Available online at www.sciencedirect.com

ScienceDirect

Social Networks 00 (2015) 1–6

Social Networks

Hierarchical Structure in Social Networks

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

[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 imporantatskljdfl asdflks f.

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," [32] 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.

H. aving 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

2.1. *Measuring Hierarchy in Groups*

2.2. *Measuring Hierarchy in Networks*

3. A Model of Network Hierarchy

3.1. *Notation*

Given a graph $G = (V, E)$, it is comprised of a set of vertexes $V = \{v_i\} = v_1, ...v_N$ and edges $E = \{e_j\} = e_1, ...e_M$. We define the set of paths through the graph as $\{\pi_k\} = \pi_1, ...\pi_P$.
Need to explain the collapsed graph for 3-5 in below...
Need to define d in 6
need to know what m is in 7

3.2. *Measures of Hierarchy*

1. Landau's $h \in [0, 1]$ is used to compare a directed network to a perfect linear hierarchy in [1], where S_i is the row sum for each node also referred to as the dominance total:

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

2. Kendall's $K \in [0, 1]$ is also used in [1] to compare a directed network to a perfect linear hierarchy. Let d be the number of cyclic triad 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. Treeness $T \in [-1, 1]$ tries to capture how pyramidal the structure is and how unambiguous the chain of command is in the directed network. Mathematically, it is the average of $f(G)$ over the set in $W(G)$, where $W(G)$ is the subset G_C and all of its subsets obtained through a leaf removal algorithm. This measure is defined in [2] as:
$$f(G) = \frac{H_f(G_C) - H_b(G_C)}{max\{H_f(G_C), H_b(G_C)\}},$$
where H_f, H_b denote the forward and backward path entropies, respectively. Path entropy is defined to be: $h_f(v_i) = -\sum P(\pi_k|v_i)logP(\pi_k|v_i)$.
4. Feedforwardness $F \in [0, 1]$ tries to penalize cycles near the top of the directed network. Mathematically, it is the average of path weights $F(\pi_k)$ where cyclic modules that are closer to the top get a higher penalty, and defined by [2]. Here the paths under consideration consist of the subset starting from the top node of G_C. If there are p such paths then:
$$F(G) = \frac{1}{p}\sum_{k=1}^{p}\frac{|v(\pi_k)|}{\sum a_i},$$
where a_i are the weights of each node along the path under consideration (i.e. the number of collapsed nodes from G in the corresponding node of G_C), and $v(\pi_k)$ is the number of nodes along the path π_k.
5. Orderability $O \in [0, 1]$ defines how orderable the directed network is. Mathematically, it is the fraction of nodes that do not belong to any cycle and defined by [2]:
$$O(G) = \frac{|v_i \in V_c \cap V|}{|V|}$$
6. Global Reaching Centrality, based on the m-reach centrality measure is adapted in [3] 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 [3] is used:
$$C_R^{'}(i) = \frac{1}{N-1}\sum_{j:0<d^{out}_{(i,j)<\infty}}\left(\frac{\sum_{k=1}^{d^{out}(i,j)} w_i^{(k)}(j)}{d^{out}(i,j)}\right)$$
   (c) When the graph is unweighted and undirected, the following version for the reaching centrality as defined by [3] is used:
$$C_R^{''}(i) = \frac{1}{N-1}\sum_{j:0<d(i,j)<\infty}\frac{1}{d(i,j)}$$
7. For social networks, [4] 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.
8. Rooted Depth in directed networks is defined in [5], where a root is a node that has only incoming edges. 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. 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.

9. Relative Depth in directed networks is defined in [5] by first collapsing the network into G_C. Then the main root must be found, which lies on the end of the longest path through the network. From here, each vertex is given a relative depth d equal to the path length between itself and the main root. Let $L = 1, ...l$ be the set of leaves and $R = 1, ...r$ be the set of roots. Then the global relative depth is found by:

$$D = \frac{1}{l}\sum_{i=1}^{l} d_i - \frac{1}{r}\sum_{i=1}^{r} d_i$$

10. Betweenness Centrality
11. Eigenvector
12. Nodal Degree
13. Closeness Centrality

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.

A. mong 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 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 outcomesi.e., the ability of particular nodes to control the actions and behaviors of subordinate nodes.

W. e 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.

(a) Adolescent Health: survey asked students to list 5 male and female friends. [7]
(b) Residence Hall: friendships between 217 students in Australian National University. [8]
(c) Taro Exchange: gift–giving relationships between households in a Papaun village. [9]
(d) Highschool: friendship relationship between boys at a small Indiana high school in 1957-1958. [10]
(e) Dutch College: friendships between 32 university freshmen. [11]
(f) Monks: preference ratings between monks in a cloister during a crisis. [12]
(g) Physicians: innovation spread between 246 physicians in Illinois. [13]
(h) Seventh graders: activity specific proximity rankings for 29 middle school students in Victoria [14].
(i) Prosper loans: loans between users of prosper.com [**?**].

(j) Libimseti.cz: likes between users on a Czech dataing site [16].
(k) Friendster: friendship adds on the online site Friendster [31].
(l) Digg: friendships on Digg [?].
(m) Youtube: connections between Youtube users [19].
(n) Epinions: who–trusts–whom between users of epinions [20].
(o) EU emails: emails for 18 months from a major European research institution [21].
(p) Facebook: friends lists from FAcebook, generated through a Facebook app survey [22].
(q) Google Plus: friends between users who selected to "share circles" on Google Plus [22].
(r) Linx kernel mailing list: communication network for the linux kernel mailing list, where each edge is a reply from a user to another [23].
(s) Livejournal: map of an online community friendships of Livejournal users [24].
(t) Manufacturing: communication network between employess of a mid–size manufacturing firm [25].
(u) Pokec: Friendship networks in the Pokec online social network, popular in Slovakia [26].
(v) Slashdot: tagging between users in slashdot for 2008 and 2009 [24].
(w) Twitter: circles between twitter users [22].
(x) UC Irvine: messages sent between students on an online community at UC Irvine [27].
(y) U. Rovira i Virgili: email communication network from University Rovira i Virgili in Tarragona [28].
(z) Wikipedia Talk: network of discussions between all users from the beginning of Wikipedia to January 2008 [29].
(aa) Wikipedia Votes: data from administrator elections [29].
(ab) Wikipedia Requests for Adminship: requests from 2003 through 2013 [30].
(ac) Friendster: network for online social site Friendster [31].

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.

T. he 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 programed 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] Shizuka, D. and McDonald, D. B. (2012), *A social network perspective on measurements of dominance hierarchies*. Animal Behavior: 83, 925-934.
[2] Murtra, B., Goni, J., and Caso, C. (2013), *On the origins of hierarchy in complex networks*. PNAS: 110, 33, 13316-13321.
[3] Mones, E., Vicsek, L., and Vicsek, T. (2012), *Hierarchy Measure for Complex Networks*. Plos ONE: 7, 3, 1-10.
[4] 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).

[5] Suchecki, K. and Holyst, J. (2013), *Hierarchy depth in directed networks*. Physica A: preprint.
[6] Liu, Y., Slotine, J., and Barabási, A (2012). *Control centrality and hierarchical structure in complex networks*. Plos ONE: 7, 9.
[7] Moody, J. (2001). *Peer influence groups: Identifying dense clusters in large networks*. Social Networks: 23, 4, 261-283.
[8] Freeman, L., Webster, C., Kirke, D. (1998) *Exploring social structure using dynamic three–dimensional color images.* Social Networks: 20, 2, 109-118.
[9] 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.
[10] Coleman, J. (1973) *Introduction to mathematical sociology*. London Free Press Glencoe.
[11] Van de Bunt, G., Van Deuijn, M., Snijders, T. (1999) *Friendship networks through time: An actor-oriented dynamic statistical network model*. Computational and Mathematical Organization Theory: 5, 2, 167-192.
[12] Breiger, R., Boorman, S., 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, 3, 1975.
[13] Coleman, J., Katz, E., Menzel, H. (1957) *The diffusion of an innovation among physicians*. Sociometry: 253-270.
[14] Watts, D., Strogatz, S. (1998) *Collective dynamics of 'small world' networks*. Nature: 393, 1, 440-442
[15] Prosper loans network dataset - KONECT, May 2015.
[16] Brozovsky, L., Petricek, V. (2007) *Recommender system for online dating service*. Proc. Znalosti: 29-40.
[17] Friendster network dataset - KONECT, May 2015.
[18] Hogg, T., Lerman, K. (2012) *Social dynamics of Digg*. EPJ Data Science: 1, 5.
[19] Mislove, A., Marcon, M., Gummadi, K., Druschel, P., Bhattacharjee, B. (2007) *Measurement and analysis of online social networks*. Proc. Internet Measurement Conference.
[20] Richardson, M., Agrawal, R., Domingos, P. (2003) *Trust Management for the Semantic Web*. ISWC
[21] Leskovec, J., Kleinber, J., Faloutsos, C. (2007) *Graph Evolution: Densification and Shrinking Diameters*. ACM Transactions on Knowledge Discovery from Data (ACM TKDD): 1, 1.
[22] McAuley, J., Leskovec, J. (2012) *Learning to Discover Social Circles in Ego Networks*. NIPS
[23] Linux kernel mailing list replies network dataset - KONECT, May 2015.
[24] Leskovec, J., Lang, K., Dasgupta, A., 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.
[25] Michalski, R., Palus, S., Kazienko, P. (2011) *Matching organizational structure and social network extracted from email communication*. Lecture Notes in Business Information Processing: 87, 196-206.
[26] Takac, L., Zabovsky, M. (2012) *Data Analysis in Public Social Networks*. International Scientific Conference & International Workshop Present Day Trends of Innovations, Lomza, Poland.
[27] Opsahl, T., Panzarasa, P. (2009) *Clustering in weighted networks*. Social Networks: 31, 2, 155-163.
[28] Guimera, R., Danon, L., Diaz-Guilera, A., Giralt, F., Arenas, A. (2003) *Self–similar community structure in a network of human interactions*. Phys. Rev. E.: 68, 6.
[29] Leskovec, J., Huttenlocher, D., Kleinberg, J. (2010) *Predicting Positive and Negative Links in Online Social Networks*. WWWW
[30] West, R., Paskov, H., Leskovec, J., Potts, C. (2014) *Exploiting Social Network Structure for Person-to-Person Sentiment Analysis*. Transactions of the Association for Computational Linguistics: 2, 297-310.
[31] Friendster network dataset - KONECT, May 2015.
[32] Mann, Michael. "Chapter 1: Societies as organized power networks." *The Sources of Social Power, Volume 1, A history of power from the beginning to AD 1760*. Cambridge University Press, 1986: 1-33.