

Hierarchical Structure in Social Networks

Cynthia Cook^{1c}, Matthew J. Denny^d, Mitchell Goist¹, Timmy Huynh¹

^a Department of Statistics cmc496@psu.edu

^b Department of Political Science, mzd5530@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

Hierarchies are important. What we did and why...what we found...

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,” [21] 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.

Having developed a framework for measuring hierarchy in networks, we will then test its internal and external validity. To test the internal validity of such a measure, we will conduct a simulation study. To test the external validity of our measure we will compare measurements across networks that we can theoretically rank by their degree of hierarchy. We will then apply our framework to better understand the implications of hierarchical network structure for organizational performance in a sample of 17 county government organizations.

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

1.1. Problem Statement

How do we define and measure hierarchy in (directed) social networks? We need to relate sociological conceptions of hierarchy and power to network measures. In developing this framework we intend to compare analytical and statistical approaches to measurement on both synthetic and real world datasets. In particular there are several key questions we must address.

1. Is an analytical or statistical measure of network hierarchy more appropriate for our goals?
2. Can we capture all or even most salient dimensions of hierarchy as defined in the sociological literature in a single measure?
3. Can our measure be extended to undirected networks?

1.2. Significance

All of our fields including political science, sociology, and statistics approach the concept of hierarchy from different angles. Yet the main goal of any researcher is the same: to accurately theoretically understand and quantify real world phenomena. 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. Measuring Hierarchy

We will denote a network's graph as $G = (V, E)$, where $V = \{v_i\} = v_1, \dots, v_N$ the set of vertices and the set of edges $E = \{e_j\} = e_1, \dots, e_M$.

2.1. Measures of Hierarchy-Global

1. Landau's $h \in [0, 1]$ is used to compare a directed network to a perfect linear hierarchy [15]. Let S_i be the row sum for each node, then:

$$h = \frac{12}{N^3 - N} \sum_{i=1}^N [S_i - \frac{N-1}{2}]^2$$

2. Kendall's $K \in [0, 1]$ like Landau's h was designed to compare a directed network to a perfect linear hierarchy [14]. Let d be the number of cyclic triads defined as: $d = \frac{N(N-1)(2N-1)}{12} - \frac{1}{2} \sum S_i^2$. Then:

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

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

3. Global reaching centrality, based on the m -reach centrality measure is adapted in [25] to be a simple measure of hierarchy for any graph:

$$GRC = \frac{\sum_{i \in V} [C_R^{max} - C_R(i)]}{N - 1},$$

- (a) When the graph is unweighted and directed, the $C_R(i)$ is the local reaching centrality defined as the proportion of all nodes in G that can be reached along outgoing edges from node i .
- (b) When the graph is weighted and directed, the following version for the reaching centrality as defined in [25] is used:

$$C'_R(i) = \frac{1}{N - 1} \sum_{j: 0 < d_{i,j}^{out} < \infty} \left(\frac{\sum_{k=1}^{d_{i,j}^{out}} w_i^{(k)}(j)}{d_{i,j}^{out}} \right),$$

where $d_{i,j}^{out}$ is the length of the directed outgoing path from v_i to v_j and $w_i^{(k)}$ is the weight of the k th edge along this path.

- (c) When the graph is unweighted and undirected, the following version for the reaching centrality as defined by [25] is used:

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

4. Rooted depth in directed networks is defined in [31], where a root is a node that has only incoming edges. Let N_r be the number of node-root pairs in the network. Then the global root depth is defined as:

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

where l is the length of the shortest path between root r and node i .

5. We apply centralization, a method used to calculate a global measure of centrality from local measures, to the following local centrality measures: betweenness, closeness, eigenvector, and degree, all of which are defined in the following section. Centralization is defined as follows for local measures c_i :

$$C = \sum_{i=1}^N (\max\{c_i\} - c_i)$$

2.2. Measures of Hierarchy-Local

1. M-reach degree was developed to identify ‘key’ players in a network [1] and is defined for a vertex v_i as the number of reachable vertices from v_i . If G is directed then this reachable vertices must lie along an outgoing path from v_i . This local measure is the same used for the *GRC* local measures in [25].
2. M-reach closeness was also developed to identify ‘key’ players [1]. It is defined as the M-reach degree weighted by the inverse geodistances.
3. A local version of rooted depth can be found for a graph. Given r roots in a network, the each node has a local root depth equal to the average length of the shortest path between itself and all roots [31].
4. Betweenness centrality is meant to capture the amount of influence a node has on the information transversed through it [8]. Define $D_{i,j}$ as the number of shortest paths 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 for v_k is:

$$\sum_{i \neq k \neq j} \left(\frac{D_{i,j}(k)}{D_{i,j}} \right)$$

5. Given $d(i, j)$ is the length of the shortest path between v_i and v_j , closeness centrality is measured as

$$\sum_{i \neq j} \left(\frac{1}{d(i, j)} \right)$$

6. Eigenvector centrality is meant to give weight to nodes that are connected to nodes that are most central in the network [3]. Defining the graphs adjacency matrix as A and the vector of local eigenvector centrality scores as $c = \{c(v_1), \dots, c(v_V)\}$ we have to solve the following equation:

$$Ac = \lambda c$$

7. Degree centrality for v_i is simply the number of outgoing edges it has.

2.3. Discussion of Measures

For the global measures, both Landau’s h and Kendall’s K have mostly been used to determine so called pecking order among animals [2]. Due to their ability to capture the degree to which a network is linearly hierarchical, we expect these measures to perform similarly. We also expect all three centralization scores to perform the same. We

note that rooted depth is only defined for graphs with at least one root. If there is no root the measure classifies the network as having no hierarchy. For the local measures, we expect M-reach degree and degree to be correlated since M-reach degree was defined based on degree centralization. Similarly we expect M-reach closeness to be correlated to closeness.

3. A Model of Network Hierarchy

1. Control Centrality in a weighted and directed graph, defined by [20], identifies the minimum number of nodes need to drive an entire network to a given final state. Consider a directed, weighted network:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$$

which is the state of each node at time t , and also denoted as (\mathbf{A}, \mathbf{B}) . The components of this controllability matrix are: $\mathbf{A} \in \mathbb{R}^{N \times N}$, where each element gives the strength that node j can affect node i ; and $\mathbf{B} \in \mathbb{R}^{N \times M}$, where each element is the strength between the input signal $u_j(t)$ and node i , and M contains independent signals imposed by an outside controller. Defining $\mathbf{C} = (\mathbf{A}, \mathbf{B})$, the control centr...(line truncated)...

$$C_c(i) \equiv \text{rank}_g(\mathbf{C}^i)$$

2. For social networks, [12] defines hierarchy $h(G) \in [0, 1]$ from inferred nodal rankings $r(v)$. Mathematically, it is defined as:

$$h(G) = 1 - \frac{1}{m} A(G),$$

where $A(G) = \sum_{(v_i, v_j) \in E} \max(r(v_i) - r(v_j) + 1, 0)$ is the total 'agony'. Since the rankings are not known, they are found by minimizing the total agony over all possible rankings r .

4. Data

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 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.

Among the network datasets we are exploring, they are already or mostly in usable format. As we're navigating through our theoretical conception of hierarchy in network, we are ruling out the use of the karate club, dolphin, football, etc. datasets because we want to be able to analyze datasets where the networks are more interesting or theoretically-relevant. In this way, there are a couple of systems that might be useful. The first is a network of cooperation among militant groups, which encompasses joint exercises, mergers, and splits among militant groups: <http://web.stanford.edu/group/mappingmilitants/cgi-bin/>. This may be interesting for us for a few reasons: (1) there is no de jure hierarchical structure (i.e., no formally-recognized chain of command or sovereignty); (2) militant groups face a classic collective action problem, and thus we can expect the dynamics Mann describes to hold; and (3) most theories of conflict would predict no hierarchy to occur in this system. An interesting system to compare this to would be military actions in Vietnam: <http://tinyurl.com/pwofooy>. The nodes here would be military units, and the edges are participation in the same battle. Of course, the main issue with this dataset is that it's undirected, which we've noted may be difficult to conceptualize within a hierarchy framework. We're focusing on conflict datasets because many of the theoretical definitions define hierarchy as essentially about outcomes i.e., the ability of particular nodes to control the actions and behaviors of subordinate nodes.

We may also use manager network data where each organization has a “county manager” who is theoretically in charge of the rest of the actors, providing an opportunity to determine if the methods we employ capture a plausible hierarchical structure. We are still exploring this and other datasets though.

1. Adolescent Health: survey asked students to list 5 male and female friends. [26]
2. Residence Hall: friendships between 217 students in Australian National University. [9]
3. Taro Exchange: gift–giving relationships between households in a Papaun village. [30]
4. Highschool: friendship relationship between boys at a small Indiana high school in 1957–1958. [7]
5. Dutch College: friendships between 32 university freshmen. [33]
6. Monks: preference ratings between monks in a cloister during a crisis. [4]
7. Physicians: innovation spread between 246 physicians in Illinois. [6]
8. Seventh graders: activity specific proximity rankings for 29 middle school students in Victoria [34].
9. Prosper loans: loans between users of prosper.com [?].
10. Libimseti.cz: likes between users on a Czech dataing site [5].
11. Friendster: friendship adds on the online site Friendster [10].
12. Digg: friendships on Digg [?].
13. Youtube: connections between Youtube users [24].
14. Epinions: who–trusts–whom between users of epinions [29].
15. EU emails: emails for 18 months from a major European research institution [16].
16. Facebook: friends lists from FACEBOOK, generated through a Facebook app survey [22].
17. Google Plus: friends between users who selected to “share circles” on Google Plus [22].
18. Linx kernel mailing list: communication network for the linux kernel mailing list, where each edge is a reply from a user to another [19].
19. Livejournal: map of an online community friendships of Livejournal users [17].
20. Manufacturing: communication network between employess of a mid–size manufacturing firm [23].
21. Pokec: Friendship networks in the Pokec online social network, popular in Slovakia [32].
22. Slashdot: tagging between users in slashdot for 2008 and 2009 [17].
23. Twitter: circles between twitter users [22].
24. UC Irvine: messages sent between students on an online community at UC Irvine [27].
25. U. Rovira i Virgili: email communication network from University Rovira i Virgili in Tarragona [11].
26. Wikipedia Talk: network of discussions between all users from the beginning of Wikipedia to January 2008 [18].
27. Wikipedia Votes: data from administrator elections [18].
28. Wikipedia Requests for Adminship: requests from 2003 through 2013 [35].
29. Friendster: network for online social site Friendster [10].

5. Analysis

The analytical portion of the problem will be conducted in R, which is known by all members of the group. We will be using both statistical and mathematical methods of quantifying and/or measuring hierarchy. We will focus on methods that have already been developed, published, and implemented in R, or are can easily be implemented by one of the group members. If time permits, we may try to develop or suggest directions for future development of our own statistical models and/or mathematical measurements. Each member of the group will be responsible for at least one method.

The statistical methods we will be looking into include hierarchical exponential graph models in the R package *hergm*. This package also includes hierarchical stochastic block models. Unlike fitting network data with exponential random graph models (ERGMs), hierarchical ERGMs focus on inducing local dependencies. Next, we will focus on latent space models, which can be fit in R using the *latentnet* package in the *statnet* suite of packages. For both the latent space and ERGM models, Bayesian inferential analysis can be conducted using the *Bergm*, *VBLPCM*, and *lvm4net* packages in R. We note that whenever fitting network data there is always the chance for computational timing and accuracy issues to come up. We have chosen a number of datasets for the purposes of capturing several types of hierarchies, but also so that we may have a few that are easily fit in R. Lastly, we will focus on mathematical measures of hierarchy. These measures primarily stem from graph theory, and can be easily programmed by ourselves in R. The measurements include the Global Reach Centrality (GRC), Triangle Transitivity, Kendall's K, and Landau's lambda.

6. Conclusions

References

- [1] An, W. (2015), "Multilevel Meta Network Analysis with Application to Studying Network Dynamics of Network Interventions." *Social Networks* 43: 48-56.
- [2] Bonacich, P. (1972), "Factoring and Weighting Approaches to Status Scores and Clique Identification." *Journal of Mathematical Sociology* 2: 113-120.
- [3] Breiger, R., Boorman, S., and Arabie, P. (1975), "An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling." *Journal of Mathematical Psychology* 12.
- [4] Brozovsky, L. and Petricek, V. (2007), "Recommender system for online dating service." *Proc. Znalosti* 29-40.
- [5] Coleman, J., Katz, E., and Menzel, H. (1957), "The diffusion of an innovation among physicians." *Sociometry* 253-270.
- [6] Coleman, J. (1973), "Introduction to mathematical sociology." *London Free Press Glencoe*.
- [7] Freeman, L. (1977), "A set of Measures of Centrality Based on Betweenness." *Sociometry* 40: 1, 35-41.
- [8] Freeman, L., Webster, C., and Kirke, D. (1998), "Exploring social structure using dynamic three-dimensional color images." *Social Networks* 20: 2, 109-118.
- [9] Friendster network dataset - KONECT, May 2015.
- [10] Guimera, R., Danon, L., Diaz-Guilera, A., Giral, F., and Arenas, A. (2003), "Self-similar community structure in a network of human interactions." *Phys. Rev. E* 68: 6.
- [11] Gupte, M., Shankar, P., Li, J., Muthukrishnan, S., and Iftode, L. (2011), "Finding Hierarchy in Directed Online Social Networks." *International World Wide Web Conference Committee (IW3C2)*.
- [12] Hogg, T. and Lerman, K. (2012), "Social dynamics of Digg." *EPJ Data Science* 1, 5.
- [13] Kendall, M. and Babington Smith B. (1940), "On the Method of Paired Comparisons." *Biometrika* 31: 324-345.
- [14] Landau, H. G. (1951), "On Dominance Relations and the Structure of Animal Societies: Effect of inherent characteristics." *Bulletin of Mathematical Biology* 13: 245-262.
- [15] Leskovec, J., Kleinberg, J., and Faloutsos, C. (2007), "Graph Evolution: Densification and Shrinking Diameters." *ACM Transactions on Knowledge Discovery from Data (ACM TKDD)* 1: 1.
- [16] Leskovec, J., Lang, K., Dasgupta, A., and Mahoney, M. (2009), "Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters." *Internet Mathematics* 6: 1, 29-123.
- [17] Leskovec, J., Huttenlocher, D., and Kleinberg, J. (2010), "Predicting Positive and Negative Links in Online Social Networks."
- [18] Linux kernel mailing list replies network dataset - KONECT, May 2015.
- [19] Liu, Y., Slotine, J., and Barabási, A. (2012), "Control centrality and hierarchical structure in complex networks." *Plos ONE*: 7: 9.
- [20] Mann, Michael. (1986), "Chapter 1: Societies as organized power networks." *The Sources of Social Power, Volume I, A history of power from the beginning to AD 1760*. Cambridge University Press: 1-33.
- [21] McAuley, J. and Leskovec, J. (2012), "Learning to Discover Social Circles in Ego Networks." *NIPS*.
- [22] Michalski, R., Palus, S., and Kazienko, P. (2011), "Matching organizational structure and social network extracted from email communication." *Lecture Notes in Business Information Processing*: 87, 196-206.
- [23] Mislove, A., Marcon, M., Gummadi, K., Druschel, P., and Bhattacharjee, B. (2007), "Measurement and analysis of online social networks." *Proc. Internet Measurement Conference*.
- [24] Mones, E., Vicsek, L., and Vicsek, T. (2012), "Hierarchy Measure for Complex Networks." *Plos ONE* 7: 3, 1-10.
- [25] Moody, J. (2001), "Peer influence groups: Identifying dense clusters in large networks." *Social Networks* 23: 4, 261-283.
- [26] Opsahl, T. and Panzarasa, P. (2009), "Clustering in weighted networks." *Social Networks* 31: 2, 155-163.
- [27] Prosper loans network dataset - KONECT, May 2015.
- [28] Richardson, M., Agrawal, R., and Domingos, P. (2003), "Trust Management for the Semantic Web." *ISWC*.
- [29] Schwimmer, E. (1973), "Exchange in the Social Structure of the Orokaiva: Traditional and Emergent Ideologies in the Northern District of Papua." *St. Martin's Press*.
- [30] Shizuka, D. and McDonald, D. B. (2012), "A social network perspective on measurements of dominance hierarchies." *Animal Behavior* 83: 925-934.
- [31] Suchecki, K. and Holyst, J. (2013), "Hierarchy depth in directed networks." *Physica A* arXiv:1303-2085.

- [32] Takac, L. and Zabovsky, M. (2012), “Data Analysis in Public Social Networks.” International Scientific Conference & International Workshop Present Day Trends of Innovations, Lomza, Poland.
- [33] Van de Bunt, G., Van Deuijn, M., and Snijders, T. (1999), “Friendship networks through time: An actor-oriented dynamic statistical network model.” *Computational and Mathematical Organization Theory* 5: 2, 167-192.
- [34] Watts, D. and Strogatz, S. (1998), “Collective dynamics of ‘small world’ networks.” *Nature* 393: 1, 440-442.
- [35] West, R., Paskov, H., Leskovec, J., and Potts, C. (2014), “Exploiting Social Network Structure for Person-to-Person Sentiment Analysis.” *Transactions of the Association for Computational Linguistics* 2: 297-310.