Leadership Insularity:
A New Measure of Connectivity Between Central Nodes in Networks

Samuel Arbesman
Department of Health Care Policy, Harvard Medical School
Cambridge, Massachusetts

Nicholas A. Christakis
Departments of Health Care Policy & Medicine, Harvard Medical School
Department of Sociology, Harvard University
Cambridge, Massachusetts

We combine two foci of interest with respect to community identification and node centrality and create a novel metric termed "leadership insularity." By determining the most highly connected nodes within each community of a network, we designate the 'community leaders' within the graph. In doing this, we have the basis for a novel metric that examines how connected, or disconnected, the leaders are to each other. This measure has a number of appealing measurement properties and provides a new way of understanding how network structure can affect its dynamics, especially information flow. We explore leadership insularity in a variety of networks.

Authors
Samuel Arbesman, Ph.D. (Harvard University) is a posdoctoral research fellow in the Department of Health Care Policy at Harvard Medical School and affiliated with the Institute for Quantitative Social Science at Harvard University.

Nicholas A. Christakis, M.D., Ph.D., M.P.H (Harvard University) is Professor of Medical Sociology in the Department of Health Care Policy at Harvard Medical School; Professor of Medicine in the Department of Medicine at Harvard Medical School; and Professor of Sociology in the Department of Sociology in the Harvard Faculty of Arts and Sciences.

Acknowledgements: *The authors would like to thank James Fowler and Alan Zaslavsky for useful discussions. This was work was supported by NIH (P-01 AG031093) and by the Pioneer Portfolio of the Robert Wood Johnson Foundation.*

Correspondence: *Contact Samuel Arbesman at Harvard Medical School, Department of Health Care Policy, 180 Longwood Avenue, Boston, MA 02115. Phone: 617-432-7421; fax: 617-432-5891; email: arbesman@hcp.med.harvard.edu.*

INTRODUCTION

In recent years, there has been considerable work in two areas of network measurement: community identification and node centrality. Communities within networks are often identified as subgraphs that are connected more tightly than the graph as a whole. The available algorithms vary widely and include traditional clustering techniques, centrality-based community detection, and modularity-based methods (Porter, Onnela, & Mucha, 2009). Furthermore, there are many methods of determining the most centrally located nodes within a network. These range from examining the node with the highest degree to the node with the highest betweenness centrality and so forth (Newman, 2003).

Here, we combine these methods and create a novel metric known as “leadership insularity.” By determining the most highly connected nodes within each community of a network, we are able to determine the ‘community leaders’ within the graph. In doing this, we have the basis for a novel metric that examines how connected, or disconnected, the leaders are connected to *each other*. This measure can be used to characterize individual leaders in a network (in terms of how isolated they are from other leaders) or it can be used to summarize the property of a whole network (in terms of how isolated its leaders are compared to other networks). This measure of insulation provides a new way of understanding how network structure can affect its dynamics, especially information flow.

Using a topographic analogy, as in Figure 1, each community may be viewed as an individual mountain within a mountain range, with its leader as the peak. The topography of the mountain range can vary wildly, and has implications for how closely connected the peaks are. Analogously, if the ‘slope’ of a community were shallow, two leaders would only be able to interact via many intermediaries. However, if the distance is much closer, then they might be able to interact more effectively. This has implications for many situations, such as coordination problems (Kearns, Suri, & Montfort, 2006).

Guimera et al. hint at something similar to leadership insularity, though their metrics are somewhat different (Guimerà, Mossa, Turtschi, & Amaral, 2005). They identify a number of different categories of nodes and even create a metric called the participation coefficient (which examines how connected nodes in a community are connected to other communities). Our measure is different in that it is mathematically simpler, by focusing only on the leaders of the communities, as opposed to all nodes. Moreover, since community leaders often have an outsized influence on the dynamics of their groups, it is useful to have a *single* metric for an entire network's leadership insularity.

Figure 1. A Metaphorical View of Leadership Insularity

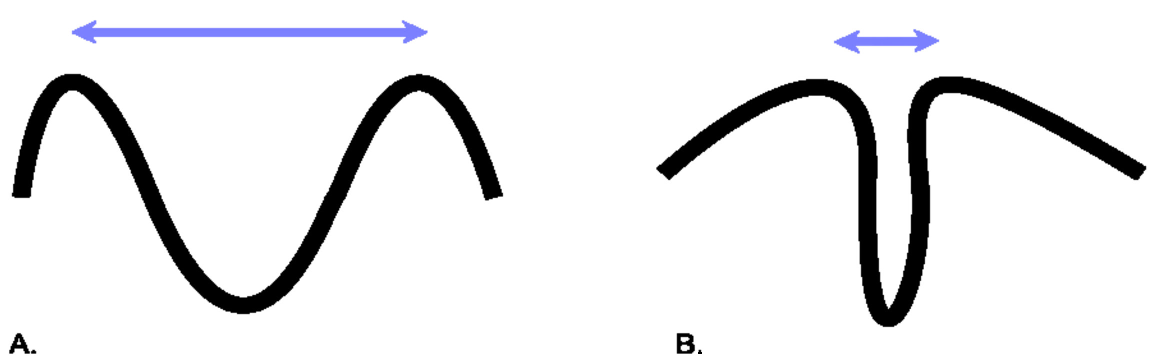

Figure 1. Using the topographical imagery provided in the text: Part A has a large distance between leaders/peaks, while Part B has a much smaller distance between leaders.

By being able to quantify the distance between these community leaders, we can understand the structure and dynamics of networks better. After explaining the metric, which has some appealing measurement properties, we explore the leadership insularity of a variety of networks and examine how it relates to the diverse functions of these networks.

METHODS

1. Description of the Metric

Leadership Insularity is simply defined as the average relative distance between the leaders of different communities. This is achieved by dividing the path length between each leader by the average path length between any two individuals of their respective communities. The overall leadership insularity then becomes the average of these relative path lengths, weighted according to the size of the communities. The equation, visualized in Figure 2, is as follows:

$$I = \frac{1}{(N_c - 1)N} \sum_{i=1}^{N_c} \sum_{j=i+1}^{N_c} \frac{d(L_i, L_j)}{d(i,j)} (N_i + N_j) \tag{1}$$

Where the variables are defined as follows:

N_c = number of communities identified
N = number of nodes in the network
N_i = number of nodes in community i
L_i = leader of community i
$d(L_i,L_j)$ = distance between community leaders L_i and L_j
$d(i,j)$ = mean distance between communities i and j

The term:

$$\frac{1}{(N_c - 1)N} \sum_{i=1}^{N_c} \sum_{j=i+1}^{N_c} (N_i + N_j)$$ equals 1 and

is used to allow a weighted average of the various relative distances between community leaders.

In addition, the leadership insularity can be calculated for a single leader within the network as follows:

$$I_i = \frac{1}{2N_c N} \sum_{j \neq i}^{N_c} \frac{d(L_i, L_j)}{d(i,j)} (N_i + N_j) \tag{2}$$

When the mean of these individual leadership insularities is taken, the leadership insularity of the entire network is obtained.

The communities can be identified by a variety of methods, as can the community leaders. For the purposes of the implementation of the metric, we used the method described in Clauset to identify communities within our networks (Clauset, 2005). The community leaders were those nodes with the highest betweenness centrality when a community was viewed as a graph, separate from the network as a whole. If there are two or more nodes with equally high betweenness centralities, then a comparison is made to the nodes with the highest degree centrality. A randomly selected node from the intersection of the nodes with the highest betweenness and degree centralities is chosen (if the intersection has no nodes, then a randomly selected node from the highest betweenness centralities is used). This use of a combination of centrality measures is similar to that used by researchers studying peer-education and food intake (D. Buller et al., 2000; D. B. Buller et al., 1999).

Figure 2. Visual Demonstration of Leadership Insularity

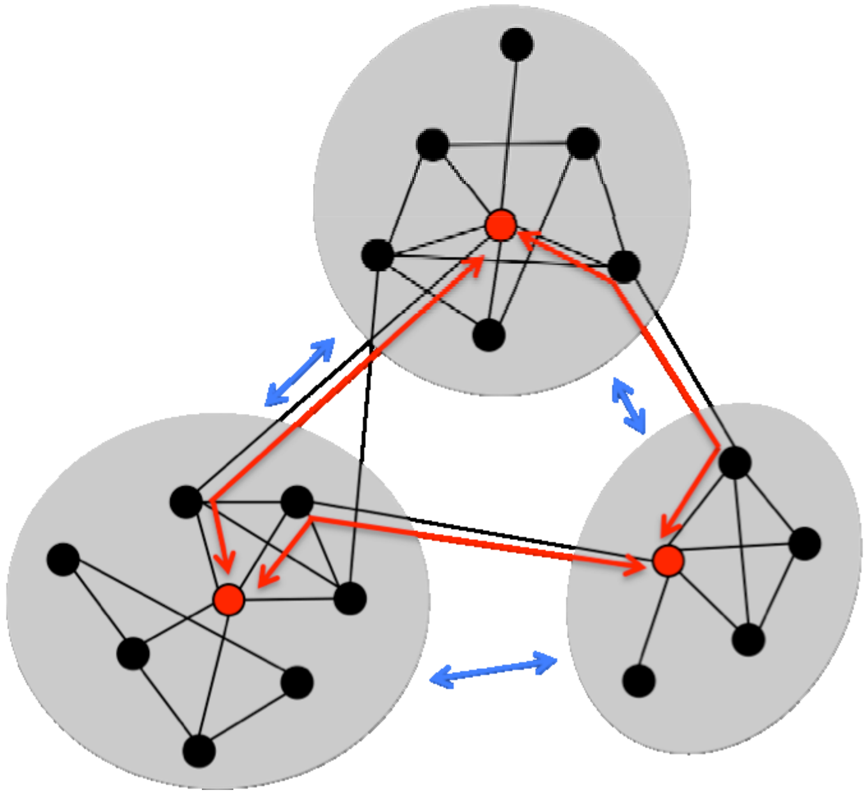

Figure 2. Red nodes indicate community leaders and red lines indicate distance between them. Blue lines indicate the mean distance of the individuals in one community to another. The calculated leadership insularity is 0.68. Figure adapted from Newman (Newman, 2003).

In the Addhealth dataset the number of communities with multiple equally good choices as leader is 3.2% of the total 1570 communities within the networks, with the majority of these only containing two possible leaders, and most of these possible leaders being the most central nodes for both measures (see section 2A). However, it seems that these numbers might be domain-specific. For example, one of our scientific collaboration datasets. Condensed Matter arXiv 2003 (see section 2B), had multiple equally good choices for the leader in about 25% of the communities, and these communities contained more than two possible leader choices (often around seven). Therefore, leader identification in different domains merits further study.

In addition, we performed a robustness test on the use of betweenness centrality for leader detections by creating a modified metric that uses degree centrality as the primary criterion (with betweenness centrality as the secondary criterion). Using this modified metric, a similar dispersion of leadership insularity was found in the Addhealth dataset as below, and similar correlations (albeit with less significant p-values).

The code has been implemented in Python and requires the packages of igraph and NetworkX (Csárdi & Nepusz, 2006; Hagberg, Schult, & Swart, 2008). It is being released under the GPL license and will be downloadable from the following locations:

http://christakis.med.harvard.edu/

http://arbesman.net/

Figure 3. Dispersion of Leadership Insularity in Schools

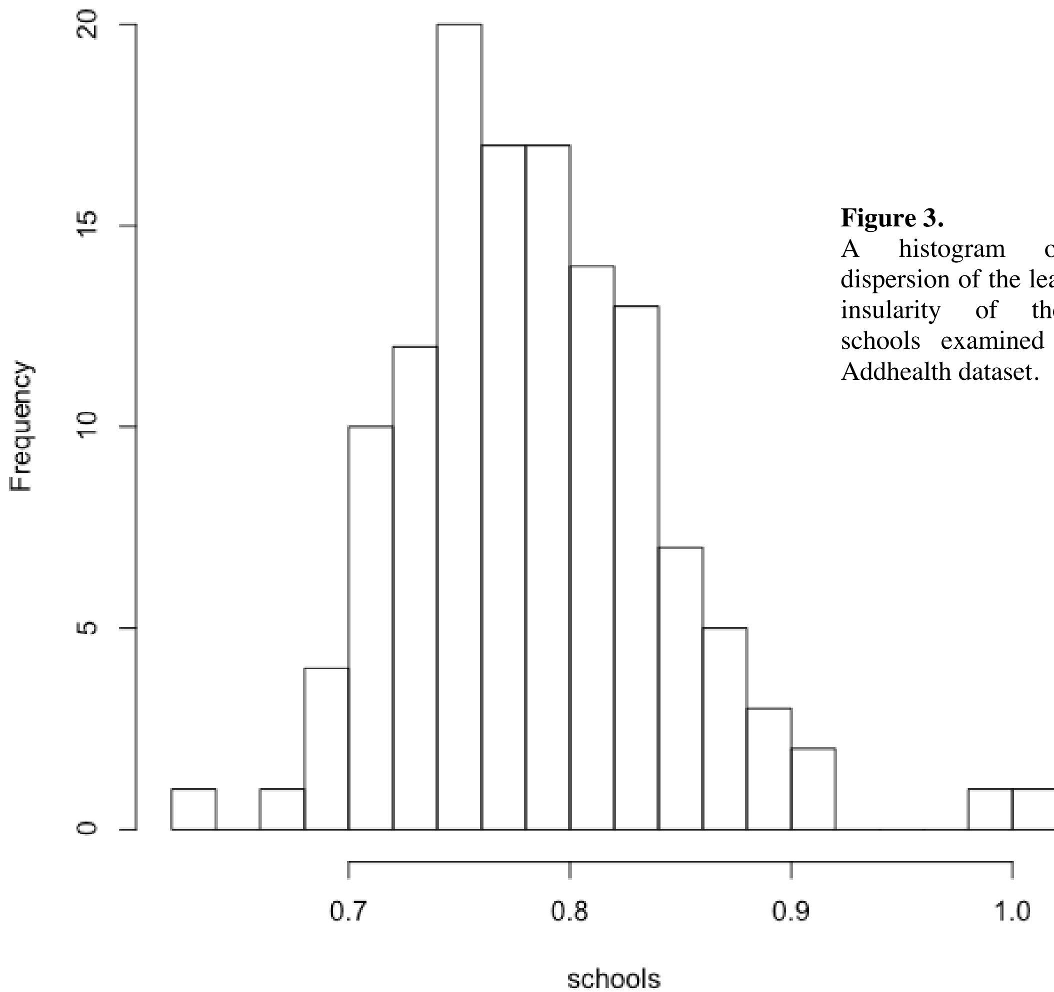

Figure 3.
A histogram of the dispersion of the leadership insularity of the 142 schools examined in the Addhealth dataset.

2. Applications

2.1 Addhealth Dataset

To test the robustness and applicability of leadership insularity, we applied the metric to a variety of networks. Our first test consisted of examining high school social networks in different schools. We expected that there would be significant variation between schools, and that this variation would be related to other differences between schools. We used the Addhealth dataset, a survey conducted in 142 American high schools (Harris, 2008). As part of the survey, adolescents were asked about their social ties, which allowed us to reconstruct the social networks for each high school.

A high degree of dispersion was found in the high schools, as seen in Figure 3. In addition, we observed a significant relationship between a high school's leadership insularity and certain other attributes of the schools, such as the extent to which students feel safe at school or the average tenure of the students in the school. For example, a simple OLS regression model reveals that schools with a high LI had a higher duration of time the students had been in the school, regression coefficient = 6.69, $p < 0.0001$ (standard error = 1.37). Schools with high LI also had students who were more likely to report feeling safe in the school, regression coefficient = 1.69, p=0.003 (standard error = 0.555). The longer the average duration of the students in a school could very easily lead to a certain amount of social insularity, which would in turn lead to leadership insularity. Less turnover in the nodes on the network also would stabilize the cliques in the schools, and their leaders. Similarly, this type of social insularity might lead to a greater feeling of safety in one's school and neighborhood, since one's social circle is cloistered and insulated from the world at large.

Scientific Coauthorship Networks. We also examined the variation in leadership insularity for various scientific coauthorship networks. These networks are constructed from authorship of scientific papers, where two individuals are connected if they coauthored a paper. We examined the coauthorship networks compiled from selected subareas within arXiv, an online preprint repository with a physics focus. The areas we looked at are theoretical high energy physics (hep-th), condensed matter (cond-mat), and astrophysics (astro-ph) (Newman, 2001). In addition, a smaller dataset composed network science articles (netscience) was also included (Newman, 2006). As a check, we also used a more recent version of the condensed matter coauthorship network (up to 2003, as opposed to 1999) to ensure that each area's leadership insularity was reasonably robust.

As seen in Table 2, there is a certain amount of variation in the leadership insularities of the different scientific disciplines. This could be due to a variety of factors, such as the degree of collaboration within the networks. Patterns of collaboration and interaction vary between scientific areas, and these differences are visible in differences between leadership insularity. In addition, with more data available, such as the number of citations (as an indication of the impact of the discipline), it could be seen whether or not the connectivity between scientific 'leaders' has an impact on the productivity of a discipline or leader.

Table 2. Leadership Insularity of Scientific Subdisciplines

arXiv Area	Leadership Insularity
hep-th	0.70
netscience	0.69
cond-mat	0.77
astro-ph	0.76
cond-mat-2003	0.76

CONCLUSIONS

Large groups configured as networks have subgroups, and subgroups typically have leaders. The ability of the group as a whole to function may be related to how integrated its leaders are with each other, and not just with their own group members, especially when communication flows between leaders are indirect (through others) and not direct (in the form of person-to-person ties). Otherwise similar networks may therefore differ meaningfully in terms of how inter-connected their leaders are, and this measure may correlate with a variety of internal and external properties of the network. We have proposed a novel metric, termed leadership insularity, to capture the degree of social isolation of central nodes of different communities within networks.

REFERENCES

Buller, D., Buller, M. K., Larkey, L., Sennott-Miller, L., Taren, D., Aickin, M., et al. (2000). Implementing a 5-a-Day Peer Health Educator Program for Public Sector Labor and Trades Employees. *Health Education & Behavior, 27*(2), 232-240.

Buller, D. B., Morrill, C., Taren, D., Aickin, M., Sennott-Miller, L., Buller, M. K., et al. (1999). Randomized Trial Testing the Effect of Peer Education at Increasing Fruit and Vegetable Intake. *JNCI Journal of the National Cancer Institute, 91*(17), 1491-1500.

Clauset, A. (2005). Finding local community structure in networks. *Physical Review E, 72*(2), 026132.

Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*.

Guimerà, R., Mossa, S., Turtschi, A., & Amaral, L. A. N. (2005). The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences, 102*(22), 7794-7779.

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy2008)*, 11-15.

Harris, K. M. (2008). The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002 [machine-readable data file and documentation]. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.

Kearns, M., Suri, S., & Montfort, N. (2006). An Experimental Study of the Coloring Problem on Human Subject Networks. *Science, 313*(5788), 824-827.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 404-409.

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review, 45*(2), 167-256.

Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), 74*(3), 036104-036119.

Porter, M. A., Onnela, J.-P., & Mucha, P. J. (2009). *Communities in Networks*.