An Improved Local Bridging Centrality Model for Distributed Network Analytics

Joseph P. Macker*

*Information Technology Division, Naval Research Laboratory, Washington DC Email: joseph.macker@nrl.navy.mil

Abstract-Classifying an edge or a node within a network graph according to its bridging characteristics is an important structural measure that has many practical analytic applications. Bridging centrality is a relatively new graph-based centrality metric for classifying nodes serving as key structural connections between dense components. A global bridging centrality has been previously introduced and requires global knowledge of the entire graph structure to perform the centrality computation. Recently, a local bridging centrality variant was also introduced that can be calculated locally requiring only limited local neighborhood graph information. We introduce an extended model of the recently introduced local bridging centrality based upon the the concept of a local "friends of friends" or 2-hop neighbor distance egocentric graph. We further develop and analyze the use of this extended centrality with both unweighted and weighted graph structures. Finally we present a series of comparative studies using a variety of graph models and examine ranking correlation comparisons with global bridging centrality results. Our analysis includes comparisons with both past literature models and newer temporal graph results of a 100-node mobile wireless network scenario with link weights representing dynamic reception quality.

I. INTRODUCTION

Complex network theory measures include the broad use and definition of a variety of centrality metrics. Centralities represent the statistical ranking of the importance or influence of vertices (i.e., nodes) or edges within a network graph based upon a particular structural or interaction model [7], [19]. Many centrality measures have a strong relationship to statistical mechanics models and are useful in predicting forms of behavior or dynamic performance of network node interaction and related traffic flows.

Bridging centrality was first introduced in [11] as means to identify nodes acting as significant *structural bridges* between dense areas of a network. Understanding such statistical relationships is an important analytical concept in applied areas of complex graph analysis. As one example, chemical biology has been invigorated by the application of such structural centralities to investigate diseases and related treatment responses when modeling such interactions as graphs [10]. Within our focus, communication network structures, the knowledge of important bridging nodes and edges helps one better determine key nodes statistically residing between existing network clusters. Clusters are defined here as areas of higher neighborhood density or affinity. Bridging characteristics of a network also help predict areas of significant potential congestion and identify key features related to structural robustness against network failures or attacks. The latter application is close to the rationale behind the use of bridging centrality measures in other disciplines, such as the study of biomedical interactions or in the study of network pharmacology to better understand the efficacy of drug treatments for a variety of diseases.

Because of the distributed nature of wireless mobile communication networks, we are often motivated to investigate localized or distributed calculations of structural metrics while realizing there exists a tradeoff between measurement accuracy, computation, and communications. Localized centrality calculations can reduce both communication and computation complexity and there are a myriad of potential applications. We are mostly interested in the application of these metrics to distributed communication systems, a related past work [18] demonstrated the potential for such approaches to assist in the operations and analysis of mobile ad hoc networks by influencing distributed forwarding control and election mechanisms.

We first present our extension to an existing local algorithm variant and then presents empirical measurements and ranking correlations between local and global estimation for a variety of scenarios. We also present and discuss important considerations and issues related to weighted graph application of the estimation.

II. BRIDGING CENTRALITY

So what does bridging mean and how is it defined? Bridging nodes within network graphs are described as nodes relatively more important in connecting and providing effective pathways for separate dense components within the graph. Centralities are relative statistical measures of certain properties within a given network graph structure, bridging being one such potential structural property. The global bridging centrality, BC, of a node was more formally defined in [11] and is formed as the product of the betweenness centrality, C_{bet} , and a term called the bridging coefficient, β_c .

$$BC = C_{bet} * \beta_c \tag{1}$$

 C_{bet} takes into account global features of a nodes' topological role in terms of how many shortest path flows it resides on between other node pairs, while the coefficient B_c takes into account local bridging structural features of the node (i.e., its function in providing connections between locally dense portions of a graph within an immediate neighborhood).

A. Betweenness Centrality

The basic Betweenness Centrality is defined in Equation 2.

$$C_{bet}(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$
(2)

In Equation 2, σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those shortest paths that pass through v. Betweenness centrality is an extremely popular complex network measure and its use as a statistical contributor in the bridging centrality calculation captures one dimension of the statistical flow properties of a node. One can easily imagine different flow models providing alternative statistical input at this point (e.g., a multipath current centrality model), but we will limit our present comparison and development to this presently defined bridging centrality model from [11].

B. Bridging Coefficient

The bridging coefficient, β_c , introduced in [11] is defined as shown in Equation 3.

$$\beta_c = \frac{\frac{1}{d(v)}}{\sum_{i \in N(v)} \frac{1}{d(i)}} \tag{3}$$

In Equation 3, d(v) is the degree of node v, and N(v) is the set of graph neighbors of node v.

In this way, the bridging coefficient attempts to model how well situated a node is between other high degree nodes within a localized topology.

III. PAST WORK ON EGO NETWORKS AND BRIDGING CENTRALITY

Egocentric network analysis focuses on something termed the *ego network* which is based upon a single node, the *ego*, and its direct neighbors or *alters*. The analysis of ego networks and their various applications was first explored in Ron Burt's 1992 "*Structural Holes: The Social Structure of Competition*" [3]. Later Everett and Borgatti in [8] developed from this a concept of egocentric betweenness centrality. A synthetic network used as a model in both [8] and in the original bridging centrality paper [11], is shown in Figure 1.



Fig. 1: Synthetic Network Example

An ego network is defined as the subgraph made up of a node itself (i.e., ego node) and its alters, or direct neighbors, and direct connections between alters. As an example, node *F*'s ego network is extracted and presented in Figure 2. In this case, *F* has a high egocentric betweenness because all shortest path pairs between *alters* contain *F*. If we added an additional edge in the ego network between nodes *G* and *H*, the relative egocentric betweenness of *F* decreases because *F* is no longer on the shortest path between some endpoint pairs (H,G)



Fig. 2: Node F Egocentric Network

A. Past Work on Local Bridging Centrality

Egocentric betweenness in [8] was combined with the concepts of bridging centrality from [11] to develop a local bridging centrality measure by Nanda, et al in [17]. Local bridging centrality is a basic follow-on to the method for global bridging centrality calculation introduced in Section II. The basic equation for calculation is shown as Equation 4, where C_{lbet} represents a localized egocentric betweenness measure rather than the global betweenness measure.

$$LBC = C_{lbet} * \beta_c \tag{4}$$

The concept for calculating C_{lbet} presented in [17] is computationally simplified over the calculation of global betweenness centrality and only requires the 1-hop adjacency matrix. Efficient means for calculating C_{lbet} were also discussed in [5], [16].

Analyses of multiple example graph datasets in [5], [16] demonstrated some reasonable correlation between *local bridging centrality* and *global bridging centrality* for the networks presented and studied. These results provided motivation for the use of *local bridging centrality* for two potential reasons:

- 1) local estimation with limited global information
- reduction in the computation complexity of bridging centrality estimates

In distributed wireless and other highly dynamic systems, we are often interested in structural algorithms with fast convergence using local estimators. Such approaches can be used for local real-time analytics and for potentially improving self-organization and routing. Computational improvements are also of interest in more general analytic applications. Reducing the computational impact of a global betweenness calculation is one potential contribution of a localized estimation approach. Freeman's betweenness centrality [9], a basic contributor to the global bridging centrality calculation, is calculable on the order of O(mn) for unweighted graphs and $O(nm + n^2 logn)$ for weighted graphs [2], where n is the number of nodes and m is the number of edges. For fully

connected networks $m \sim n^2$ but may be much smaller for sparser networks.

IV. EXTENDING THE EGO NETWORK CONSTRUCT FOR USE IN WIRELESS NETWORK TOPOLOGIES

When applying ego network concepts to distributed wireless networks or mobile ad hoc network (manet) use as suggested in [17], we consider extending the adopted betweenness model to include additional local information often available within such networks. Upon consideration, it turns out this same information is also often available in many social network datasets. For instance, most manet or wireless ad hoc routing protocols include a local discovery protocol mechanism that collects 2-hop topology information at each node via local neighbor discovery messaging [1], [4], [13], [20]. Neighbors include information about their neighbors in such discovery messages. Sociologically this can also be thought of as having a dynamic model of an egocentric friends of friends network. Going back to the example from Section III we now construct and provide an example of the 2-hop egocentric network for node F as shown in Figure 3.



Fig. 3: Node F Extended Egocentric Network

We define our version of local betweenness on this 2hop egocentric network as C_{lbet2} . We show the calculation of 1-hop and 2-hop egocentric betweenness for the synthetic reference network in Table I. We can see from the results that C_{lbet2} reveals more statistical details of relative topological flow structure. Of special note are the differences for nodes (B, E, D, I, J) between C_{lbet} and C_{lbet2} . The relative structural importance of E is completely missed within the 1-hop egocentric model. What is important to note when examining the local centrality results is not actual values but the relative rankings of the values of betweenness. In C_{lbet} there are only 4 relative levels of ranking for this unweighted network model. Whereas, with C_{lbet2} there are 8 relative ranking levels. It should be pointed out, as discussed in [8], we do not apply betweenness normalization to our local betweenness calculations.

V. EXTENDED LOCAL BRIDGING CENTRALITY

In a straightforward manner, we modify the previous local bridging centrality formula LBC to include the use of the 2-hop ego network betweenness, C_{lbet2} . We call this 2-hop local bridging centrality, LBC2, and equation 5 shows its formulation.

FABLE	I:	Synthetic	Net	C_{lbet}	and	Clbet?
						- 10001/2

Name	C_{lbet}	C_{lbet2}
А	6	11.5
С	0	0.0
В	1	3.0
Е	1	12.0
D	1	3.0
G	1	1.0
F	3	6.5
Ι	1	0.5
Н	1	1.0
Κ	0	0.0
J	3	3.5

$$LBC2 = C_{lbet2} * \beta_c \tag{5}$$

VI. COMPARISON TO PREVIOUS RESULTS AND LARGER STATIC MODELS

Prior to doing more complex network studies, we compare results for *LBC*2 against the results for the simple example networks presented in the reference work [17]. We include the small synthetic network, the small bank wiring example, and our own larger interconnected gaussian cluster examples (5-Gauss-100 and 10-Gauss-200). Our interconnected 5-Gauss-100 example is a 100 node network with 5 gaussian clusters with an interconnection ratio of edges within a cluster to interconnecting edges set at 0.9. An example generated network is depicted in Figure 4. Similarly, 10-Gauss-200 is a 200 node network with 10 gaussian clusters with the same 0.9 interconnection ratio as 5-gauss-100. This defines a moderately sized network with some amount of *bridging* nodes with edges that interconnect more densely connected clusters.



Fig. 4: 100 Node 5 Gaussian Cluster Network

We present the results of our initial study in Table II. We took each unweighted network model and calculated the global bridging centrality BC, the local bridging centrality, LBC, and the 2-hop extended local bridging centrality, LBC2. We present Pearson, Spearman, and Kendall [12] correlation between each local bridging centrality and the global bridging centrality results. To compare the classification of bridging features, relative rankings are most important so we include Spearman and Kendall coefficients along with typical Pearson coefficients for correlating data sets. Spearman and Kendall measures are intended specifically for correlation comparison of relative rankings. From examining the results presented in Table II, we note that while the Pearson correlation results remain relatively the same between techniques we see improvement in the ranking correlation (both Spearman and Kendall scores) for LBC2 as compared with LBC. This is most notable in the referenced synthetic and clustered Gaussian network cases.

TABLE II. <i>EDC</i> , <i>EDC</i> 2 to <i>DC</i> conclution measures									
Correlation Graph	$\begin{array}{c} \text{Pearson} \\ LBCp \end{array}$	LBC2p	Spearman LBCs	LBC2s	Kendall $LBCk$	LBC2k			
Synthetic Bank Wiring 5-Gauss-100 10-Gauss-200	0.98 1.0 0.99 0.78	0.99 1.0 0.99 0.64	0.62 0.98 0.95 0.90	0.99 1.0 0.99 0.94	0.53 0.92 0.78 0.76	0.65 0.96 0.85 0.81			

TABLE II: LBC,LBC2 to BC Correlation Measures

VII. NOTES ON USE WITH WEIGHTED GRAPH MODELS

We will now discuss use of localized bridging centrality with weighted graph models and to our knowledge some of these important issues have not been presented in past bridging centrality literature. By re-examining the original formulation of bridging centrality derived from [11], one can see it is missing important applied discussion when considering its use in weighted graph models. A key technical issue is the fact that graph edge weights or costs often need to be represented mathematically in different ways depending upon the purpose of the metric application. For example additive edge metrics appropriate for classical shortest cost path computation, C_{bet} , should not be used for weighted degree calculations that would contribute to the bridging coefficient, β_c in this case. These metrics often have inverse relationships in that smaller weights representing poor edge quality are equivalent to a higher weights appropriate for least cost path purposes. If link cost metrics were used for bridging coefficient calculation, as defined, a single high cost edge (e.g., linking two neighbors) could make it appear that there were large clustered areas surrounding a node, thus inflating bridging coefficient values artificially. In the case of uniform weights (e.g., weight=1) this is a non-issue.

In our approach, we will apply a floating point edge quality weight as shown in Equation 6 so that the edge weights, E_{vi} , are between the values (0,1) and represent some notion of edge quality not cost for use in bridging coefficient calculation and will use a related additive cost metric for betweenness we introduce in Section VIII.

$$Deg(v) = \sum_{i \in N_v} E_{vi} \le v_{neigh} \tag{6}$$

VIII. DYNAMIC MOBILE NETWORK MODEL

Encouraged by our initial reference model findings in Table II, we now investigate a larger and more complex set of weighted networks. For the next set of analyses we apply an existing temporal mobile wireless communication network model. A snapshot topology of the modeled JavaSea (JS) network scenario we will be using is shown in Figure 5. This network model was previously introduced and discussed in detail for graph-based temporal clustering and routing path studies within [14], [15], [21]. In Figure 5, the upper right cluster represents mobile nodes within a littoral region and the lower left region represents nodes moving with a sea-based region. There are nodes performing aerial flight patterns amongst these areas throughout the scenario. The average degree-based assortativity coefficient, r, for the JS scenario is r = 0.57, meaning vertex degree distributions are reasonably assortative and because of the maritime, littoral, and airborne nodes there are clearly nodes at various times with bridging functionality between dense areas of the scenario. Due to mobility and link quality variations, network structural characteristics modulate throughout the scenario.



Fig. 5: JavaSea Scenario Snapshot

To capture the temporal nature of the network, we model JS as a time series of weighted graphs G = (V, E, w, t), where t is time or the sequence value of our data set, V is the set of vertices (representing the data or nodes in the network) and E is the set of edges connecting the vertices in V. wrepresents a set of edge weights in E that may be asymmetric. In our analysis, w represents a stochastic probability of packet reception given a wireless link quality and mobility model. The weighted graph betweenness calculation required for bridging centrality needs an additive cost metric within the weighted graph model so we convert the neighbor link quality-based asymmetric w weights (i.e., packet reception probability) in the original JS model to an additive cost metric. We chose to use the well-cited expected transmission count (ETX) metric [6] as the additive cost metric for the dynamic edges within our scenario. Prior to analysis, we filter the graph edge quality retaining only edges with a minimum probability of successful reception threshold ≥ 0.8 in both directions. This models the neighbor adjacency establishment of typical routing protocols through localized neighborhood protocol exchanges. We then convert the asymmetric probability of successful packet reception weights (calculated from range and fading models) in E to a set of ETX weights for edge in the graph. While details and dynamics of the overall model are interesting, we are most interested in using it as an example of a complex, weighted temporal graph model relevant to potential real world dynamic scenarios.

For each temporal graph, we calculate local and global bridging centrality ranking measures for both weighted and unweighted cases using the JS scenario temporal graph data. We analyze a series of one second temporal snapshots of the scenario over a total of 600 seconds. For the weighted case, we use ETX for betweenness cost and a packet reception probability metric for bridging coefficient calculations. The unweighted case treats all edge weights as unit value = 1. We then compute ranking correlations over the time series for each of the local bridging centrality values, (LBC, LBC2), and the global bridging centrality values, BC.

The weighted graph results are shown in Figures (6,7) the correlation between local and global bridging centrality is always positive for these trials and the *LBC*2 model improves the ranking correlation results over *LBC* for all of the temporal graph snapshots analyzed. Pearson correlation results are removed as ranking correlation is more important for applied purposes of classifying and identifying structural characteristics.



Fig. 6: JavaSea Spearman Rank Correlation



Fig. 7: JavaSea Kendall Rank Correlation

The unweighted graph results demonstrated similar trends and results to the weighted graph studies although rank correlation values tended to be slightly higher in general. Again the LBC2 ranking correlations were improved over LBC across the temporal graph snapshots studied. As an illustration of unweighted results, we present the Spearman correlation in Figure 8 and one can compare this against the weighted graph results from Figure 6.



Fig. 8: JavaSea Spearman Rank Correlation

IX. LBC2 vs. LBC RANKING CORRELATION IMPROVEMENT RESULTS

We present the comparison more directly by plotting the global BC correlation improvement to LBC2 vs. LBC. Figure 9 plots the rank correlation value increase of LBC2 over LBC for both Spearman and Kendall coefficients over the entire temporal scenario. Both the Spearman and Kendall results show very similar trends in this regard. Also, while rank correlations fluctuate throughout the 600 temporal graph experiment the Spearman coefficient for LBC2 to BC stays positive within the value range [0.6-0.9]. The average and standard deviation of the Spearman coefficient between LBC2 and BC for weighted JS is 0.93 ± 0.02 .



Fig. 9: LBC2 vs. LBC Rank Correlation Improvement

While we performed a general ranking correlation measure between local centralities, *LBC* and *LBC2*, and a global centrality (*BC*), it is important to compare results for the highest ranking results. Therefore, we examined a running count of the top 10% of centrality values throughput the entire scenario for BC, LBC, and LBC2. We found that LBC had 6 nodes in common with BC's top 10 scoring nodes throughout the scenario while LBC2 had a slight increase with 7 nodes in common with BC's top 10. Looking even closer at the top 3 node results, there are 2 in common between LBC2 and BC, while zero are in common for LBC and BC. While not a comprehensive study, this provides some indication of LBC2 ability to improve the identification of global bridging characteristics using the extended local estimation model. In many practical applications, identifying the highest ranking percentage of bridging nodes helps identify a community of nodes for some specialized purpose in network management, routing, and topological optimization (e.g., cacheing, prioritized forwarding).

X. ISSUES AND FUTURE WORK

Bridging centrality was initially developed using shortest path betweenness as a main contributor. Since wireless network topologies and applications may use multiple paths simultaneously to communicate flows, a shortest path model may not be the most optimal model for some network data flow models or network deployment scenarios. We are interested in exploring the extension of the bridging centrality models to include other flow-based centralities (e.g., current or communicability-based centralities).

Other areas we plan to address include the potential for using weighted, directed network models more effectively and understanding the relationship between localized bridging centrality models and manet-based distributed control structures and techniques. It is expected that the concept of localized bridging centrality can improve existing distributed relay or optimization algorithms. Self-electing algorithms in use with manet routing control and management, such as the essential connected dominating set (E-CDS) algorithm, are basically forms of egocentric ranking methods using limited local topology knowledge. Similar improvements for ad hoc routing with data flow modeling extensions were demonstrated using LBC and Localized Load-aware Bridging Centrality (LLBC) accounting for some traffic flow loading in [18].

XI. CONCLUSIONS

We have developed and analyzed a recommended extension to local bridging centrality (LBC) called LBC2 for use in distributed wireless network analytics. Within manet and other self-organizing networks 2-hop neighbor topology information is often available at each network node through localized neighbor message exchanges, so it is strongly justified to extend the idea of an alters-only egocentric betweenness in [17] to a 2-hop neighbor betweenness calculation. We described using this new formulation to develop a 2-hop local bridging centrality, LBC2, that can be generally applied when 2-hop topology knowledge or "friends of friends" connection information is available. Comparing against past simple toy networks presented in previous related work and a basic gaussian clustered network model we first showed that LBC2 improved some ranking correlations with global bridging centrality over the previous egocentric local bridging centrality, LBC.

We next discussed issues in calculating bridging centrality for weighted graphs and presented a working model. Next, we used an existing 100-node temporal network model to compare approaches including link quality weights and we showed that LBC2 provided significant improved ranking correlation over LBC for all sets of temporal graphs analyzed.

From this initial work, we demonstrated that application of localized bridging centrality to a distributed wireless network model benefitted significantly from local 2-hop egotistical topology knowledge often available to local nodes in such wireless networks. A similar approach is generally applicable to k-hop egocentric network analysis within general network models but the complexity of communications and computation will increase for larger k-hop variants and may provide limited gains in accuracy.

REFERENCES

- S. Basagni, M. Conti, S. Giordano, and I. Stojmenović. Mobile Ad Hoc Networking: Edited by Stefano Basagni...[et al.]. IEEE, 2004.
- [2] Ulrik Brandes. A faster algorithm for betweenness centrality*. Journal of Mathematical Sociology, 25(2):163–177, 2001.
- [3] Ronald S Burt. Structural Holes: The Social Structure of Competition. Harvard University Press, 2009.
- [4] T. Clausen and P. Jacquet. Optimized Link State Routing Protocol (OLSR), RFC 3626. Internet Engineering Task Force (IETF), October 2003.
- [5] Elizabeth M Daly and Mads Haahr. Social Network Analysis for Routing in Disconnected Delay-tolerant Manets. In *Proceedings of the 8th ACM international symposium on Mobile Ad hoc Networking and Computing*, pages 32–40. ACM, 2007.
- [6] Douglas SJ De Couto, Daniel Aguayo, John Bicket, and Robert Morris. A High-throughput Path Metric for Multi-hop Wireless Routing. *Wireless Networks*, 11(4):419–434, 2005.
- [7] Ernesto Estrada. The Structure of Complex Networks: Theory and Applications. Oxford University Press, 2011.
- [8] Martin Everett and Stephen P Borgatti. Ego Network Betweenness. Social Networks, 27(1):31–38, 2005.
- [9] Linton C Freeman. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, pages 35–41, 1977.
- [10] Andrew L Hopkins. Network Pharmacology: The Next Paradigm in Drug Discovery. *Nature Chemical Biology*, 4(11):682–690, 2008.
- [11] Woochang Hwang, Young-rae Cho, Aidong Zhang, and Murali Ramanathan. Bridging Centrality: Identifying Bridging Nodes in Scalefree Networks. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 20–23, 2006.
- [12] Maurice George Kendall. Rank Correlation Methods. Griffin, 1948.
- [13] Joseph P Macker. Simplified Multicast Forwarding, RFC 6621. Internet Engineering Task Force (IETF), 2012.
- [14] Joseph P Macker, David Claypool, and Nathan Hughes. Improving Routing Path Stability in Mobile Ad Hoc Networks That Use a CDS Control Plane. In Proceedings of the 2014 IEEE Military Communications Conference (MILCOM), pages 1099–1104. IEEE, 2014.
- [15] Joseph P Macker and David J Claypool. Dynamic Communities in Evolving Network Graphs. In Proceedings on the IEEE 2012 Military Communications Conference, MILCOM 2012, pages 1–6. IEEE, 2012.
- [16] Peter V Marsden. Egocentric and Sociocentric Measures of Network Centrality. Social Networks, 24(4):407–422, 2002.
- [17] Soumendra Nanda and David Kotz. Localized Bridging Centrality for Distributed Network Analysis. In *Computer Communications and Networks*, 2008. *ICCCN'08.*, pages 1–6. IEEE, 2008.
- [18] Soumendra Nanda and David Kotz. Social Network Analysis Plugin (SNAP) for Mesh Networks. In Wireless Communications and Networking Conference (WCNC), 2011 IEEE, pages 725–730. IEEE, 2011.
- [19] Mark EJ Newman. The Structure and Function of Complex Networks. SIAM Review, 45(2):167–256, 2003.
- [20] R. Ogier and P. Spagnolo. Mobile ad hoc network (MANET) extension of OSPF using connected dominating set (CDS) flooding) flooding. *RFC* 5614, IETF, 2009.
- [21] John M Shea and Joseph P Macker. Automatic selection of number of clusters in networks using relative eigenvalue quality. In *Proceedings of the 2013 IEEE Military Communications Conference, MILCOM 2013.*, pages 131–136. IEEE, 2013.