Available online at www.sciencedirect.com
ScienceDirect
Social Networks 00 (2015) 1–??
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. Landau's $h \in [0, 1]$ and Kendall's $K \in [0, 1]$ are both used to compare a network to a perfect linear hierarchy in [1], where $i = 1...N$ is the number of nodes and S_i is the row sum for each node also referred to as the dominance total:

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

IF we 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

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

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

2. Triangle transitivity is shown to be higher in dominance relationships in [1], but not used as a measure here.

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

3. Treeness $T \in [-1, 1]$ 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].

$$f(G) = \frac{H_f(G_C) - H_b(G_C)}{max\{H_f(G_C), H_b(G_C)\}},$$

where H_f, H_b denote the forward and backward path entropies, respectively, where $h_f(v_i) = -\sum P(\pi_k|v_i) log P(\pi_k|v_i)$.

4. Feedforwardness $F \in [0, 1]$ 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 are all paths starting at from the top of G_C denoted π_k. Let $k = 1, ...M$ be this number of paths then:

$$F(G) = \frac{1}{M} \sum_{k=1}^{M} \frac{|v(\pi_k)|}{\sum a_i},$$

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 $O \in [0, 1]$ 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 where the graph is unweighted and directed is defined by [3] as:

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

where $C_R(i)$ is the local reaching centrality defined as the proportion of all nodes in G that can be reached along outgoing edges from node i.

7. Global Reaching Centrality where the graph is weighted and directed is defined by [3] using the following version for the reaching centrality:

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

8. Global Reaching Centrality where the graph is unweighted and undirected is defined by [3] using the following version for the reaching centrality:

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

9. Control Centrality in a weighted and directed graph, defined by [4], identifies the minimum number of nodes need to drive an entire network to a given final state. Consider a directed, weighted network:

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

which is the state of each node at time t, and also denoted as $(\boldsymbol{A}, \boldsymbol{B})$. The components of this controllability matrix are: $\boldsymbol{A} \in \mathbb{R}^{N \times N}$, where each element gives the strength that node j can affect node i; and $\boldsymbol{B} \in \mathbb{R}^{N \times M}$, 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 $\boldsymbol{C}$ = ($\boldsymbol{A}$,$\boldsymbol{B}$), the control centrality of node i is:

$$C_c(i) \equiv rank_g(\boldsymbol{C}^i)$$

4. Data

1. Adolescent Health: survey asked students to list 5 male and female friends. [5]

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

5. Analysis

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] Liu, Y., Slotine, J., and Barabási, A (2012). *Control centrality and hierarchical structure in complex networks*. Plos ONE: 7, 9.
[5] Moody, J. (2001). *Peer influence groups: Identifying dense clusters in large networks*. Social Networks: 23, 4, 261-283.
[6] Freeman, L., Webster, C., Kirke, D. (1998) *Exploring social structure using dynamic three–dimensional color images.* Social Networks: 20, 2, 109-118.
[7] 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.
[8] Coleman, J. (1973) *Introduction to mathematical sociology*. London Free Press Glencoe.
[9] 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.
[10] 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.
[11] Coleman, J., Katz, E., Menzel, H. (1957) *The diffusion of an innovation among physicians*. Sociometry: 253-270.

[12] Watts, D., Strogatz, S. (1998) *Collective dynamics of 'small world' networks*. Nature: 393, 1, 440-442
[13] Prosper loans network dataset - KONECT, May 2015.
[14] Brozovsky, L., Petricek, V. (2007) *Recommender system for online dating service*. Proc. Znalosti: 29-40.
[15] Friendster network dataset - KONECT, May 2015.
[16] Hogg, T., Lerman, K. (2012) *Social dynamics of Digg*. EPJ Data Science: 1, 5.
[17] Mislove, A., Marcon, M., Gummadi, K., Druschel, P., Bhattacharjee, B. (2007) *Measurement and analysis of online social networks*. Proc. Internet Measurement Conference.
[18] Richardson, M., Agrawal, R., Domingos, P. (2003) *Trust Management for the Semantic Web*. ISWC
[19] Leskovec, J., Kleinber, J., Faloutsos, C. (2007) *Graph Evolution: Densification and Shrinking Diameters*. ACM Transactions on Knowledge Discovery from Data (ACM TKDD): 1, 1.
[20] McAuley, J., Leskovec, J. (2012) *Learning to Discover Social Circles in Ego Networks*. NIPS
[21] Linux kernel mailing list replies network dataset - KONECT, May 2015.
[22] 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.
[23] 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.
[24] Takac, L., Zabovsky, M. (2012) *Data Analysis in Public Social Networks*. International Scientific Conference & International Workshop Present Day Trends of Innovations, Lomza, Poland.
[25] Opsahl, T., Panzarasa, P. (2009) *Clustering in weighted networks*. Social Networks: 31, 2, 155-163.
[26] 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.
[27] Leskovec, J., Huttenlocher, D., Kleinberg, J. (2010) *Predicting Positive and Negative Links in Online Social Networks*. WWWW
[28] 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.
[29] Friendster network dataset - KONECT, May 2015.