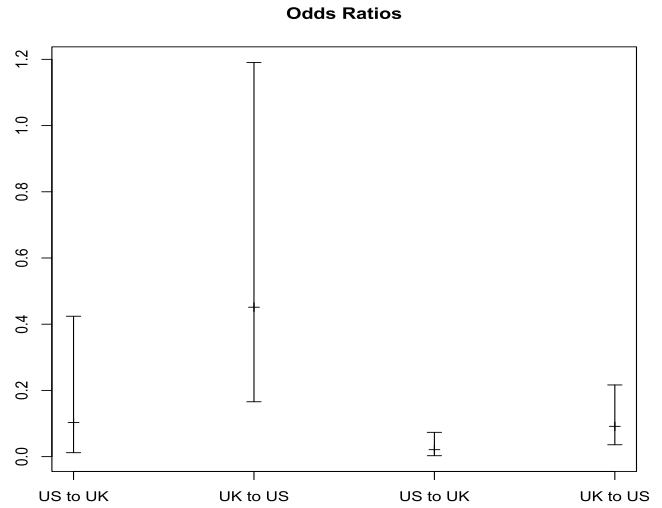


*Figure 2. Plot of the medians and 95% CI's for various odds ratios; the first two intervals are compared against a reference group of UK-to-UK and the second two are compared against a reference group of US-to-US.*

The medians and the 95% posterior predictive credible intervals based for the network analysis measures described in Sections 3.1.2 and 3.1.4 are presented in Tables 6 and 7, respectively. Here, the network was replicated using draws from (2) and the measures were computed for each sampled network as described in Section 3.2.2. This generated posterior predictive samples for each of the network analysis measures from which we computed point estimates and the 95% posterior predictive credible intervals.

The point estimates of the vertex analysis measures in Table 6 seem to be consistent with earlier findings reported in [21], hence not reproduced here, but quantify uncertainty which will enable social scientists to attach confidence to their conclusions. In general, there was much overlap between all the credible intervals for a specific vertex measure, especially the intervals for betweenness, closeness, Harary centrality, and eigenvector centrality. The structure statistics table indicates that at least half of all possible site-pairs have a geodesic distance of 4 or less.

Looking at the table for Krackhardt's axiomatic analysis in Table 7, the network appears to be very connected, since the 95% CI for connectedness is $(0.79, 1)$. The network also seems to be fairly efficient, with an interval of $(0.75, 0.88)$; Krackhardt's measures range from 0 to 1. The point estimates appear consistent with what was concluded from the structure statistics table reported in [21]. The network does not seem to be too hierarchical, with a CI of $(0.00, 0.68)$ for that measure. The CI for LUBness indicates a lack of information from the network on that property, since it spans almost the entire range of possible values for LUBness.

Finally, the summaries of the posterior distributions for $\boldsymbol{\beta}$ in Models 4 and 5 are seen in Tables 8 and 9, respectively.

Table 6. Medians and 95% CI's for the vertex centrality measures

| Vertex | Outdegree | Indegree | Freeman | Betweenness | Closeness | Harary | E-center |
|--------|-----------|----------|---------|-------------|-----------|--------|----------|
| 1 | 2 [0, 6] | 1 [0, 4] | 4 [1, 8] | 5.3 [0.0, 65.5] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.1 [0.0, 0.3] |
| 2 | 2 [0, 6] | 1 [0, 4] | 4 [1, 8] | 5.5 [0.0, 64.2] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.1 [0.0, 0.3] |
| 3 | 2 [0, 6] | 1 [0, 4] | 4 [1, 8] | 5.3 [0.0, 65.0] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.1 [0.0, 0.3] |
| 4 | 2 [0, 6] | 1 [0, 4] | 4 [1, 8] | 5.9 [0.0, 64.5] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.1 [0.0, 0.3] |
| 5 | 2 [0, 6] | 1 [0, 4] | 4 [1, 8] | 5.7 [0.0, 64.8] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.1 [0.0, 0.3] |
| 6 | 2 [0, 6] | 1 [0, 4] | 4 [1, 8] | 5.4 [0.0, 64.0] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.1 [0.0, 0.3] |
| 7 | 2 [0, 6] | 1 [0, 4] | 4 [1, 8] | 5.3 [0.0, 63.1] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.1 [0.0, 0.3] |
| 8 | 2 [0, 6] | 1 [0, 4] | 4 [1, 8] | 5.5 [0.0, 64.0] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.1 [0.0, 0.3] |
| 9 | 5 [2, 8] | 6 [2, 9] | 11 [6, 15] | 13.0 [1.3, 63.5] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.3 [0.1, 0.4] |
| 10 | 5 [2, 8] | 6 [2, 9] | 11 [6, 16] | 13.1 [1.4, 64.7] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.3 [0.1, 0.4] |
| 11 | 5 [2, 8] | 6 [2, 9] | 11 [6, 15] | 13.0 [1.2, 63.7] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.3 [0.1, 0.4] |
| 12 | 5 [2, 8] | 6 [2, 9] | 11 [6, 16] | 13.0 [1.3, 65.3] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.3 [0.1, 0.4] |
| 13 | 5 [2, 8] | 6 [2, 9] | 11 [6, 15] | 13.2 [1.3, 65.1] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.3 [0.1, 0.4] |
| 14 | 5 [2, 8] | 6 [2, 9] | 11 [6, 16] | 13.3 [1.3, 63.1] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.3 [0.1, 0.4] |
| 15 | 5 [2, 8] | 6 [2, 9] | 11 [6, 15] | 12.9 [1.2, 64.0] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.5] | 0.3 [0.1, 0.4] |
| 16 | 5 [2, 8] | 6 [2, 9] | 11 [6, 16] | 13.1 [1.4, 65.0] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.3 [0.1, 0.4] |
| 17 | 5 [2, 8] | 6 [2, 9] | 11 [6, 16] | 13.0 [1.3, 67.0] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.3 [0.1, 0.4] |
| 18 | 5 [2, 8] | 6 [2, 9] | 11 [6, 16] | 13.3 [1.3, 65.4] | 0.0 [0.0, 0.5] | 0.0 [0.0, 0.3] | 0.3 [0.1, 0.4] |
| Centralization | 0.2 [0.1, 0.3] | 0.3 [0.2, 0.4] | 0.2 [0.1, 0.3] | 0.1 [0.05, 0.3] | 0.07 [0.0, 0.6] | 0.02 [0.0, 0.4] | 0.2 [0.1, 0.3] |

Table 7. Medians and 95% CI's for the Krackhardt's axiomatic analysis measures

| Measures from Krackhardt's axioms | 95% CI |
|-----------------------------------|--------|
| Connectedness | 1.00 [0.79, 1.00] |
| Efficiency | 0.82 [0.75, 0.88] |
| Hierarchy | 0.33 [0.00, 0.68] |
| LUBness | 1.00 [0.11, 1.00] |

Table 8. Posterior summaries for Model 4

| Coefficient | Mean | SD | 95% CI |
|-------------|------|-----|--------|
| $\beta_0$ | 0.85 | 0.38 | [0.10, 1.56] |
| $\beta_1$ | -1.08 | 0.22 | [-1.46, -0.69] |

Table 9. Posterior summaries for Model 5

| Coefficient | Mean | SD | 95% CI |
|-------------|------|-------|--------|
| $\beta_0$ | 11.39 | 92.12 | [-163.71, 188.39] |
| $\beta_1$ | 14.02 | 94.39 | [-173.11, 198.68] |
| $\beta_2$ | -0.51 | 0.16 | [-0.82, -0.19] |
| $\beta_3$ | 0.61 | 0.34 | [-0.13, 1.19] |



**Dirichlet−1−Rank Posterior Predictive CI's**

Figure 3. Plot of the medians and 95% CI's for the Dirichlet-1-Ranks for the 18 sites. Blue bars indicate UK sites and red bars indicate US sites.

Here, it is worth pointing out that the interpretations of the regression coefficients are different from those in Models 1–3. In fact, the purpose of incorporating the covariates in the $\mu$'s in Models 4 and 5 are solely to allow the Dirichlet distribution to vary by the nodes, thereby adding flexibility. Our main inferential goal concerns the ranks. The means, standard deviations, and 95% CI's for the Dirichlet-1-Ranks are presented in Figure 3. A summary plot for the Dirichlet-2-R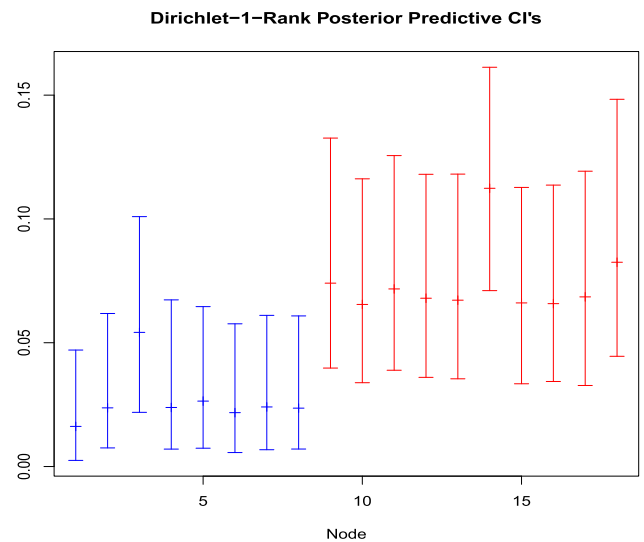anks is presented in Figure 4. For the most part, these two plots are similar. The posterior medians seem to be more or less consistent with the PageRank measures in Table 2, but the inherent uncertainties gleaned here (but not from the deterministic PageRank algorithm) should add caution to making substantial conclusions regarding these rankings.

While there is still overlap between most of the 95% CI's seen here, there is, as in the other results, a split between the US and the UK sites. The UK sites have median estimates that are consistently below those of the US sites. There is also a spike in mean estimates for sites 3 and 14, the two

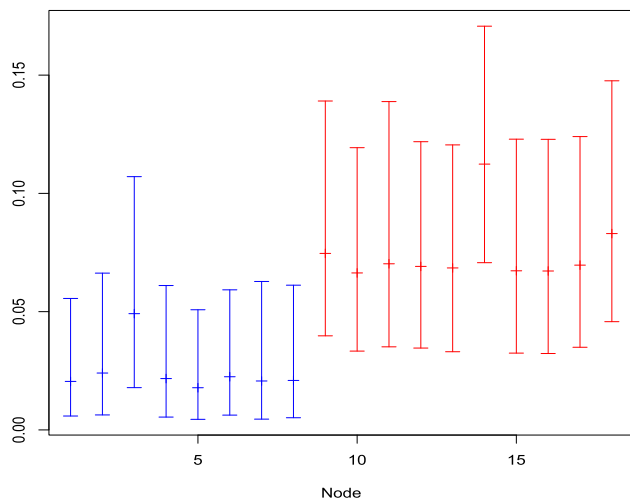**Dirichlet−2−Rank Posterior Predictive CI's**



Figure 4. Plot of the medians and 95% CI's for the Dirichlet-2-Ranks for the 18 sites. Blue bars indicate UK sites and red bars indicate US sites.

national sites, which was not seen in the other statistical results seen here. It is, however, noteworthy that there is no overlap between the CI for the US-National site and any of the UK sites, except for the UK-National website. These last two facts demonstrate that both of the Dirichlet-Ranks appear to be more sensitive or tuned than the other statistical approaches presented in this paper.

# 5. CONCLUSION

This paper has presented some parametric Bayesian methods for structural network analysis of small networks with limited information. The intended contribution is to offer sociologists some easily implementable Bayesian models that can offer uncertainty quantification while carrying out descriptive SNA. Basic descriptive measures of directed graphs, PageRank algorithms and Krackhardt's axiomatic analysis are considered in conjunction with parametric Bayesian hierarchical models. The Bayesian paradigm is particularly attractive as we do not need to rely upon asymptotic validation of our inference based on the numbers of vertices (or edges) becoming larger and larger. The methods are easily implemented using R packages `sna` and `rjags`. Bayesian inference provides new insights into a rather structurally sparse network of Hindu student groups, previously studied without uncertainty quantification, to better understand transnational identities.

We clarify that our treatment here is by no means exhaustive. For example, there are numerous other structural quantities that could be added to the measures we considered here in Section 3 as described, for example, in [3, 5, 20, 25, 26, 27]. Our choice of the measures and models investigated here has been dictated to a substantial extent

by their accessibility to practicing social scientists in the form of R packages. Investigations with these networks can be part of future studies. We can also easily incorporate latent random effects in the mean structures for our models. This can be done for each of Models 1–5 here. We have deliberately not attempted to fit such, more complex models here, given that the random effects would not be able to learn much from the small network. However, this is central to the objectives of this paper where we have focused upon methods that work well for small networks, and will be equally applicable to large networks as well.

In addition, there are much richer stochastic models such as exponential random graph models and Erdos-Renyi models [see, e.g., 4, 23]. The logistic regression models described here can be looked upon as heterogeneous versions of the Erdos-Renyi models where the probability of a directed edge between two nodes is not uniform but depends upon attributes (covariates) associated with the sender and receiver. There are other models, such as the Barbási-Albert models [22], that are generative and can simulate artificial networks given certain characteristics, but it is less clear how to carry out model-based Bayesian inference from observed networks, as we do here, with these models. We identify these as possible areas for future research.

# 6. SUPPLEMENTARY MATERIAL

Supplement to the paper (http://intlpress.com/site/pub/pages/journals/items/sii/content/vols/0012/0001/s004) comprises the R programs used for the analysis.

# REFERENCES

[1] BANERJEE, SUDIPTO and ROY, ANINDYA. (2014). *Linear Algebra and Matrix Analysis for Statistics*, 1st ed. CRC Press, Boca Raton, FL. MR3222172

[2] BEDRICK, E.J., CHRISTENSEN, R. and JOHNSON, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, **91**, 1450–1460. MR1439085

[3] BOCCALETTI, S., LATORA, V., MORENO, Y., CHAVEZ, M. and HWANG, D.-U. (2006). Complex networks: Structures and dynamics. *Physics Reports*, **424**, 175–308. MR2193621

[4] BOLLOBÁS, B. (2001). *Random Graphs* Second Edition. Cambridge University Press: Cambridge UK. MR1864966

[5] BORGATTI, S.P. (2005). Centrality and network flow. *Social Networks*, **27**, 55–71.

[6] BRIN, SERGEY and PAGE, LAWRENCE. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*. **30** 107–117.

[7] CHEN, M.-H. and SHAO, Q.-M. (2000). Propriety of posterior distribution for dichotomous quantal response models. *Proceedings of the American Mathematical Society*, **129**, 293–302. MR1694452

[8] CHIKUSE, Y. (2003). *Statistics on Special Manifolds*. vol. 174, Lecture Notes in Statistics. Springer-Verlag, New York. ISBN 0-387-00160-3. MR1960435

[9] FREEMAN, L.C. (1977). A set of measures of centrality based on betweenness. *Sociometry* **40** 35–41.

[10] Freeman, L.C. (1979). Centrality in social networks I: Conceptual clarification. *Social Networks* **1** 215–239.

[11] Gelman, A., Jakulin, A., Pittau, M.G. and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, **2**, 1360–1383. MR2655663

[12] Gelman, Andrew, Carlin, John B., Stern, Hal S., Dunson, David B., Vehtari, Aki, and Rubin, Donald B. (2013). *Bayesian Data Analysis*, 3rd ed. CRC Press, Boca Raton, FL. MR3235677

[13] Costa, L. da F., Rodrigues, F.A., Travieso, G. and Villas Boas, P.R. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, **56**, 167–242.

[14] Harville, D.D. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag, New York. MR1467237

[15] Hoff, P.D., Raftery, A.E. and Handcock, M.S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, **97**, 1090–1098. MR1951262

[16] Ibrahim, J.G. and Laud, P.W. (1991). On Bayesian analysis of generalized linear models using Jeffrey's prior. *Journal of the American Statistical Association*, **86**, 981–986. MR1146346

[17] Kemeny, J.G. and Snell, L.J. (1976). *Finite Markov Chains*. New York: Springer-Verlag. MR0410929

[18] Krackhardt, D. (1994). Graph theoretical dimensions of informal organizations. In Kathleen M. Carley and Michael J. Prietula (Eds.), *Computational Organization Theory* pp. 89–111. Lawrence Erlbaum Associates, Hillsdale, NJ.

[19] Langville, A.N. and Meyer, C.D. (2006). *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton, NJ: Princeton University Press. MR2262054

[20] Newman, M.E.J. (2010). *Networks: An Introduction.* Oxford University Press. Oxford, UK. MR2676073

[21] Narayan, A., Purkayastha, B. and Banerjee, S. (2011). Constructing transnational and virtual ethnic identities: A study of the discourse and networks of ethnic student organisations in the USA and UK. *Journal of Intercultural Studies* **32**, 515–537.

[22] Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, **74**, 47–97. MR1895096

[23] Erdős, P. and Rényi, A. (1959). On random graphs. I. *Publicationes Mathematicae* **6**, 290–297. MR0120167

[24] Robins G., Snijders T., Wang P., Handcock M., and Pattison P. (2007). Recent developments in exponential random graph (p*) models for social networks. *Social Networks*, 29, 192–215.

[25] Scott, J. and Carrington, P.J. (2011). *The SAGE Handbook of Social Network Analysis*. SAGE Publications Ltd. London, UK.

[26] Victor, J.N., Montgomery, A.H. and Lubell, M. (2018). *The Oxford Handbook of Political Networks*. Oxford University Press, New York.

[27] Vigna, S. and Boldi, P. (2014). Axioms for centrality. *Internet Mathematics*, **10**, 222–262. MR3259267

[28] Wasserman, S. and Frost, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, New York.

[29] Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research*, **14**, 867–897. MR3049492

[30] Zellner, A. and Rossi, P.E. (1984). Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics*, **25**, 365–393. MR0752447

Thomas Nemmers
Lockton Companies, LLC
4275 Executive Square, Suite 600
La Jolla, CA 92037
USA
E-mail address:  TNemmers@lockton.com

Anjana Narayan
Associate Professor
Department of Psychology and Sociology
California State Polytechnic University
3801 West Temple Avenue, Pomona, CA 91768
USA
E-mail address: anarayan@cpp.edu

Sudipto Banerjee
Professor and Chair
UCLA Department of Biostatistics
650 Charles E. Young Dr. South
Los Angeles, CA 90095
USA
E-mail address: sudipto@ucla.edu