1-30-2008

Localized Bridging Centrality for Distributed Network Analysis

Soumendra Nanda
Dartmouth College

David Kotz
Dartmouth College

Localized Bridging Centrality for Distributed Network Analysis

Soumendra Nanda* and David Kotz

Department of Computer Science

and Institute for Security Technology Studies,

Dartmouth College, Hanover, NH 03755

Dartmouth Computer Science Technical Report TR2008-612

January 30, 2008

Abstract

Centrality is a concept often used in social network analysis to study different properties of networks that are modeled as graphs. We present a new centrality metric called Localized Bridging Centrality (LBC). LBC is based on the Bridging Centrality (BC) metric that Hwang et al. recently introduced. Bridging nodes are nodes that are located in between highly connected regions. LBC is capable of identifying bridging nodes with an accuracy comparable to that of the BC metric for most networks. As the name suggests, we use only local information from surrounding nodes to compute the LBC metric, while, global knowledge is required to calculate the BC metric. The main difference between LBC and BC is that LBC uses the egocentric definition of betweenness centrality to identify bridging nodes, while BC uses the sociocentric definition of betweenness centrality. Thus, our LBC metric is suitable for distributed computation and has the benefit of being an order of magnitude faster to calculate in computational complexity. We compare the results produced by BC and LBC in three examples. We applied our LBC metric for network analysis of a real wireless mesh network. Our results indicate that the LBC metric is as powerful as the BC metric at identifying bridging nodes that have a higher flow of information through them (assuming a uniform distribution of network flows) and are important for the robustness of the network.

*Corresponding author: snanda@cs.dartmouth.edu

Keywords: wireless mesh networks, network management, network monitoring, network diagnosis, network analysis, social network analysis, centrality, distributed algorithms

1 Introduction

Our initial motivation for this work was to discover metrics and develop tools that can help a system administrator manage a wireless mesh network or would allow an automated management system understand the state of a network. We provide below a list of questions asked from a system administrator's point of view, that we initially set out to answer and that we consider as relevant to our scenarios.

1. Which nodes should the system administrator be most concerned about from a robustness point of view? That is, the loss of which nodes would have a significant impact on the connectivity of the network?
2. How many nodes can fail before my network is partitioned into multiple components?
3. Which nodes are the most "important" in my network?
4. Similarly, which nodes are the least important and why?
5. If I could or should add or move a node to enhance the network, which one should it be?
6. Similarly, if I had to update a subset of nodes and reboot them, in which order should I perform the update?

One technique to identify which nodes are critical from a network management perspective is to identify all articulation points and bridges in the network topology. When applied to wireless mesh networks, in our experience, we found that articulation points are rare in practice unless the network has a low density.

Our goal is to apply social-network analysis techniques to identify properties of individual nodes that can aid a system administrator to manage a mesh network in a more effective manner. While

the system administrator is primarily asked to perform absolute tasks (e.g., to fix or replace a non-functioning node), there may be situations when relative decisions must be made. For example, we posed the following question earlier: "If the system administrator had to update a subset of nodes and reboot them, then in which order should he or she perform the update?". Since we are interested in relative comparisons between seemingly similar nodes to answer such questions, we need to develop techniques and metrics that differentiate between nodes and rank them. We apply techniques from social-network analysis to attempt to answer these types of questions. In a wireless mesh network context, a system administrator should pay attention to bridging nodes since they are important from a robustness perspective (as they help bridge connected components together) and their failure may increase the risk of network partitions.

The main contribution of this work is the development of the LBC metric, which is equivalent in functionality to the Bridging Centrality [11] metric at identifying bridging nodes, yet can be calculated quickly in a distributed manner. BC is calculated in a centralized manner and has an order of magnitude higher computational complexity. To calculate its own LBC value, each node only needs to know its 1-hop neighbor set and the degree of each of its neighbors.

The outline of the rest of this paper is as follows. We describe the basics of social-network analysis in Section 2, and explain three common social centrality metrics. In Section 2.3, we explain the key differences between sociocentric and egocentric betweenness, since our approach essentially builds upon the difference between these two metrics. Readers familiar with centrality may jump to Section 3 where we present the definition of Bridging Centrality and introduce our definition of the Localized Bridging Centrality metric. Finally, we look at our initial results and present our conclusions.

2 Social-network analysis

We believe that techniques borrowed and enhanced from the domain of social network analysis can help in providing answers to some of the questions we pose. We aim to use "centrality" metrics from social-network analysis to study the roles of individual nodes in the network and the relationship of these nodes to their neighbors. Social-network analysis is normally applied to the study of social

networks of actors, usually people and their relationships with other people. In our domain, we are interested in the positions and roles of individual mesh nodes and the relationships between different mesh nodes such as connectivity, which can be characterized in different ways such as direct or indirect, weak or strong. Many social-network analysis techniques and metrics are based on graph theory. Humans tend to form clusters of communities within social networks. Similarly, mesh networks may have groups of nodes that share a common relationship or structure, which may be worth identifying.

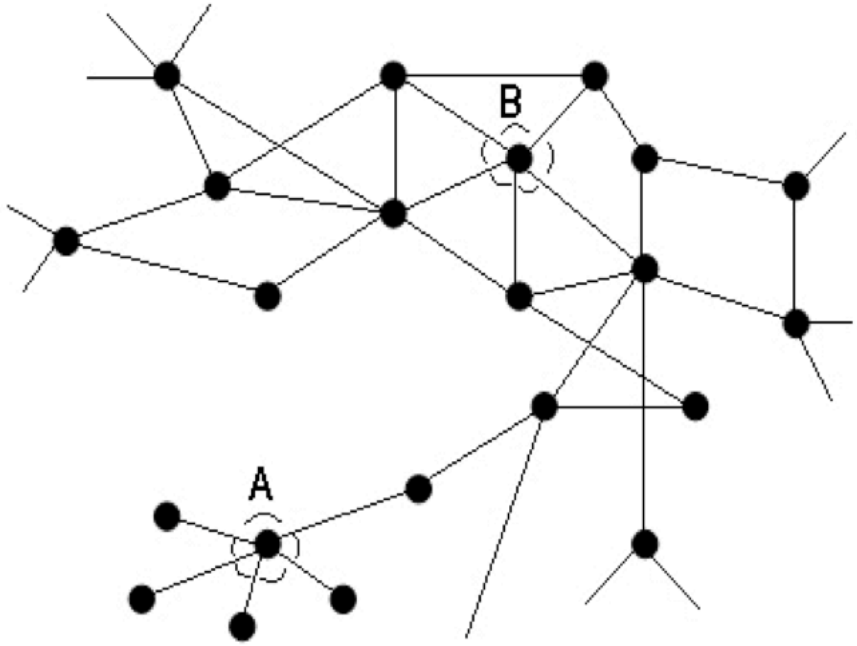
2.1 Degree centrality

One simple way to characterize an individual node in a topological graph is by its *degree*. The *degree* of a node in a graph in the mesh context is the number of links the node shares with its neighbors, which are available for routing purposes. A well-connected mesh network is a healthy network. If a node has many neighbors then the failure of a single neighbor should not affect the routing health of the regional network adversely. A node with a high degree can be considered as being well connected and a node with a relatively low degree can be considered weakly connected. The degree of an individual node and the minimum, maximum and average degree over all the nodes are standard characterization metrics in graph theory.

If the global topology is available at a central location, then all the nodes can be quickly ranked according to their degree. However, this degree-based ranking does not convey a good picture of the nature of connectivity in the network since all links are rarely identical. For instance, different links may have varying capacity levels and different latencies. In addition, the existence of neighbor links and their respective qualities fluctuate over time. In a wireless network a link with a poor-quality connection has lower effective capacity and a link using a lower bit-rate may have a higher latency.

Even two nodes with the same degree but need not have similar characteristics [1] (for example, see Figure 1). There are other centrality metrics, such as eigenvector centrality, which can help distinguish between nodes A and B that have the same degree centrality.

Figure 1: Limitations of degree centrality (note that nodes A and B each have degree 5) [1]



2.2 Eigenvector centrality

Eigenvector Centrality (EVC) is a concept often used in social-network analysis and was first proposed by Bonacich [2]. Eigenvector Centrality is defined in a circular manner. The centrality of a node is proportional to the sum of the centrality values of all its neighboring nodes. In the social-network context, an important node (or person) is characterized by its connectivity to other important nodes (or people). A node with a high centrality value is a well-connected node and has a dominant influence on the surrounding network. Similarly, nodes with low centrality values are less similar to the majority of nodes in the topology and may exhibit similar characteristics and behavior and share common weaknesses. Google uses a similar centrality ranking technique called Pagerank [6] to rank the relevance of pages in search results.

Eigenvector centrality is calculated using the adjacency matrix to find central nodes in the network. Let v_i be the i_{th} element of the vector $\vec{v}$, representing the centrality measure of node i, where $N(i)$ is the set of neighbors of node i and let A be the $n \times n$ adjacency matrix of the

undirected network graph. Eigenvector centrality is defined using the following formulas [1]:

$$v_i \propto \sum_{j \in N(i)} v_j \tag{1}$$

which can be rewritten as

$$v_i \propto \sum_{j=1}^{n} A_{ij} v_j \tag{2}$$

which can be rewritten in the form

$$A\vec{v} = \lambda\vec{v} \tag{3}$$

Since A is an n x n matrix, it has n eigenvectors (one for each node in the network) and n corresponding eigenvalues. One way to compute the eigenvalues of a square matrix is to find the roots of the characteristic polynomial of the matrix. It is important to use symmetric positive real values in the matrix used for calculations [3].

The principle eigenvector is recommended for use in rank calculations. The principle eigenvector is the eigenvector with the highest eigenvalue. After the principle eigenvector is found, its entries are sorted from highest to lowest values to determine a ranking of nodes. The most central node has the highest rank and most peripheral node has the lowest rank.

This metric is often used in the study of the spread of epidemics in human networks. In the mesh context, a node with a high eigenvector centrality represents a strongly connected node. A worm or virus propagated from the most central node could spread to all reachable nodes in the most efficient manner as opposed to one that was spreading from a node on the extreme periphery. Thus, the central node is a prime target for preventive inoculation or for prioritized software update.

In any network, and especially in an ad hoc or mesh network where nodes must cooperate with each other to route packets, the connectivity of a node depends on the connectivity of its neighbors and EVC can help capture this property. The main drawback of eigenvector centrality is that it can only be calculated in a central manner.

2.3 Betweenness centrality

In addition to the above two centrality metrics, several other definitions of centrality measures exist, such as closeness centrality, graph centrality and betweenness centrality. We focus now on betweenness centrality [10], which is also called sociocentric betweenness centrality. Betweenness centrality is a key component of the bridging centrality metric.

2.3.1 Sociocentric betweenness centrality

The sociocentric betweenness centrality of a node is calculated as the fraction of shortest paths between all node pairs that pass through the node of interest. A node with a high betweenness centrality value is more likely to be located on the shortest paths between multiple node pairs in the network, and thus more information must travel through that node (assuming a uniform distribution of information across node pairs).

Although the betweenness centrality calculation appears to be computationally intensive since all pairs of shortest paths must be computed (typically $\theta(n^3)$), Brandes presents a fast technique to compute betweenness centrality that runs in $O(VE)$ time and uses $O(V + E)$ space for undirected unweighted graphs with V nodes and E edges [5].

2.3.2 Egocentric betweenness centrality

A more computationally efficient approach is to calculate betweenness on the ego network as opposed to the global network topology. In social networks, egocentric networks are defined as networks of a single actor together with the actors they are directly connected to, that is, their neighbors. Thus, for wireless mesh networks we need to calculate betweenness on the one-hop adjacency matrix of a node. This metric can be calculated in a distributed manner and is called egocentric betweenness.

Marsden [12] discovered empirically that egocentric betweenness values have a strong positive correlation to sociocentric betweenness values (calculated on the complete network graph) for many different network examples. Everett and Borgatti [9] also present a similar conclusion that the two metrics are strongly correlated for most networks. The authors also provide a few synthetic

examples where egocentric and sociocentric betweenness values do not have a positive correlation [9].

Daly and Haahr [8] recently applied egocentric betweenness centrality as the basis for a distributed routing protocol in a delay tolerant network. Our approach used to calculate LBC is inspired by the work of Marsden and the recent work by Daly and Haahr.

2.4 Summary

It is important to remember that centrality measures can only provide *relative* measures that can be used to compare nodes against each other at that instant of time for a specific network topology. This ranking may allow a system administrator to prioritize management tasks on several nodes, such as deciding which nodes should be patched first and in which order.

3 Bridging Centrality

Bridging Centrality is a relatively new centrality metric. It was introduced in 2006 by Hwang et al. [11]. Bridging centrality can help discriminate bridging nodes, that is, nodes with higher information flow through them, and locations between highly connected regions (assuming a uniform distribution of flows).

The Bridging Centrality of a node is the product of its sociocentric betweenness centrality C_{Soc} and its bridging coefficient $\beta(v)$. The Bridging Centrality $BC(v)$ for a node v of interest is thus defined as:

$$BC(v) = C_{Soc}(v) \times \beta(v) \tag{4}$$

The bridging coefficient of a node describes how well the node is located between high-degree nodes. The bridging coefficient of a node v is thus defined as:

$$\beta(v) = \frac{\frac{1}{d(v)}}{\sum_{i \in N(v)} \frac{1}{d(i)}} \tag{5}$$

where $d(v)$ is the degree of node v, and $N(v)$ is the set of neighbors of node v.

According to the authors,"Betweenness centrality decides only the extent how important the node of interest is from information flow standpoint, but it does not consider the topological

locations of the node. On the other hand, bridging coefficient measures only the extent how well the node is located between highly connected regions, and it does not deliberate the nodes importance from information flow standpoint. Bridging nodes should be positioned between modules and also located on important positions in information flow standpoint. Thus Bridging Centrality combines these two distinct metrics, giving equal weight to both factors." [11]

Based on their empirical studies, the authors recommend labeling the top 25th percentile of nodes as ranked through Bridging Centrality as "bridging nodes"; nodes that are more bridge-like and lie between different connected modules. The authors present results on which nodes are selected by this metric for different networks, and study the impact of removing the highest-ranked bridging nodes from a yeast metabolic network with 359 nodes and 435 edges, as measured by changes in the clustering coefficient, average path length and number of singletons generated.

An alternative definition for the bridging coefficient is to use eigenvector centrality as a substitute for degree centrality in both the numerator and denominator. The disadvantage is the high computational cost of calculating eigenvector centrality and its lack of a distributed alternative. We may explore this technique in future work.

4 Localized Bridging Centrality

We introduce a variant of Bridging Centrality that we call Localized Bridging Centrality (LBC). As the name suggests, we define $LBC(v)$ of a node v using only local information, as the product of egocentric betweenness centrality $C_{Ego}(v)$ and its bridging coefficient $\beta(v)$. LBC is thus represented symbolically as:

$$LBC(v) = C_{Ego}(v) \times \beta(v) \tag{6}$$

The benefit of our approach is that LBC is computationally easier to calculate than BC, and can be calculated in a parallel or distributed manner. Secondly as shown by Marsden [12], and by Everett and Borgatti [9], there is a strong correlation between egocentric betweenness and global betweenness values for most networks, so LBC values should correlate well with BC values. Indeed, our results in Section 5 show this to be the case. While individual nodes can calculate their own

LBC metric in a fully distributed manner, in order to to determine the global rank of each node, a central node must aggregate all LBC values or all nodes must use a distributed consensus-based ranking algorithm.

We explore the utility of the LBC metric in our evaluation. As with the Bridging Centrality metric, the LBC metric can help the system administrator identify clusters, their boundaries and the bridging nodes in the mesh network. These bridging nodes (which are different from the articulation points in a topological analysis) provide the system administrator with prioritized set of nodes to monitor from a robustness perspective.

5 Evaluation

We present our initial results from the application of the BC and LBC metrics on three distinct networks. The first example is a synthetic network, the second is a social network and the third the topology of a wireless mesh network we deployed in our department. All calculations were verified using a popular social-network analysis tool called UCINET [4]. Two or more nodes with the same centrality value were assigned the same rank.

5.1 Synthetic network example

We first tested our metric using a synthetic network example presented in Figure 2. This network was also used by Hwang et al. [11]. The rankings produced by Bridging Centrality and Localized Bridging Centrality shown in Table 1 are nearly identical, although we note that the BC and LBC values are clearly not identical and nor are the betweenness measures used. Since both BC and LBC are used as a "relative" measure of how nodes differ from each other, the induced ranking is more important than the magnitude of the BC or LBC value and thus in this example our metric is equivalent to BC.

5.2 Social-network example

This example (presented in Figure 3) represents game-playing relationships in a bank wiring room and is popular in social-network studies. Marsden [12] presented this example to show how so-

Figure 2: A small synthetic network example. Top 6 high bridging score (BC) nodes are shaded [11]

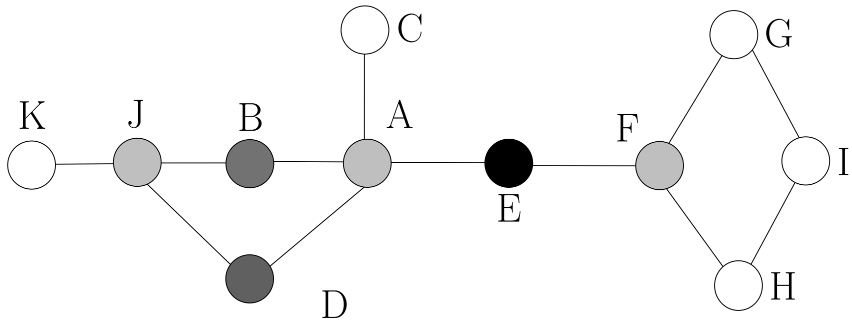


Table 1: Top six centrality values for Figure 2, including Sociocentric Betweenness (C_{Soc}), Egocentric Betweenness (C_{Ego}), Bridging Coefficient (β), Bridging Centrality (BC) and Localized Bridging Centrality (LBC)

Node	Degree	C_{Soc}	C_{Ego}	β	BC	LBC	Rank of BC	Rank of LBC
E	2	0.533	1	0.857	0.857	0.457	1	1
B	2	0.155	1	0.857	0.133	0.857	2	1
D	2	0.155	1	0.857	0.133	0.857	2	1
F	3	0.477	3	0.222	0.106	0.666	4	4
A	4	0.655	6	0.100	0.065	0.600	5	5
J	3	0.211	3	0.166	0.035	0.499	6	6

ciocentric and egocentric betweenness measures correlate. Again the relative ranking of nodes calculated by BC and LBC as shown in Table 2 are identical. Visible inspection of Figure 3 shows that nodes W5 and W7 are bridging nodes, and the tie between them is a bridge between two connected components.

5.3 Real-world mesh network example

We applied our LBC metric on the topology derived from a live wireless mesh network we have deployed in our department. The mesh nodes use the Optimized Link State Routing (OLSR) [7] mesh routing protocol implemented by Tonnesen [13] on Linux. The topology of the network is shown in Figure 4. The ovals and rectangles represent mesh nodes identified by their individual IP addresses. The diamond box is a virtual node representing the Internet. Thus nodes 192.168.1.50 and 192.168.1.20 are Internet Gateways. The BC and LBC results are presented in Table 3 and

Figure 3: Bank wiring room games example [12]

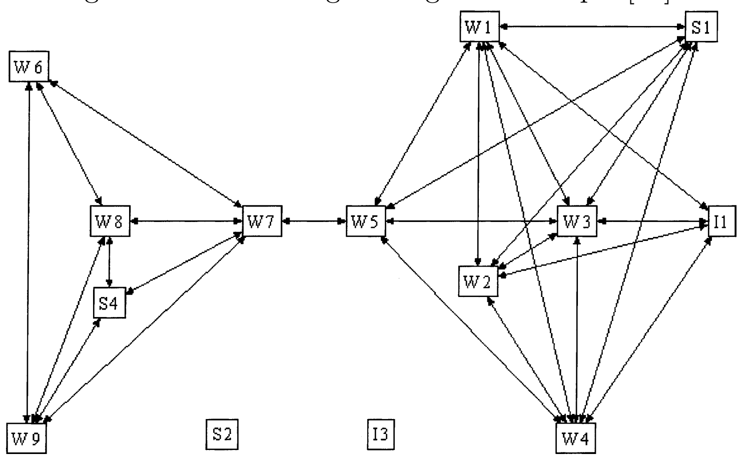
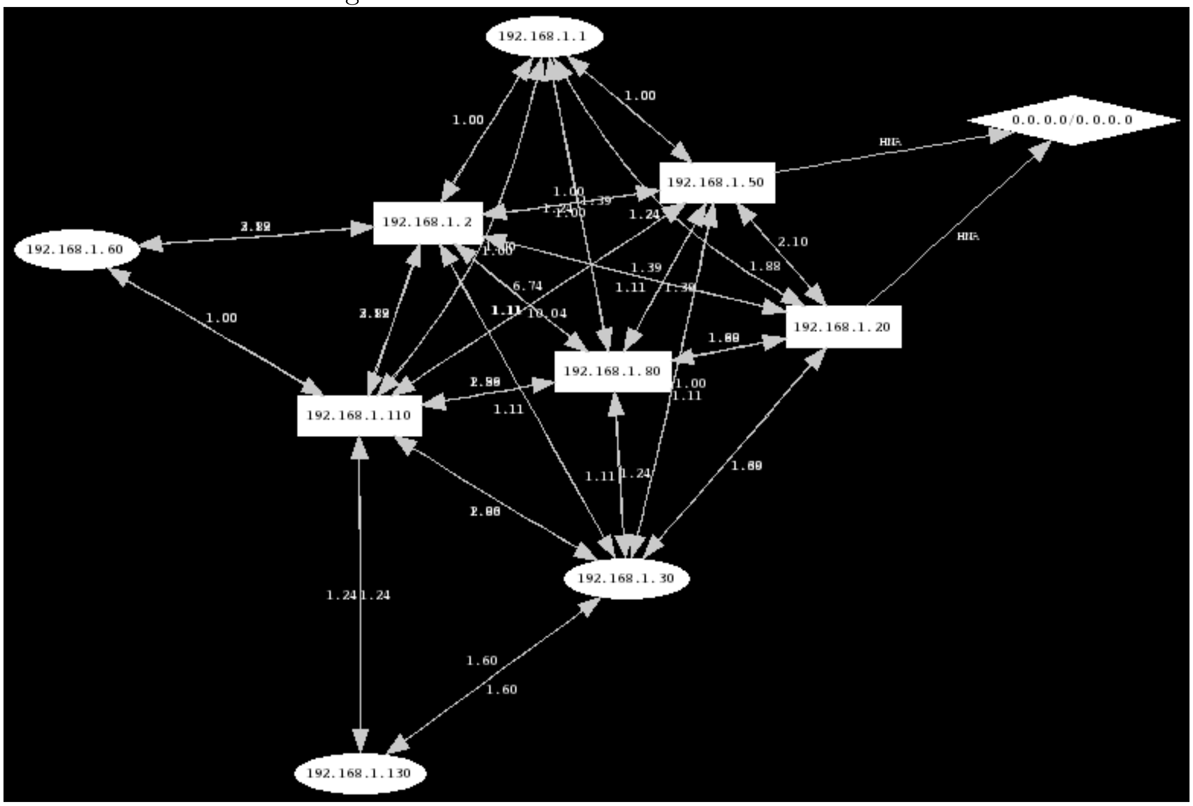


Table 2: Centrality values for Figure 3 sorted by BC values

Node	Degree	C_{Soc}	C_{Ego}	β	BC	LBC	Rank of BC	Rank of LBC
W5	5	30	4	0.222	6.667	0.889	1	1
W7	5	28.33	4.33	0.179	5.074	0.775	2	2
W1	6	3.75	0.83	0.140	0.528	0.117	3	3
W3	6	3.75	0.83	0.140	0.528	0.117	3	3
W4	6	3.75	0.83	0.140	0.528	0.117	3	3
S1	5	1.5	0.25	0.222	0.333	0.055	6	6
W8	4	0.33	0.33	0.223	0.073	0.073	7	7
W9	4	0.33	0.33	0.223	0.073	0.073	7	7
W2	5	0.25	0.25	0.210	0.052	0.052	9	9
W6	3	0	0	0.476	0	0	10	10
S4	3	0	0	0.476	0	0	10	10
I1	4	0	0	0.357	0	0	10	10
I3	0	0	0	0	0	0	10	10
S2	0	0	0	0	0	0	10	10

the nodes are sorted in decreasing order by BC values.

Figure 4: A small real-world mesh network



While in this example the two rankings produced by the two metrics are not identical, they are quite close. The top 5 ranked nodes are common to both metrics and if we remove any of these bridging nodes, then at least one of the other bridging nodes becomes an articulation point, so if that node is now removed, we will have a network partition. However, if you analyze the original network graph in Figure 4, you will find that it is fully connected and has no articulation points. Thus LBC can help detect nodes that may not presently be articulation points but with certain perturbations in the network are most likely to become articulation points. Our LBC metric allows the system administrator to gather this information using few computational resources in a distributed manner.

Although LBC classifies nodes 192.168.1.110, 192.168.1.30 and 192.168.1.2 (co-ranked with 192.168.1.50) as its top three bridging nodes (using the top 25th percentile rule), and BC classifies 192.168.1.110, 192.168.1.50 and 192.168.1.30 as its top three bridging nodes, qualitatively

Table 3: Ranked centrality values for Figure 4, sorted by BC values

Node IP	Degree	C_{Soc}	C_{Ego}	β	BC	LBC	Rank of BC	Rank of LBC
192.168.1.110	7	6.367	5.75	0.078	0.496	0.4485	1	1
192.168.1.50	7	4.733	3.4	0.096	0.454	0.326	2	3
192.168.1.30	6	3.367	2.75	0.132	0.444	0.363	3	2
192.168.1.2	7	4.067	3.4	0.096	0.391	0.326	4	3
192.168.1.20	6	2.867	2.25	0.126	0.361	0.283	5	5
192.168.1.80	6	0.4	0.4	0.173	0.069	0.069	6	6
192.168.1.1	5	0.2	0.25	0.262	0.052	0.065	7	7
192.168.1.60	2	0	0	1.615	0	0	8	8
192.168.1.130	2	0	0	1.65	0	0	8	8
0.0.0.0	2	0	0	1.615	0	0	8	8

there is little difference between the choices, since all of these nodes lie equally on the boundaries between connected components and removal of any of these nodes will leave some other node as an articulation point.

6 Conclusion

In this paper we introduce a new centrality metric called the Localized Bridging Centrality. Our initial investigation indicates that the utility of LBC is equivalent to that of the Bridging Centrality metric. Our LBC metric is easy to compute and is designed to be computed in a distributed manner. We have demonstrated the usefulness of our metric in identifying critical bridging nodes in a wireless mesh network from a network management perspective. We note just one potential drawback, that our metric may not work well in cases where the egocentric and sociocentric betweenness values do not correlate. We are in the process of testing the properties of this metric on larger data sets and exploring its utility in other scenarios.

Acknowledgment

This research program is a part of the Institute for Security Technology Studies, supported by a gift from Intel Corporation, by Award number 2000-DT-CX-K001 from the U.S. Department of

Homeland Security (Science and Technology Directorate) and by Grant number 2005-DD-BX-1091 awarded by the Bureau of Justice Assistance. Points of view in this document are those of the authors, and do not necessarily represent the official position or policies of any of the sponsors. The authors wish to thank Cindy Torres for her help in proofreading this paper.

References

[1] K. Begnum and M. Burgess. Principle Components and Importance Ranking of Distributed Anomalies. *Machine Learning*, 58(2):217–230, 2005.

[2] P. Bonacich. Power and Centrality: A Family of Measures. *The American Journal of Sociology*, 92(5):1170–1182, 1987.

[3] P. Bonacich and P. Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23(3):191–201, 2001.

[4] S.P. Borgatti, M.G. Everett, and L.C. Freeman. UCINET for Windows: Software for Social Network Analysis. *Harvard: Analytic Technologies*, 2002.

[5] U. Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.

[6] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, 1998.

[7] T. Clausen and P. Jacquet. Optimized Link State Routing Protocol (OLSR). RFC 3626 (Experimental), October 2003.

[8] Elizabeth M. Daly and Mads Haahr. Social network analysis for routing in disconnected delay-tolerant MANETs. In *MobiHoc '07: Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 32–40, New York, NY, USA, 2007. ACM.

[9] M. Everett and S.P. Borgatti. Ego network betweenness. *Social Networks*, 27(1):31–38, 2005.

[10] L.C. Freeman. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1):35–41, 1977.

[11] W. Hwang, Y. Cho, A. Zhang, and M. Ramanathan. Bridging Centrality: Identifying Bridging Nodes in Scale-free Networks. Technical Report 2006-05, Department of Computer Science and Engineering, University at Buffalo, March 15 2006.

[12] P.V. Marsden. Egocentric and sociocentric measures of network centrality. *Social Networks*, 24(4):407–422, 2002.

[13] Andreas Tønnesen. Implementing and extending the Optimized Link State Routing protocol. Master's thesis, UniK - University Graduate Center, 2004.