Available online at www.sciencedirect.com

ScienceDirect

Social Networks 00 (2015) 1–17

Social Networks

Hierarchical Structure in Social Networks

Cynthia Cook[1a], Matthew J. Denny[b], Mitchell Goist[c], Timmy Huynh[d]

[a]*Department of Statistics cmc496@psu.edu*
[b]*Department of Political Science, mdenny@psu.edu*
[c]*Department of Political Science mlg307@psu.edu*
[d]*Department of Sociology and Criminology, tnh133@psu.edu*

Abstract

Keywords: Hierarchy, Network, Power

1. Introduction

Hierarchy is an important feature of many organizations, such as firms, social clubs, and military units. Formally, we can define a hierarchy as a system where people or groups are ranked according to status or authority. Yet it is difficult to operationalize this definition for measurement and comparison. There has been a great deal of research on power and status in groups and organizations, but most of this research relies on measurements defined over domain specific rankings, such as job titles. At the same time, networks scholars have defined a number of broadly applicable hierarchy metrics based on network structure, but these metrics are not necessarily grounded in meaningful sociological concepts of status and authority. Contrastingly, social theorists like Michael Mann have noted the messiness of society and that a network-oriented perspective of the "sociospatial and organizational model [of a network]" can explicate the "sources of social power," [**?**] but they have generally not delved into the methodologies through which to fully explore such power dynamics. In this paper, we seek to bring together these two areas of research, and to develop a framework for measuring hierarchy in social networks that is both generally applicable and exhibits a high degree of construct validity.

Without statistical models/mathematical measurements for hierarchy which are theoretically based, and vice versa; theory that can be statistical/mathematically quantified and verified, the conceptual idea of hierarchy cannot be fully understood. We do not suggest that this project will achieve an overreaching theory and methods, but we strive to take the first step. At the very least, we will try to demonstrate the need for a united theory and corresponding methods. As an interdisciplinary team, we are in the unique position to accomplish our goals.

2. Sociological Theories of Hierarchy

We are still working through evaluating a few different datasets to best suit our purposes. However, at present, this is a little difficult because we really want our measure to be theoretically-grounded, but we haven't yet developed a solid theoretical conception for hierarchy. Thus far, theory-wise, the Mann (1986) definition seems closest to the

[1]Authors are listed in alphabetical order but all contributed equally to this publication.

Liu-Driver measures discussed in the Mones et al. (2012) article: i.e., hierarchical networks are those in which the actions of a few nodes are needed to take control of the graph. Another potential definition, also implied, is hierarchy means the mechanisms of collective actions (i.e., the ability of different nodes to connect with one another) hinges on a small number.

3. Measuring Hierarchy

A number of analytical measures of hierarchy have been proposed for directed networks. Most proposed measures return a scalar value that is meant to capture the "hierarchical-ness" of a given network [1, 2**?** **?** , 3**?** , 4]. This theoretically facilitates comparison between measures calculated on the same network, as well as comparison on the same measure across multiple networks. However some approaches to measuring hierarchy only provide a local measure of importance or position in the hierarchy for each node in the network [**?**]. For these measures, one can calculate a Gini coefficient [**?**] from the individual level scores and use these coefficients as a proxy for a global measure.

In this section, we introduce and describe twelve candidate measures of network hierarchy that have been previously used in the networks literature. It is important to note that most of the measures we consider are only defined for directed networks, and thus for the remained of this paper we assume all networks under consideration are directed. We begin by introducing some terminology that will be common across all measures. For a given network $G = (V, E)$, let $V = \{v_i\}_{i=1}^{N}$ be the set of N vertices (nodes) associated with G, and $E = \{e_j\}_{j=1}^{M}$ be the set of M edges associated with G. Furthermore, for a given edge e_j, let $e_j^{(s)} \in 1 : N$ be the index of the sender of the edge and $e_j^{(r)} \in 1 : N$ be the index of the recipient.

One other important point is that most measures of network hierarchy are meant to be applied to networks where the edge sets capture relations other than "has power over". In this special case, it is theoretically easier to construct a measure of network hierarchy since the network must be directed and acyclic (preventing circular chains of command). However, obtaining such information is usually impossible in most cases (with military personell networks being an obvious exception). Furthermore, if the researcher has collected such an edge-set, then the need for summary measures of the "hierarchical-ness" of the network is likely obviated, as deeper insights could be gained from applying inferential network analysis tools to the raw network. Therefore, we focus or attention on the measurement of hierarchy on networks where edges are not explicitly power relations.

3.1. Analytical Measures of Hierarchy

The most basic measures of hierarchy or differential importance of a nodes in a network can all be derived from basic node-level measures of network centrality [5]. To aggregate from the node-level measures up to a single measure on a network, one can calculate C, the centralization of the network. Centralization captures the degree of inequality in the distribution of a given centrality measure over the network. In general, we should then expect that networks that are more centralized are also likely to be more hierarchical. However, this measure attains its maximal value for any centrality measure when one node has a maximal value of the given centrality measure and all the rest of the nodes have the minimal possible value. This will tend to give star networks maximal centralization scores. This implicit assumption in measurement is important to consider when evaluating the validity of centralization based measures of hierarchy. The four centrality measures we consider are: degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality.

The **degree centrality** of node v_i is simply the number of outgoing or incoming edges incident to it. Formally, this can be calculated as:

$$\text{in-degree centrality}_i = \text{In}_i = \sum_{j=1}^{M} \mathbb{1}(e_j^{(r)} = i) \tag{1}$$

$$\text{out-degree centrality}_i = \text{Out}_i = \sum_{j=1}^{M} \mathbb{1}(e_j^{(s)} = i) \tag{2}$$

Degree centrality captures the number of friends or connections a node has, and intuitively, we should expect more powerful nodes will have more incoming and outgoing connections. However, degree centrality does not account for the identity of a node's alters. This can lead to difficulties when it is used to asses a node's position in a social hierarchy. For example, in a large company, we might expect that an administrative assistant may have a higher degree centrality than the CEO of a company if the edges being measured are work-related interaction. Furthermore, if there are many administrative assistants at the bottom of the power structure in an organization we might qualitatively consider to be extremely hierarchically structured, the degree centralization of the interaction network might be lower than in a comparably sized "organizationally flat" organization where a handfull of people serve as coordinators. Thus we must take great care in interpreting this statistic, depending on the type of edges it is defined over. For given node-level in-degree or out-degree centrality measures $\mathbf{c} = \{c_i\}_{i=1}^N$, the corresponding in-degree or out-degree centralization measure is defined as:

$$\text{Degree Centralization} = \frac{\sum_{i=1}^N (\max\{c_i\} - c_i)}{(N-1)(N-2)} \tag{3}$$

where $(N-1)(N-2)$ normalizes the measure for the size of the network.

The **betweenness centrality** of node v_i is a measure of the amount of influence a node has on the information transversed through it [**?**]. Define $D_{i,j}$ as the number of shortest paths in G between v_i and v_j, and $D_{i,j}(k)$ as the number of these shortest paths that pass through v_k, then the betweenness centrality of node k is:

$$\text{betweenness centrality}_i = \sum_{i \neq k \neq j} \left(\frac{D_{i,j}(k)}{D_{i,j}}\right) \tag{4}$$

Betweenness centrality captures how in-the-middle-of-things a node is and when edges involve sharing information, how important the node is to information flowing quickly across the network, Intuitively, we should expect the nodes with higher betweenness centrality will be more powerful, and that greater inequality on this measure should signal a greater degree of hierarchical structure in a network. For given node-level betweenness centrality measures $\mathbf{c} = \{c_i\}_{i=1}^N$, the betweenness centralization measure is defined as:

$$\text{Betweenness Centralization} = \frac{\sum_{i=1}^N (\max\{c_i\} - c_i)}{(N-1)^2(N-2)} \tag{5}$$

where $(N-1)^2(N-2)$ normalizes the measure for the size of the network. Similarly, the **closeness centrality** of node v_i is a measure of how few intermediate edges a given node must traverse to reach all other nodes in the network. Define $d(i, j)$ as the length of the shortest path between v_i and v_j, then closeness centrality of node i is:

$$\text{closeness centrality}_i = \sum_{i \neq j} \left(\frac{1}{d(i, j)}\right) \tag{6}$$

Again, intuitively, powerful members of a hierarchy will tend to be able to reach others in the network more easily, and inequality in this measure (as captured by the closeness centralization of the network) should theoretically signal the degree of hierarchy in the network. For given node-level closeness centrality measures $\mathbf{c} = \{c_i\}_{i=1}^N$, the closeness centralization measure is defined as:

$$\text{Closeness Centralization} = \frac{\sum_{i=1}^N (\max\{c_i\} - c_i)}{\frac{(N-1)(N-2)}{(2N-3)}} \tag{7}$$

where $\frac{(N-1)(N-2)}{(2N-3)}$ normalizes the measure for the size of the network.

The last of the classical centrality based measures is **eigenvector centrality**, which is meant to capture the degree to which a node is connected to other well connected nodes [**?**]. Let the adjacency matrix of the network be defined as $\mathbf{A}$ and the vector of local eigenvector centrality scores for each node be defined as $\mathbf{w} = \{w(v_1), ..., w(v_V)\}$. Then to calculate the eigenvalue centralities $\mathbf{w}$, one must solve the following eigenvector equation:

$$\mathbf{A}\mathbf{w} = \lambda\mathbf{w} \tag{8}$$

where λ is a vector of positive eigenvalues. The challenge is to find the *dominant eigenvector*, as only the largest eigenvalue results in the desired centrality measure for each node. Intuitively, higher eigenvector centrality should be associated with greater social or organizational importance. For given node-level eigenvector centrality measures $\mathbf{c} = \{c_i\}_{i=1}^N$, the eigenvector centralization measure is defined as:

$$\text{Eigenvector Centralization} = \frac{\sum_{i=1}^N (\max\{c_i\} - c_i)}{(N-1)} \tag{9}$$

where $(N-1)$ normalizes the measure for the size of the network. Interestingly, the eigenvector centralization of a network can still be large even if there are a relatively large proportion of higher degree nodes, as long as the edges are organized such that only a few nodes are connected to all of these higher degree nodes.

Landau's h and **Kendall's K** are two closely related measures of hierarchy that operate on a *dominance* network – a transformation of the weighted sociomatrix of a given network [**?**]. Intuitively, this *dominance* network is meant to capture dominance-subordination relationships (between animals). The $[i, j]$ entry of the corresponding sociomatrix is coded as 1 if i is dominant over j, 0.5 if i and j are equals, and 0 if j is dominant over i. If the underlying network does not capture dominance relationships, then an weighted network can be transformed into a *dominance* network by assigning a value of 1 in the $[i, j]$ entry of of the corresponding sociomatrix if the $[i, j]$ entry of the original sociomatrix is greater than the $[j, i]$ entry of the original sociomatrix, 0.5 in the $[i, j]$ entry of of the corresponding sociomatrix if the $[i, j]$ entry of the original sociomatrix is equal to the $[j, i]$ entry of the original sociomatrix, and zero otherwise.

Landau's h is used to compare a directed network to a perfect linear hierarchy (a strict dominance-ordering of nodes) [**?** 2]. This measure is defined as follows:

$$h = \frac{12}{N^3 - N} \sum_{i=1}^{N} \left[\text{Out}_i - \frac{N-1}{2} \right]^2 \tag{10}$$

where $h \in [0, 1]$. Note that Landau's h does not provide an individual level metric of importance or relative power. This measure was specifically designed to operate on networks of "has power over" edges, and thus may be difficult to interpret when the network is not explicitly defined on power relations. Because of a preference for chain-like structures, this measure will likely provide poor performance when edges measure social relations. **Kendall's K** is also designed to compare a directed network to a perfect linear hierarchy [3]. As noted in [2], it often gives an identical value to Landau's h, but is theoretically distinct. Begin by defining the number of *cyclic triads* (CyT) in the network as:

$$CyT = \frac{N(N-1)(2N-1)}{12} - \frac{1}{2}\sum \text{Out}_i^2 \tag{11}$$

Then we can define Kendal's K as

$$K = 1 - \frac{d}{d_{max}} \tag{12}$$

where the value d_{max} is defined as follows:

$$d_{max} = \begin{cases} \frac{1}{24}(N^3 - N) & \text{if } N \text{ is odd} \\ \frac{1}{24}(N^3 - 4N) & \text{if } N \text{ is even} \end{cases} \tag{13}$$

Kendal's $K \in [0, 1]$ is also only defined as a global measure, and no individual level analogue exists. It was also specifically designed to operate on networks of "has power over" edges, and is therefore likely a poor choice for networks defined over social interactions.

Triangle transitivity (t_{tri}**)** was proposed as an improvement over Landau's h and Kendal's K in [2]. This measure captures the degree to which triads in a network are not cyclic, with the intuition that more hierarchical networks will contain a lower proportion of cyclic triads relative to the total number of triads. To calculate this measure, we begin by calculating the proportion of triangles that are not cycles.

$$P_t = \frac{N_{transitive}}{N_{transitive} + N_{cycle}} \tag{14}$$

The authors show that in a random network $P_t = 0.75$ in expectation so they normalize their statistic as follows:

$$t_{tri} = 4(P_t - 0.75) \tag{15}$$

where $t_{tri} \in [-3, 1]$ can be negative if there is a particularly high proportion of cyclic triads in the network. This is likely a better measure of hierarchy than Landau's h and Kendal's K when the network is weighted or ties do not represent dominance relationships. However, if there are violations of a linear hierarchy that involve larger cycles (spanning more than three nodes) this measure will fail to pick them up, so care should be used in interpreting this measure for complex networks.

M-reach degree was developed to identify 'key' players in a network [**?**]. It is defined as a measure for each node v_i as the number of alters that are reachable from v_i. If G is directed then the reachable alters must lie along an outgoing path from v_i. A closely related measure, **M-reach closeness** is defined as the M-reach degree of a node, but with the contribution of each alter j that is reachable, weighted by the inverse of the shortest path length between i and j. Both of these measures are only defined for individual nodes, so to calculate a global measure for a network, we take the Gini coefficient of the measures calculated for each node i. The Gini coefficient [**?**] is a measure of inequality originally developed to measure income inequality in a society. For a vector of values $x : \{x_i\}_{i=1}^N$ the Gini coefficient G of that vector is:

$$G = \frac{\sum_i \sum_j |x_i - x_j|}{2 \sum_i \sum_j x_i} \tag{16}$$

In words, it is half of the *average absolute difference* of all pairs of entries in the vector, divided by the average, which acts as a normalizing constant. We take the Gini coefficient of the M-reach degree and M-reach closeness values for each node in the network and treat this as our global measure of network hierarchy. Intuitively, this measure is capturing the degree of inequality in access to other nodes in the network. We should expect that a higher degree of inequality in this metric would be associated with a more hierarchical network. One potential shortcoming of these measures is that they only account for out-degree, when incoming ties may be a better signal of power relations in many real-world social networks.

Global reaching centrality (GRC) is a generalization of M-reach degree centrality measure [1], and was designed specifically to measure on any network. When the network is unweighted and directed, let $C_R(i)$ be the local M-reach degree centrality of node i, then the global reaching centrality of the network can be defined as:

$$GRC = \frac{\sum_{i=1}^N [\max(C_R) - C_R(i)]}{N - 1} \tag{17}$$

When the graph is weighted and directed, the authors in [1] propose an alternative formulation of $C_R(i)$:

$$C_R^{'}(i) = \frac{1}{N-1} \sum_{j: 0 < d^{out}_{(i,j)} < \infty} \left(\frac{\sum_{k=1}^{d^{out}(i,j)} w_i^{(k)}(j)}{d^{out}(i,j)} \right) \tag{18}$$

where $d^{out}_{i,j}$ is the (directed) path length from node i to node j, and $w_i^{(k)}$ is the weight of the kth edge along this path. This alternative formulation can then be plugged into equation 17 to calculate the GRC for a weighted,directed network. Finally, when the network is unweighted and undirected, the authors in [1] propose an additional alternative formulation of $C_R(i)$:

$$C''_R(i) = \frac{1}{N-1} \sum_{j:0<d(i,j)<\infty} \frac{1}{d(i,j)} \tag{19}$$

where $d_{i,j}$ is the (undirected) path length from node i to node j. The intuition of for the interpretation of GRC as a measure of network hierarchy is almost identical to the interpretation of the Gini coefficient of the M-reach degree and M-reach closeness values for a network, with only difference being the way the individual measures are aggregated, and its extension to weighted and undirected networks. Again, a potential major shortcoming is that this measure only makes theoretical sense when the outgoing edges in the network signal a power relation where the sender has power over the recipient.

rooted depth [4] is only defined for networks where a root (a node that has only incoming edges) exists. Let N_r be the number of node-root pairs in the network. Then the rooted depth of the network can be defined as:

$$D = \frac{1}{N_r} \sum_{i=1}^{N_r} l_{ri} \tag{20}$$

where l is the length of the shortest path between root r and node i. This measure can only be calculated for the network as a whole. However, given R roots in a network, we can calculate a local root depth for each node that is equal to the average length of the shortest path between itself and all roots. One of the major problems with this measure is that it is undefined for networks without a root (a node with only incoming edges). This makes it very difficult to apply to most networks (we were only able to calculate it for a small fraction of networks in our sample). Furthermore, since this measure relies on finding nodes with no outgoing ties, it can be very sensitive to the way the network is measured. In general, we do not find rooted depth to be a useful measure in our empirical analysis.

The final analytical measure we consider is the **Krackhardt** hierarchy score of a network. [6]. The Krackhardt hierarchy score is only defined for directed networks, as the proportion of un-reciprocated dyads in the reachability matrix of a network. More formally let R be the reachability matrix of G, where R is an $N \times N$ matrix with a one in the $[i, j]$ entry if node j can be reached via an outgoing path from node i. Then the Krackhardt hierarchy score can be defined as:

$$\text{Krackhardt} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{1}(R_{i,j} == 1 \& R_{j,i} == 0)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{1}(R_{i,j} == 1)} \tag{21}$$

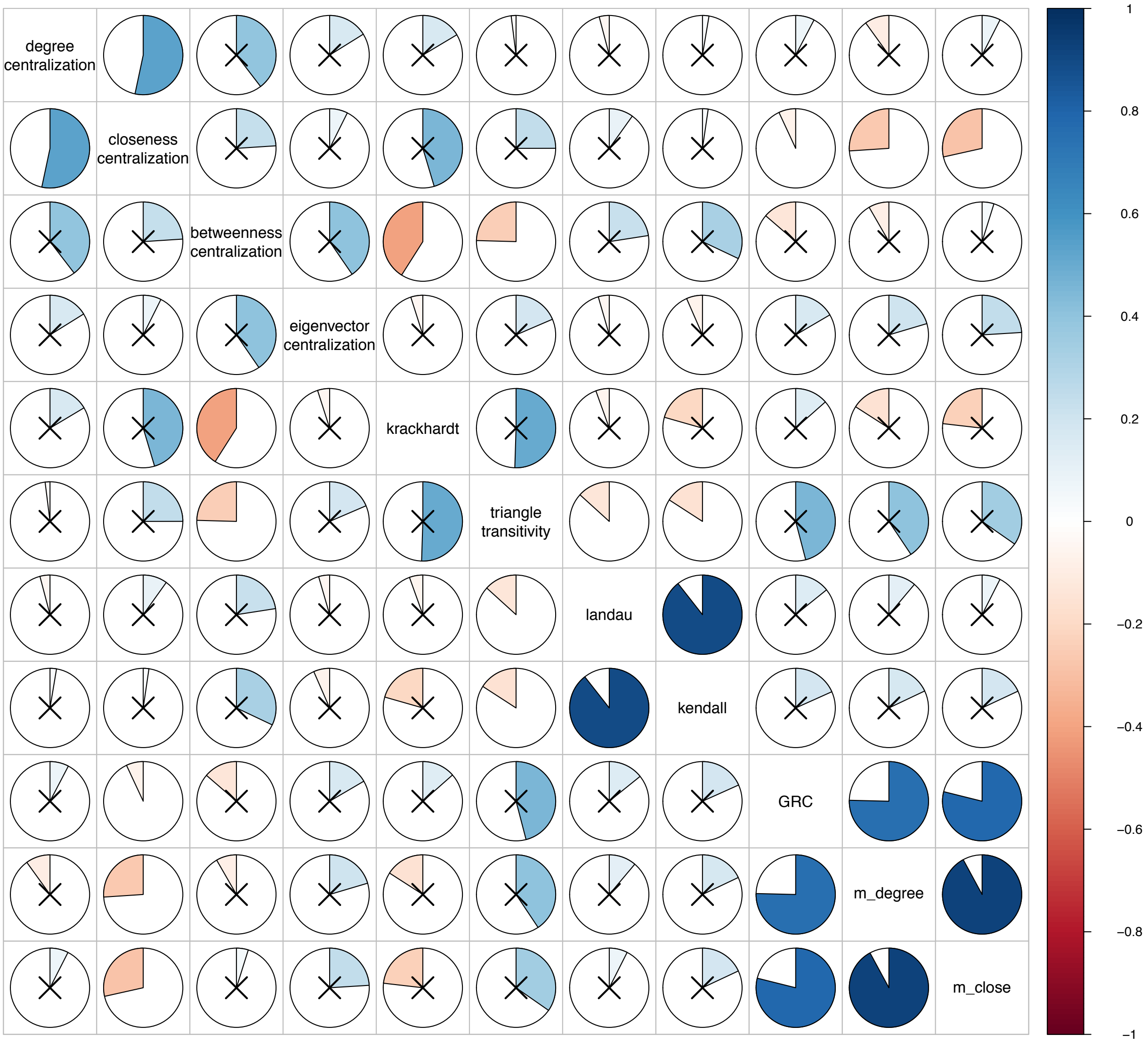
This measure takes a value of one for all directed acyclic graphs, and zero for cycles and cliques. This is a theoretically attractive measure if ties indicate a power relation, but may also be useful for analyzing communication networks, as we might expect that those higher up in the organizational hierarchy will be able to reach a larger part of the network with outgoing messages.

3.2. *Relationships Between Hierarchy Measures*

To begin to understand how these measures are related, we calculate each of these twelve measures on a total of 136 social, organizational, and information networks described in greater detail in Section 4. All of these networks are directed, and for almost all of them, we were not able to calculate the rooted depth of the network, and we therefore exclude it from our further analysis[2]. Figure 1 illustrates the correlation coefficients between the eleven hierarchy measures we were able to calculate on these networks. As we can see, Landau's h and Kendal's K are highly correlated, as are M-reach degree and M-reach closeness Gini coefficients. This makes these two pairs of measures largely redundant, but we include them in the remainder of our analysis for completeness. Landau's h and Kendal's K are also significantly correlated with the simple degree centralization of the network, and M-reach degree and M-reach closeness Gini coefficients are also significantly correlated with GRC (by construction). Also of interest, Krackhardt's hierarchy measure and triangle transitivity are both significantly negatively correlated with betweenness centralization.

[2]This is because the Rooted Depth measure can only be calculated for networks where at-least one node only has incoming ties.

Figure 1. Pearson correlation coefficients between nine measures of network hierarchy. A black **X** in a cell indicates that there was not a significant correlation between measures at the $\alpha = 0.05$ level of significance. Note that the correlation between Landau's h and Kendal's K was ≈ 0.9996 but was not exactly 1.



Perhaps surprisingly, most of the rest of the correlations depicted in Figure 1 are relatively small and statistically insignificant, indicating that if all of these measures are valid, they are likely capturing different dimensions of hierarchy in a network. This finding suggests that either there are multiple dimensions to hierarchy in a network [7], some of these measures do not measure hierarchy, or some combination of both. To investigate this finding further, we perform several statistical and qualitative comparisons between these measures in Section 5.

4. Data

The data used in this study comprise 136 social networks, collected from three primary sources. The first of these is a set of seventeen email communication networks among department managers in North Carolina county governments. These data were collected as part of a field experiment described in Ben Aaron et al. (forthcoming), and comprise all department manager to department manager email communications over a three month period in 2013. In total, these networks include 17,863 emails sent between 362 department managers. The second set of networks record cosponsorship patterns in the United States Senate between 1973 and 2009. There are a total of eighteen networks, each recording the number of times Senator *i* cosponsors (officially records support for) a piece of legislation sponsored by (introduced by) legislator *j* during a two year session of Congress. The third primary source from which we obtain network data is the UCI-Net online network data repository[3]. We were able to obtain a total of 101 network datasets from this website, comprising a wide range of social, communication, and economic networks. Links to source data, and references for all networks used in this study will be made available in an appendix[4].

Theoretically, we might expect different types of networks (social, information, biological, etc.) to display distinct relational structures, and perhaps exhibit common hierarchical structures. We therefore decided to hand code each of the 136 networks in our sample into one of eight broad categories. Descriptive statistics for networks in each of these categories are provided in Table 1. We can see that most networks have between twenty five and forty nodes on average, with the notable exception being the cosponsorship networks, which average roughly one hundred nodes. Note that the networks classified as *unknown* have yet to be categorized due to the difficulty in navigating the UCI-Net data archives, but will be classified into one of the eight categories before publication.

Table 1. Network Descriptive statistics for all 136 networks in our sample, aggregated by the network type. All columns are averages over networks of that type.

Type	# of Networks	Nodes	Edges	Density	Clustering Coefficient
communication	20	25.40	1764.25	2.87	0.55
cosponsorship	18	101.22	13358.89	1.32	0.79
co-membership	2	21.50	22.50	0.05	0.07
interaction	40	23.07	1944.95	1.99	0.59
unknown	33	38.70	445.36	0.38	0.43
friendship	6	29.83	92.00	0.12	0.35
affect	11	17.64	95.18	0.32	0.33
terrorism	1	63.00	308.00	0.08	0.36
trade	5	24.00	285.60	0.52	0.73

5. Analysis

5.1. Principal Components Analysis

In order to find patterns in the various measures of hierarchy presented above, we rely on Principal Components Analysis (PCA). PCA is a dimension reduction technique commonly used in the social sciences. This works by

[3]Data are freely available online here: https://sites.google.com/site/ucinetsoftware/datasets

[4]We did not had time to compile this appendix before the end of the semester, but it will be included in the fina version of the paper.

finding the eigenvalues and eigenvectors of a set of variables, such that the components identified by the model maximize the variance accounted for. For the purposes of this analysis, we have 11 hierarchy measures calculated across the 136 networks described above. PCA allows us to detect the patterns behind this 11-dimension data, and constructs principal components such that most of the variance in the 11 dimension problem can be accounted by a fewer number of components.

Of course, there will be as many principal components as there are dimensions in the original data. PCA component eigenvalues are illustrated in Figure 2. As is clear from this figure, while there are nominally 11 principal components, a vast majority of the variance present in the original problem is captured by the first four principal components. This four dimensional problem is much easier to interpret than the original 11D problem. A graphical comparison of components one and two is provided in Figure 3. A graphical comparison of components one and three is provided in Figure 4. A graphical comparison of components two and three is provided in Figure 6. A graphical comparison of components one and four, two and four, and three and four are included in Figure 5, Figure 7, and Figure 8, respectively.

Table 2. PCA Rotation for Components with Eigenvalues above 1

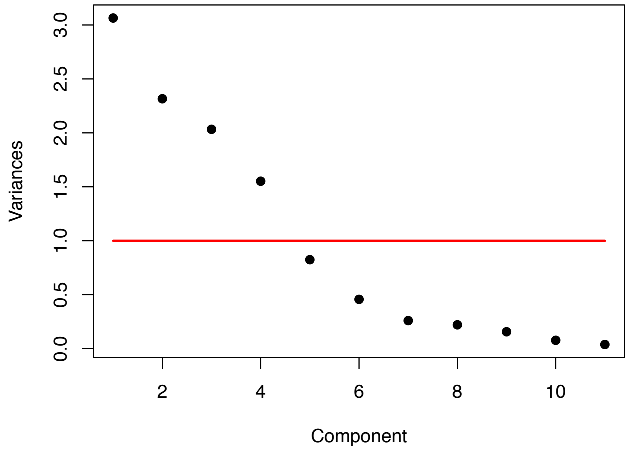
	P.C. 1	P.C. 2	P.C. 3	P.C. 4
degree centralization	-0.017	0.062	-0.522	0.222
closeness centralization	-0.130	-0.056	-0.606	-0.107
betweenness centralization	-0.009	0.437	-0.292	0.387
eigenvector centralization	0.166	0.040	-0.239	0.466
Landau's h	0.130	0.454	-0.126	-0.481
Kendall's K	0.171	0.506	-0.098	-0.395
GRC	0.501	-0.109	-0.085	-0.088
m-closeness	0.539	0.011	0.045	0.151
m-degree	0.538	-0.036	0.087	0.044
Krackhardt	-0.055	-0.408	-0.346	-0.370
Triangle Transitivity	0.273	-0.397	-0.237	-0.127

Table 2 presents the rotations, or factor loadings, which describe to what extent the original variable relates to the principal components. The first principal component, which accounts for XX% of the variance, is most highly correlated with M–Reach Degree, M–Reach Closeness, and GRC. It is not surprising that these three measures load together, as GRC relies on M–Reach Closeness, which in turn relies on M–Reach Degree. As discussed in the description of the measures, this family of hierarchy measures describes how many nodes in a dominance network can be reached from a given node. In other words, these measures describe a type of hierarchy in which diffusion of authority is recognized by higher values. This corresponds closely to Mann's "diffusion" type described in the introduction; a type of power that varies according to how many subordinates can be immediately reached by a dominant actor.

Second, the third principal component is heavily loaded on degree and closeness centrality. As discussed in the analytical measures of hierarchy section, the former measures the number of connections for a given node, while the latter measures how easy it is to reach each node in the network from a given node. Returning to Mann's typology, this seems to measure a distinct concept from his two dimensions of hierarchy: diffusion (described by the first principal component) and authoritative. Instead, it seems to measure the extensiveness of a hierarchy. In other words, we could expect that networks which have a higher value on principal component 3 to have a deeper and more extensive hierarchical system, irrespective of the type of hierarchy they represent.

Finally, principal components 2 and 4 eschew a clear explanation and mapping on to Mann's typology. Components 2 and 4 are similar in that landau and kendall load about evenly between the two. Component 2 is more heavily loaded on Krackhardt and triangle transitivity, whereas component 4 is loaded heavily on Eigenvector centrality. Landau's h and Kendall's K are indicative of Mann's authoritative type of power, in that each measures the treeness of a network, with the highest value representing a perfect linear hierarchy. In social terms, this could be understood as a military type of hierarchy, where there are clear and rigid layers of authority.

Figure 2. Eigenvalues for 9 largest principle components in our analysis indicate that we should examine the first three components, which all have eigenvalues greater than one.



Component 2 loads more heavily on Krackhardt's hierarchy score and Triangle Transitivity. The former measures the extent that dyads are not reciprocal, while the latter has a very substantively similar interpretation to Landau's h and Kendall's K, because it was intended to be an improvement on those two measures. The concept of unreciprocated dyadic relationships matches closely onto Mann's conception of authoritative hierarchies, as these are clear dominant–subordinate relationships. Therefore, we argue that component 2 most closely maps to the authoritative type of hierarchy discussed by Mann.

Component 4 differs from Component 2 in that it loads heavily on eigenvector centralization, rather than triangle transitivity or Krackhardt's hierarchy score. Eigenvector centralization describes how well a node is connected to other well–connected nodes. A network with a high degree of eigenvector centralization is one in which a small coterie of well–connected nodes, exhibit control over a system. This is a slight deviation from the authoritative power type described by Mann. To put this into social terms, while Component 4 would describe a military system, Component 2 might describe an oligarchical system, which is comprised of both explicit dominant–subordinate relationships (captured by Kendall's K and Landau's h), but a group of individuals at the top, rather than a perfect tree which extends from the lowest rung to the highest rung.

5.2. Ground Truth Rankings

	Average Rank
Degree Centrality	0.75
Closeness Centrality	0.82
Betweenness Centrality	0.78
Eigenvector Centrality	0.78
m-degree	0.81
m-close	0.63
GRC	0.36
D-root	0.86

6. Conclusions

test

Figure 3. Principle components plot for components one and two.

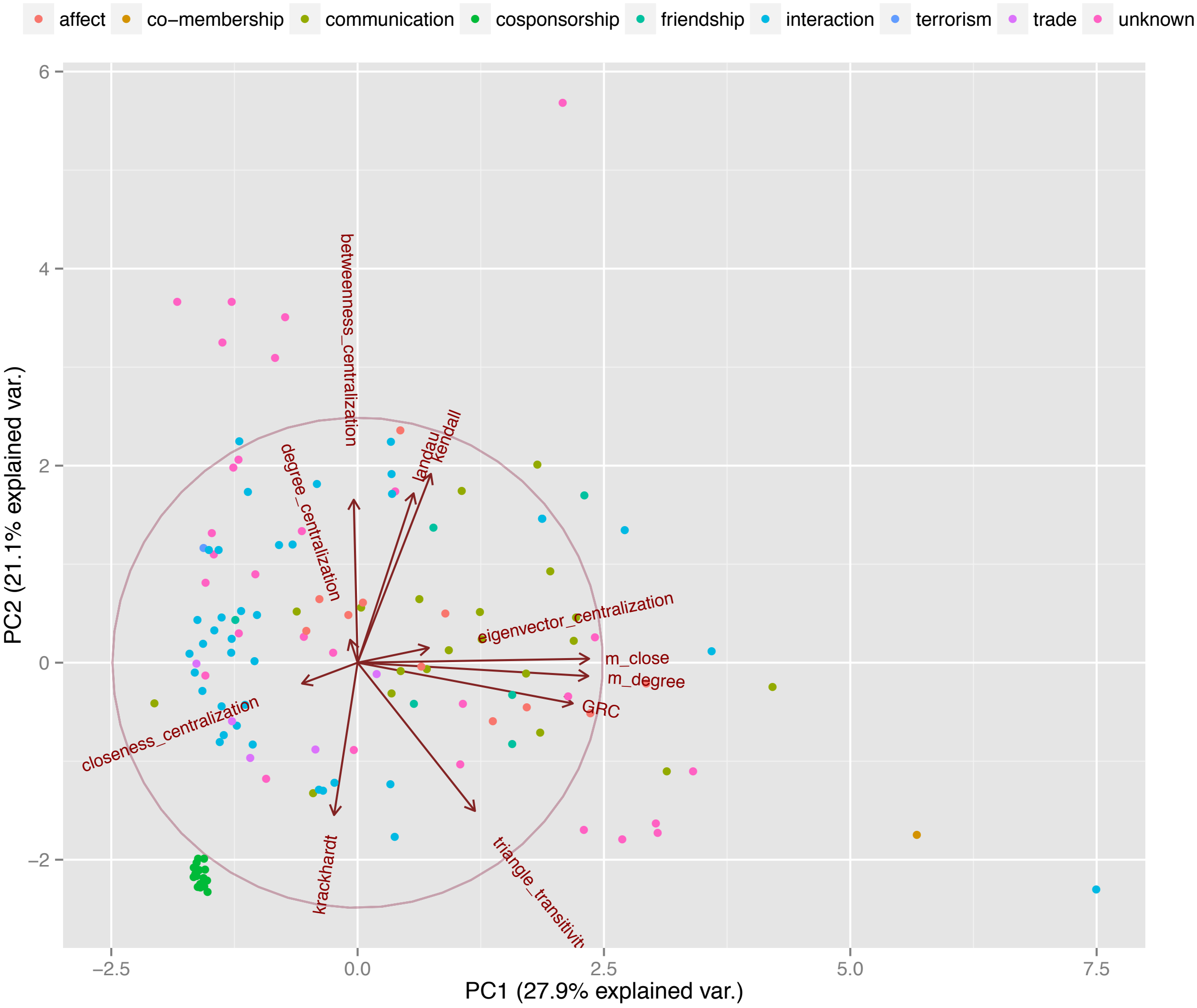


Figure 4. Principle components plot for components one and three.

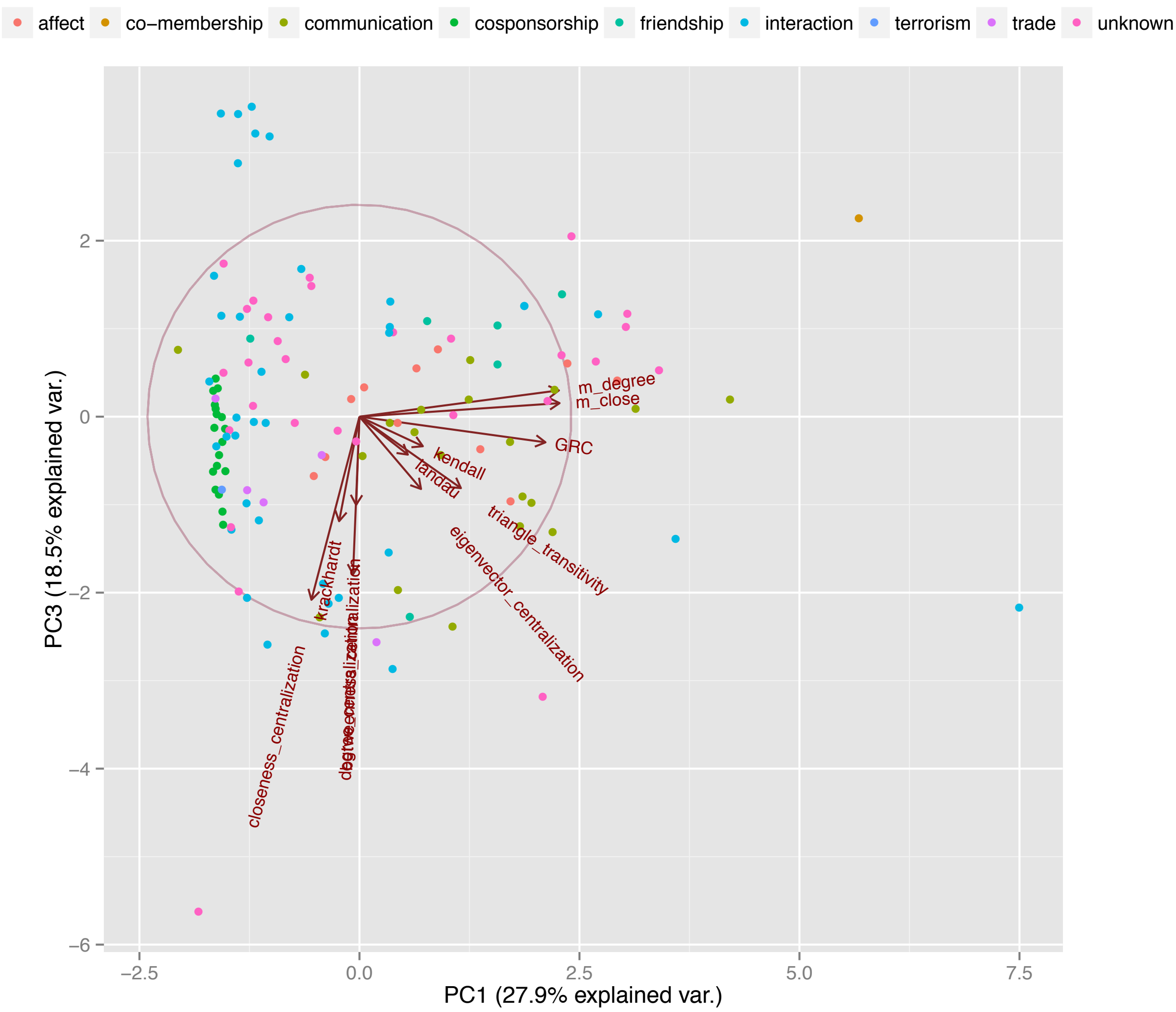


Figure 5. Principle components plot for components one and four.

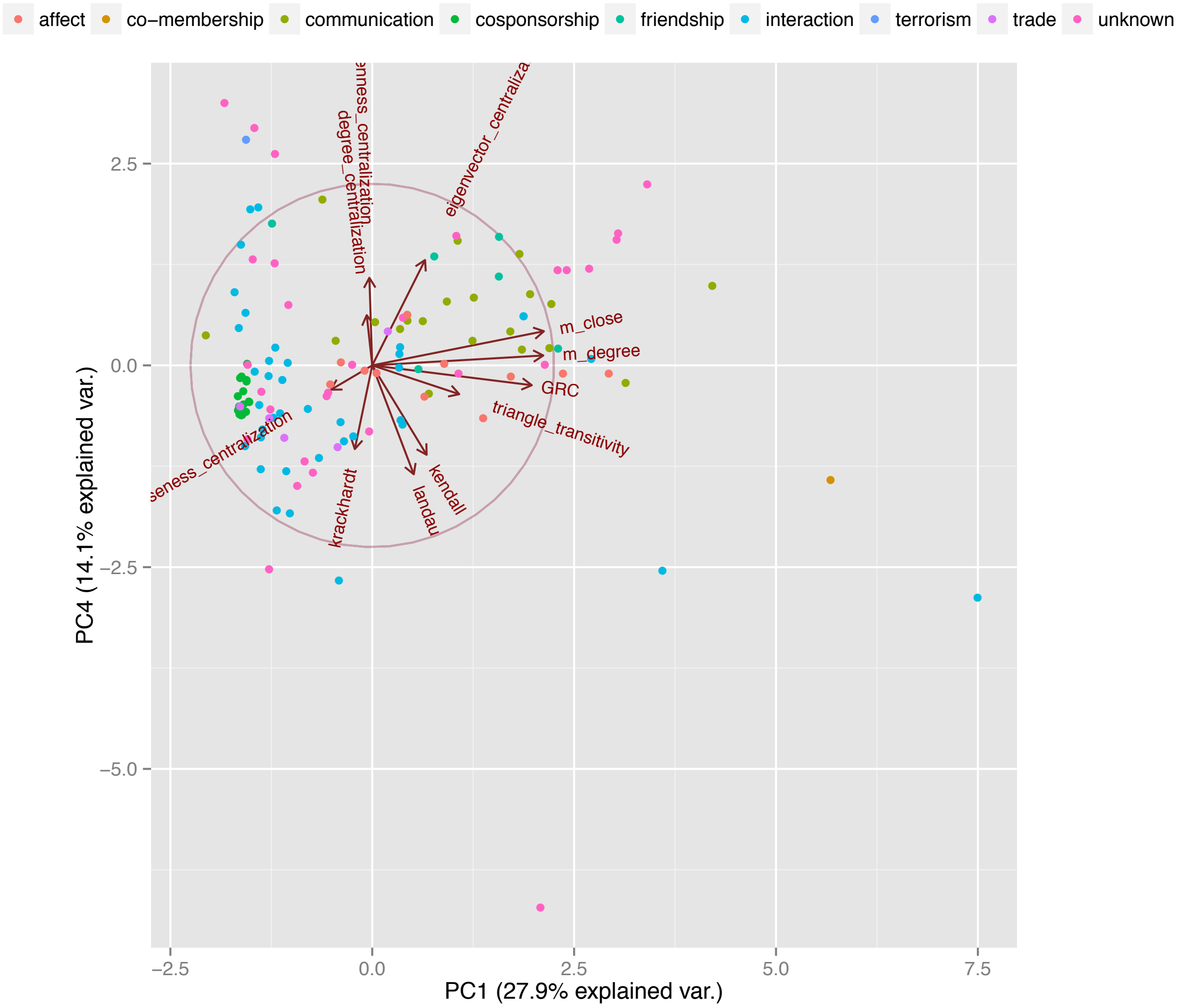


Figure 6. Principle components plot for components two and three.

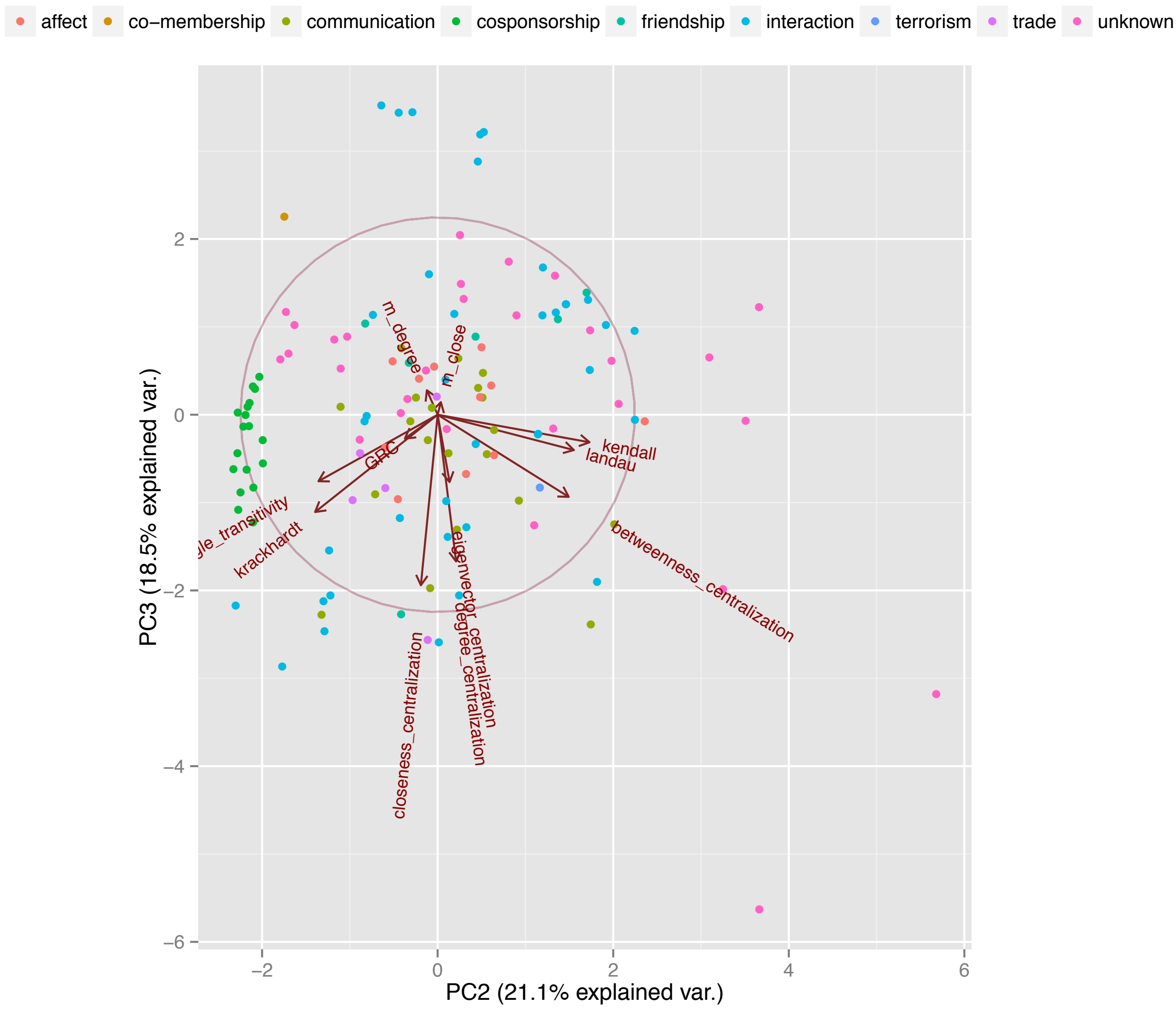


Figure 7. Principle components plot for components two and four.

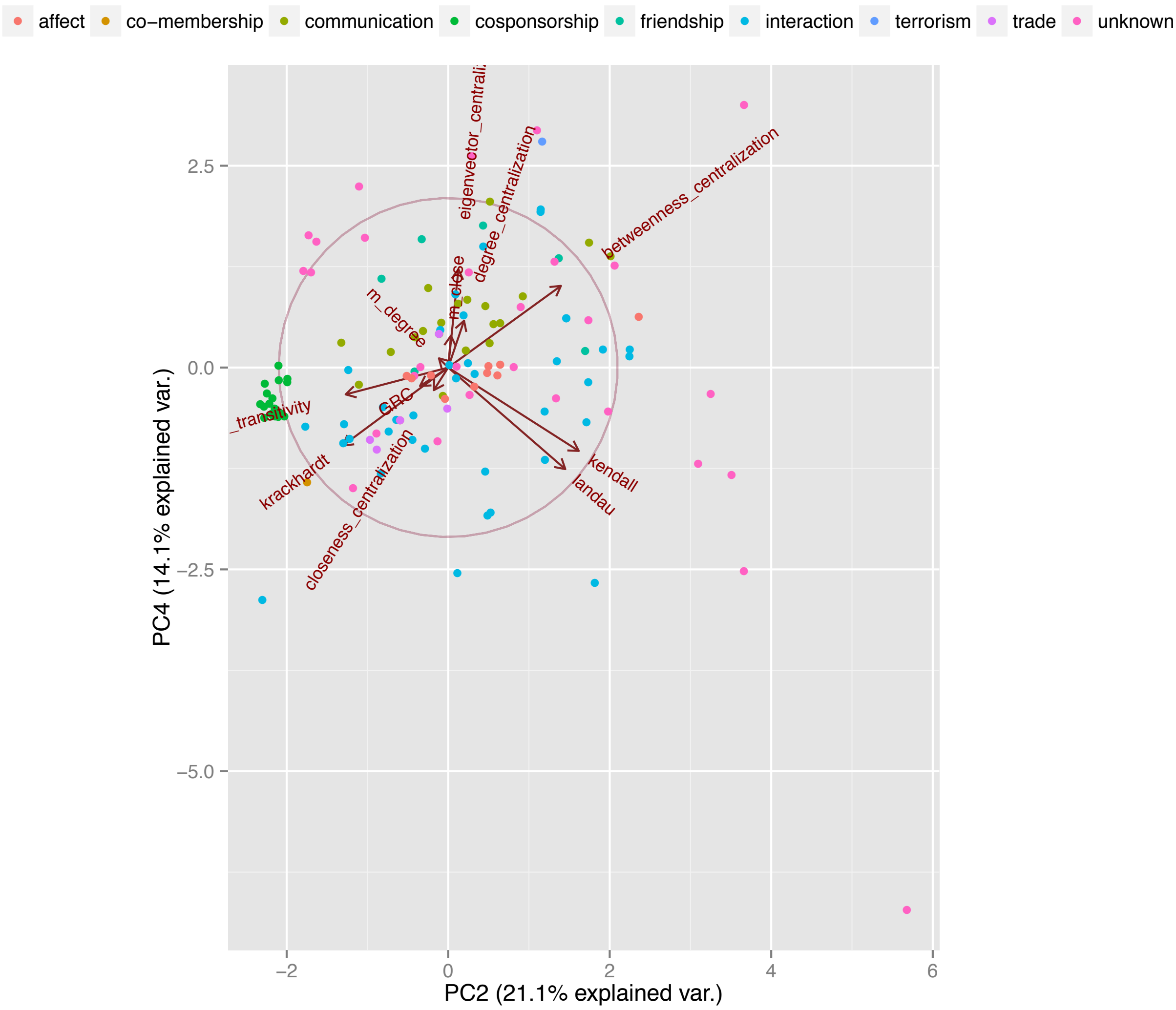
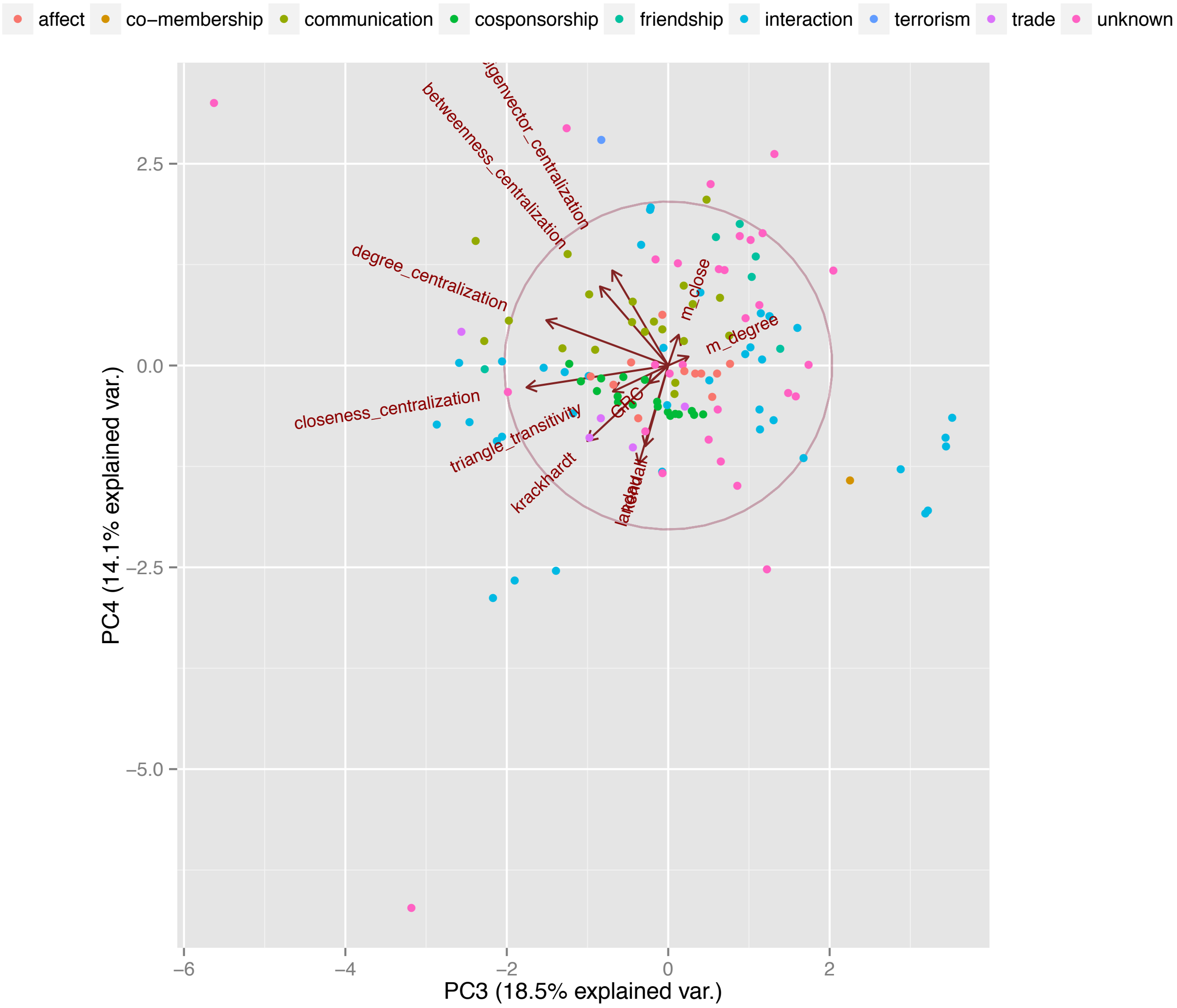


Figure 8. Principle components plot for components three and four.



References

[1] E. Mones, L. Vicsek, T. Vicsek, Hierarchy measure for complex networks., PloS one 7 (3) (2012) e33799. doi:10.1371/journal.pone.0033799.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3314676{&}tool=pmcentrez{&}rendertype=abstract

[2] D. Shizuka, D. B. McDonald, A social network perspective on measurements of dominance hierarchies, Animal Behaviour 83 (4) (2012) 925–934. doi:10.1016/j.anbehav.2012.01.011.
URL http://dx.doi.org/10.1016/j.anbehav.2012.01.011

[3] M. G. Kendall, B. B. Smith, On the Method of Paired Comparisons, Biometrika 31 (3/4) (1940) 324–345.
URL http://www.jstor.org/stable/2332613

[4] K. Suchecki, J. Holyst, Hierarchy depth in directed networks, arXiv preprint arXiv:1303.2085arXiv:arXiv:1303.2085v1.
URL http://arxiv.org/abs/1303.2085

[5] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, Vol. 8 of Structural analysis in the social sciences, 8, Cambridge University Press, 1994. doi:10.2307/2077235.
URL http://www.amazon.com/dp/0521387078

[6] D. Krackhardt, Graph theoretical dimensions of informal organizations (1994).
URL http://books.google.ca/books?hl=en{&}lr={&}id=JCSoVyl9HIUC{&}oi=fnd{&}pg=PA89{&}dq=Krackhardt,+D.++hierarchy{&}ots=1hO9KRfrY{_}{&}sig=6bPSWWnxyQqGjPGJL97Zqxu-d7Y

[7] B. Corominas-Murtra, J. Goñi, R. V. Solé, C. Rodríguez-Caso, On the origins of hierarchy in complex networks., Proceedings of the National Academy of Sciences of the United States of America 110 (33) (2013) 13316–21. doi:10.1073/pnas.1300832110.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3746874{&}tool=pmcentrez{&}rendertype=abstract