CBDI: Combined Banzhaf & Diversity Index for Finding Critical Nodes

Waqar Asif*‡, Hassaan Khaliq Qureshi*, Muttukrishnan Rajarajan‡ and Marios Lestas §
‡School of Engineering and Mathematical Sciences, City University, London, UK.
Email: {waqar.asif.1, r.muttukrishnan}@city.ac.uk
* National University of Sciences & Technology (NUST), Islamabad, Pakistan.
Email: {09mscsewasif, hassaan.khaliq}@seecs.edu.pk
§Department of Electrical Engineering, Frederick University, Nicosia, Cyprus.
Email: eng.lm@frederick.ac.cy

***Abstract*—Critical node discovery plays a vital role in assessing the vulnerability of a network to an abrupt change, such as an adversarial attack or human intervention. In this paper, we propose a new metric to characterize the criticality of a node in an arbitrary network which we refer to as the Combined Banzhaf & Diversity Index *(CBDI)*. The metric utilizes a diversity index which is based on the variability of a node's attributes relative to its neighbors and the Banzhaf Power Index which characterizes the degree of participation of a node in forming shortest paths. The Banzhaf power index is inspired from the theory of voting games in game theory. We evaluate the performance of the new metric using simulations. Our results indicate that in a number of network topologies, the proposed metric outperforms other proposals which have appeared in the literature. The proposed *CBDI* index chooses more critical nodes which, when removed, degrade network performance to a greater extent than if critical nodes based on other criticality metrics were removed.**

I. Introduction

Critical node discovery is an important process for understanding network vulnerability. A node is deemed as critical, if it plays a vital role in maintaining network performance and by removing that node, the overall performance deteriorates and in some cases leads to network partitioning [1] which is highly undesirable. Evaluating the criticality of node is significant in various complex networks. In Wireless Sensor Networks (WSNs) employing geographical routing, for example, malicious attack or malfunction of a few beacon nodes leads to fallacious node discovery for the remaining nodes in the network, thus jeopardizing the stable operation of the routing protocol [2]. Similarly, in road networks, intersections which can be considered as nodes in a graph theoretic framework, might experience heavy traffic loads when in proximity to a major landmark. Identifying such critical nodes is significant when investigating possible extensions of the existing infrastructure [3]. Moreover, in power networks some grid stations are considered to be critical as their loss, either due to element failure or an unexpected natural disaster, might lead to a major breakdown of the entire power network.

Several studies have addressed the node criticality problem and various metrics have been proposed to characterize the criticality of a node in the network. The degree centrality metric [4] and the degree of suspected nodes or edges [5] are based on the number of neighbors of each node. The average shortest path length metric in [6] characterizes the criticality of a node by calculating the average shortest path length over all possible node destination paths, while the global clustering coefficient metric [7] evaluates the criticality of a node by weighing its participation in cluster formation. The shortest path is also utilized in [8] where a node is characterized as critical based on the number of times it participates in the shortest paths throughout the network. The participation of a node in path formation is also accounted for in [9] where nodes, the removal of which causes a reduction in the rank of the routing matrix, are considered as critical. Finally, the pairwise connectivity of the network approach in [1], considers node pair properties instead of individual node properties and characterizes as critical, node pairs which contribute mostly to network partitioning when removed.

In this work, we propose a new criticality metric which is shown to be more successful in identifying nodes, the removal of which, significantly affects network operation. The metric encompasses three main node attributes: the weighted node degree, the variation in link length of the node from its neighbors and its contribution in forming shortest paths. Unlike previous proposals which take into account the absolute node degree, in this proposal we consider the node degree weighted by the average common neighbors of the node with all its neighbors. The presence of common neighbors is an indication of the presence of path alternatives which undermine the criticality of a node. In addition, in order to account for long range links which cause nodes to act as relay nodes thus accommodating heavy traffic and becoming critical for the whole network operation, we introduce the notion of the variation in link length between neighboring nodes. The diversity in the number of neighbors and the diversity in link lengths thus contribute to the criticality of a node and are used to form the diversity index. We then account for the contribution of each node in forming the routing paths by employing a new technique which is inspired by voting games in game theory. The metric emanating from this technique is known as the Banzhaf Power index. The combination of the latter with the diversity index yields the proposed criticality metric which we refer to as the *Combined Banzhaf & Diversity Index (CBDI)*.

We evaluate the performance of the proposed metric using simulations. In a number of network topologies, we identify the critical nodes using the criticality metric under consideration and we evaluate the network performance when these nodes are removed. Correct selection of the critical nodes create a significant degradation in network performance. Network performance is measured in terms of the average path length, the average node degree and the number of isolated nodes. We observe that in the scenarios under consideration, the proposed *CBDI* index chooses critical nodes which, when removed, degrade performance to a greater extent than if critical nodes based on other criticality metrics were removed. The proposed metric has been shown to outperform other metrics such as the *Hybrid Interactive Linear Programming Rounding (HILPR)* proposed in [1] and *Controllability of complex networks (Cont)* in [9].

The rest of the paper is organized as follows: in Section II we describe the proposed criticality metric, in Section III we evaluate its performance using simulations and finally in Section IV we offer our conclusions and future research directions.

II. Proposed Criticality Metric

As mentioned in the introduction, in this work, we propose a new criticality metric which is the combination of the Banzhaf power index and the diversity index. In this section, we explain the reasoning behind our design choices and formally define the diversity index and the Banzhaf power index. We then show how we combine the two to form the proposed criticality index.

A. Diversity index

Diversity index is a measure of the variation of node properties between neighboring nodes. We consider variation of two attributes of neighboring nodes which are logically related to their criticality: the variability in link lengths and the variability in their list of neighbors. Increasing both the variability of link lengths and the variability in the list of neighbors implies greater node criticality. Below we give a detailed description of the two and explain how they are combined to form the diversity index.

1) Variation in link length: This attribute measures the variation in the length of the links between neighboring nodes. A greater variation in link length certifies the existence of both long distance and short distance links. A node with the aforementioned property is capable of acting as a relay node between the nodes in proximity and the distant ones. This will aid neighboring nodes in getting their data relayed to distant nodes and vice versa at a reduced network energy and time cost [10]. Since a node with a higher variation in link length has a higher probability of acting as a relay node hence, it is deemed as critical for information dissemination.

We define the variation of link length as the average difference between the transmission radii of neighboring nodes. We assume a graph $G = (V, E)$, where V represents the set of Nodes and E represents the set of Edges. Each node x in V is characterized by its transmission radius T_x. For each node x, the set of nodes which lie within the transmission range of x is the set of its neighbors and is denoted by $N(x)$. The variation in link length of x is denoted by $D_d(x)$ and is given by:

$$D_d(x) = \frac{1}{|N(x)|} \sum_{u \in N(x)} (T_x - T_u) \tag{1}$$

2) Weighted Node Degree: Node degree was used by Freeman in [4] for determining the criticality of a node. Despite the simplicity of the method it fails to take into consideration self loops and one hop reachability of neighboring nodes which leads to overestimates of the node criticality. Therefore, in this work, we avoid the consideration of these redundant paths by elaborating on the variability of the list of neighbors of neighboring nodes, leading to the notion of weighted node degree. The weighted node degree takes values between 0 and 1, and increases as the number of common neighbor decreases. A greater number of common neighbors implies more one hop paths between neighboring nodes which undermines the criticality of a node. The weighted node degree of x is represented by $D_n(x)$ and is given by:

$$D_n(x) = \sum_{u \in N(x)} \frac{|N(u) \backslash N(x)|}{|N(u)|} \tag{2}$$

where $\backslash$ denotes the set difference and $|.|$ denotes the cardinality of the set. So, the weighted node degree of a node x is calculated by summing the dissimilarity ratios of all of its neighbors. The dissimilarity ratio for a particular neighbor u is the ratio of number of neighbors of u which are not neighbors of x over the set of all neighbors of u.

Both the variation in link length and the weighted node degree of a node described above are used to calculate the diversity index of that node. The diversity index $H(x)$ is defined as the product of the two metrics such that:

$$H(x) = D_d(x) D_n(x) \tag{3}$$

If follows from the discussion above that the greater the diversity index, the more critical a node is. The criticality of a node is further refined by weighing its participation in path formation. To this end, we use the Banzhaf power index which is described below.

B. Banzhaf power index

In game theory, different assumptions have led to different definitions for determining the importance of an agent in a game. One of the most prominent among these is the Banzhaf power index [11]. This index has been widely used primarily for the purpose of weighted voting games. In a voting game, each voter is assigned a weight and the coalition of these voters determines the outcome of the game. A game is considered as a winning game, if the sum of all the weights of the nodes in a coalition is greater than or equal to a predefined threshold weight. A node has a pivotal role if its removal transforms a winning game into a loosing game. Nodes with the aforementioned property are called swing nodes. A node that acts as a swing node in maximum coalitions is the most critical node and is assigned the highest Banzhaf power index.

We adapt the above ideas in a communication network setting in order to characterize the criticality of nodes participating in the network. In the same way that weights are being used to select coalitions in a voting game setting, we use the link bandwidths in a communication network setting to select the nodes participating in shortest path formation. A coalition of nodes is considered as a winning coalition, if the path they form satisfies the bandwidth requirements of a particular source destination pair. We thus disregard links which cannot support these bandwidth requirements. Once a shortest path has been established, a node is called a swing node if it participates in the shortest path. The removal of a node that participates in maximum shortest path routes, will have a higher impact on network performance and is thus considered a critical node in the network. So, in analogy to the voting games setting, a node which acts as a swing node in maximum coalitions is the most critical node and is assigned the highest Banzhaf power index formally defined below.

In the graph $G = (V, E)$, I denotes the set of all source destination pairs $w = (i, j)$, $i, j \in V$. For each $w \in I$, $L(w)$ contains the set of nodes which constitute the shortest path route that fulfills the bandwidth requirements. A node k that belongs in $L(w)$ acts as a swing node for the source destination pair w. The Banzhaf power index for a node is the ratio between, the number of times a node acts as a swing node, over the total number of times all the nodes in V act as swing nodes. The Banzhaf power index is denoted by C_k and is given by:

$$C_k = \frac{\sum_{w \in I}(|L(w)| - |L(w)\backslash k|)}{\sum_{p \in V}\sum_{w \in I}(|L(w)| - |L(w)\backslash p|)} \tag{4}$$

C. *Combined Banzhaf & Diversity Index* (CBDI)

The proposed criticality metric is obtained by multiplying the diversity index and the Banzhaf Power Index as shown below:

$$CBDI(x) = C_x H(x) \tag{5}$$

The metric is referred to as Combined Banzhaf & Diversity Index (*CBDI*) and refines the mechanism of critical node detection. According to this index, a node is critical not only if it participates in maximum shortest path routes but if it is also prominent among its neighbors due to a higher variation in node attributes. The index, unlike previous approaches, is able to refine nodes which participate in the same number of shortest paths by differentiating between nodes which relay information from multiple inputs to multiple outputs and nodes which relay information from a single input to a single output. Further, it can identify nodes which can relay data to distant nodes thus having a high probability of experiencing heavy traffic. Finally, it is able to refine the information obtained by the node degree by excluding neighboring nodes whose participation in path formation is not critical.

III. Performance Evaluation

In this section, we evaluate the performance of the proposed criticality index using simulations conducted on Matlab [12]. We conduct a comparative study to investigate the performance of the proposed index against two other approaches that have appeared in the literature: the Hybrid Interactive Linear Programming Rounding (HILPR) algorithm proposed in [1] and the algorithm in [9] (Cont) which attempts to reduce the rank of the routing matrix. Among all criticality indices proposed in literature we have chosen the above as they contain some of the features included in our approach, namely the diversity, the node degree and the participation in shortest paths. In addition, they have been shown to outperform the other proposals in a number of scenarios. In each conducted simulation experiment, nodes participating in the network are assigned a criticality measure based on the criticality index under consideration. A fixed percentage of the most critical nodes are removed and the degradation in network performance is evaluated. The most effective criticality index is the one that leads to a greater degradation in performance. The network performance is evaluated in terms of the following performance metrics: the Average Node Degree, the Average Path Length and the Number of Isolated Nodes.

- *Average Node Degree:* It is the average number of neighbors of all nodes participating in the network. Small average node degree values imply smaller connectivity and so the smaller the average node degree, the greater is the degradation in network performance.
- *Average Path length:* This is obtained by calculating the average of all path lengths over all source destination paths in the network. High average path length in a network implies lack of critical nodes which can participate in shortest path routes. So, the higher the average path length, the greater is the degradation in network performance.
- *Number of Isolated Nodes:* The number of nodes, that have no connections with any other node in the network. High number of isolated nodes is undesirable as it implies greater network partitioning.

Three different network topologies were considered in an area of $1000 \times 1000m^2$.

1) *Random Network Topology:* In this network topology, the x and y coordinates of the nodes were uniformly distributed in the area under consideration. The number of nodes were chosen in the range of $10 - 80$ and among them 90% of the nodes were assumed to have a constant transmission range equal to $300m$ whereas, some randomly selected 10% of nodes were assigned a transmission range of $450m$ in order to enable long distance links [13].
2) *WaxMan Network Topology:* WaxMan Network topology was introduced by WaxMan in 1988 [14]. In this model, the probability that a connection is established between any two randomly distributed nodes u, v in the network $P(u, v)$ depends on the distance d between the nodes as shown below:

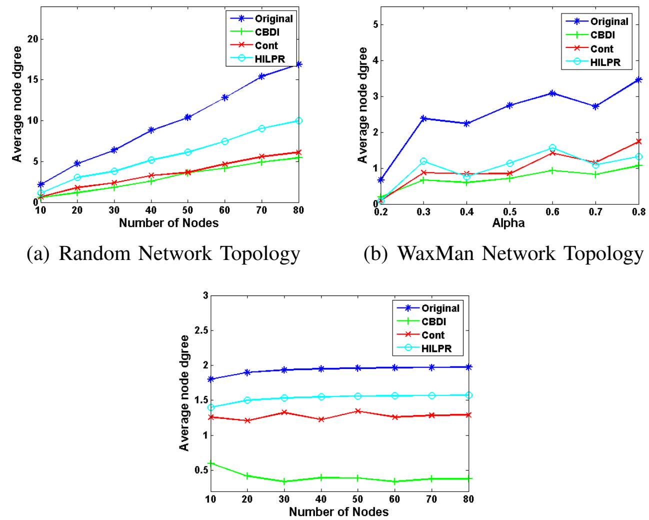
$$P(u,v) = \alpha e^{-d/bL} \tag{6}$$

where $0 < \alpha < 1$ and $b <= 1$ are constants and L is the maximum distance between any two nodes. As α increases, the probability of having edges between two nodes increases, whereas, with the increase in b, the ratio of long distance to short distance edges increases. In our simulations, we fix, the total number of nodes to 80 and consider a constant value of $b = 0.5$. In order to analyze the effect of node density on the performance of the network, we vary the value of α from $20 - 80\%$.

3) *Small World Network Topology:* Small World model was proposed by Watts and Strogatz in [15]. In a Small World network, N nodes form a one-dimensional lattice with each node placed uniformly on the boundary of a circle. Each node in the network forms a direct connection with its k^{th} nearest neighbors, where k is a constant and it represents the edge connectivity of the network. In this network topology, a network size varying from $20 - 80$ was considered, with a fixed edge connectivity of $k = 2$. In addition, 10% of the edges are randomly re-wired to introduce the long range links in the network. These long range links reduce the average path length between the nodes.

In each of these topologies, the criticality metric was evaluated by removing the selected critical nodes from the network and then measuring the network performance. In order to reduce the variance of the obtained results, each simulation experiment was repeated 50 times and the values presented, are averages over all obtained outputs. We assume a fluid flow model of the network and the bandwidth of each node is randomly selected according to a uniform distribution with a maximum value of 2Gbits/sec. Information sources are assumed to be non-responsive and their data rate is chosen from a uniform distribution in the range 0-2Gbits/sec. In each experiment, the performance of the reference network (we refer to it as the original network) is evaluated and then compared with the performance of the network when 20% of the total nodes are removed. The nodes which are removed are the ones which have been assigned the highest criticality value according to the criticality index under investigation.

In Fig. 1 for each network topology we show the average node degree values obtained in the original network and compare it with the values obtained when the most critical nodes are removed using the three criticality metrics under investigation. For the Random Network Topology, and the Small World Topology, the average node degree is plotted against the number of nodes within the network. In the WaxMan Topology, the average node degree is plotted against the parameter α of the model which is a measure of the edge density within the network. The greater the value of α, the greater is the edge density and thus the number of edges. We observe that in all cases, the proposed CBDI criticality metric achieves a larger reduction in the average node degree, a strong indication of a greater degradation in network performance. This implies that the nodes removed using the proposed CBDI metric are more critical. The highest impact of our approach compared to the



(a) Random Network Topology (b) WaxMan Network Topology

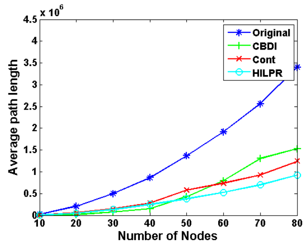
(c) Small World Network Topology

Fig. 1. Average Node Degree versus the number of nodes and α for the *Original* network and when nodes are removed using the *CBDI*, *Cont* and *HILPR* algorithms, in three different network topologies.

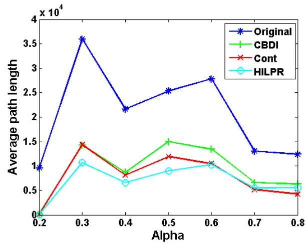
others is observed in the Small World Topology whereas the smallest impact is reported in the Random Network Topology. It is worth noting that in the Random Topology as the number of nodes increases, so does the average node degree. This is expected due to the increase in node density. A similar pattern is observed in the WaxMan Topology, however, the increase rate is smaller. For the Small World topology, the average node degree is fairly constant with increasing number of nodes due to the nature of the model which assumes a constant value for the average node degree equal to 2.

In Fig. 2, for each considered network topology, we show the Average Path Length reported in the original network and the network resulting from the removal of the critical nodes. The critical nodes are chosen using the proposed criticality metric and the other two metrics under consideration. Higher Average Path Length values are desirable, when removing critical nodes, as they imply the removal of nodes which participate in shortest paths. We observe that the proposed metric, is able to slightly increase the average path length in the WaxMan and Random Network topologies, at high α and number of node values respectively. This is expected due to a higher variability in node attributes when increasing the node density. In the Small Network Topology almost zero path length values are reported by the CBDI metric due to the large number of isolated nodes that it creates. This is highlighted below.

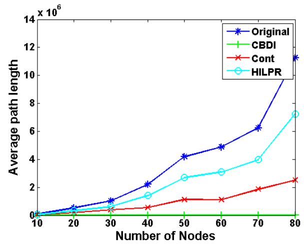
Finally in Fig. 3 we show the number of isolated nodes reported in each of the network topologies under consideration. The number of isolated nodes is shown for increasing values of the number of nodes and α in the original network and when the critical nodes have been removed using the considered criticality metrics. The results demonstrate the superiority of the proposed metric, especially in the case of the Random



(a) Random Network Topology

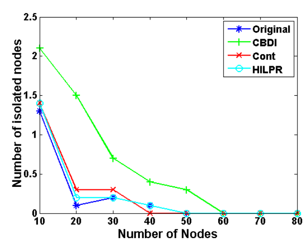


(b) WaxMan Network Topology

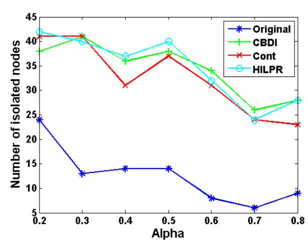


(c) Small World Network Topology

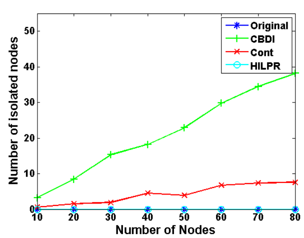
Fig. 2. Average Path Length versus the number of nodes and α, for the *Original* network and when nodes are removed using the *CBDI*, *Cont* and *HILPR* algorithms, in three different network topologies.



(a) Random Network Topology



(b) WaxMan Network Topology



(c) Small World Network Topology

Fig. 3. Number of Isolated nodes versus the number of nodes and α, for the *Original* network and when nodes are removed using the *CBDI*, *Cont* and *HILPR* algorithms, in three different network topologies.

Network topology and the Small World topology. In all three topologies, the removal of critical nodes using the proposed CBDI criticality metric yields a larger number of isolated nodes implying a severe degradation in network performance. Increasing number of isolated nodes suggests that the network becomes increasingly intermittent in nature. It is worth noting that, in the Random Network Topology and the WaxMan network topology, as the number of nodes and α increase, the isolated nodes decrease. This is expected due to the fact that an increase in the node or edge density makes isolation of nodes more improbable. On the other hand, in the case of the Small World Topology as the number of nodes increases, so does the number of isolated nodes. This is due to the fact that in this topology the average node degree is fixed, which means that as the number of nodes increases, the number of nodes removed also increases which renders more nodes to become isolated. The fact that the node degree is originally fixed yields zero isolated nodes in the original network, as shown in Fig. 3.

IV. Conclusions

In this work we highlight the contribution of critical nodes in network operation and demonstrate how the network reacts when these critical nodes are affected. We propose a new crticiality index which is based on the diversity of node attributes within the network and the participation of each node in forming shortest path routes. The proposed metric outperforms existing approaches by showing a greater degradation in network performance when the critical nodes, selected using this index, are removed from the network. In the future, we aim at further evaluating the performance of the proposed criticality index by considering additional performance metrics and by using event based simulation tools and practical network examples. We will also pursue analytical evaluation of the proposed scheme. Finally, we aim at proposing countermeasures in order to reduce the impact of removing critical nodes in communication networks.

References

[1] Y. Shen, N. P. Nguyen, Y. Xuan, and M. T. Thai, “On the discovery of critical links and nodes for assessing network vulnerability,” *Networking, IEEE Transactions on*, vol. 21, no. 3, pp. 963–973, 2012.

[2] D. Liu, P. Ning, and W. Du, “Detecting malicious beacon nodes for secure location discovery in wireless sensor networks,” in *Distributed Computing Systems, 2005. ICDCS 2005. Proceedings. 25th IEEE International Conference on*, pp. 609–619, IEEE, 2005.

[3] R. Narayanam, “A game theoretic approach to identify critical components in networked systems,” in *SRII Global Conference (SRII), 2012 Annual*, pp. 515–524, IEEE, 2012.

[4] L. C. Freeman, “A set of measures of centrality based on betweenness,” *Sociometry*, pp. 35–41, 1977.

[5] R. Albert, H. Jeong, and A.-L. Barabási, “Error and attack tolerance of complex networks,” *nature*, vol. 406, no. 6794, pp. 378–382, 2000.

[6] R. Albert, I. Albert, and G. L. Nakarado, “Structural vulnerability of the north american power grid,” *Physical review E*, vol. 69, no. 2, p. 025103, 2004.

[7] L. d. F. Costa, F. A. Rodrigues, G. Travieso, and P. Villas Boas, “Characterization of complex networks: A survey of measurements,” *Advances in Physics*, vol. 56, no. 1, pp. 167–242, 2007.

[8] L. C. Freeman, “Centrality in social networks conceptual clarification,” *Social networks*, vol. 1, no. 3, pp. 215–239, 1979.

[9] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, “Controllability of complex networks,” *Nature*, vol. 473, no. 7346, pp. 167–173, 2011.

[10] Z. Zhang, L. Rong, and C. Guo, “A deterministic small-world network created by edge iterations,” *Physica A: Statistical Mechanics and its Applications*, vol. 363, no. 2, pp. 567–572, 2006.

[11] P. Dubey and L. S. Shapley, “Mathematical properties of the banzhaf power index,” *Mathematics of Operations Research*, vol. 4, no. 2, pp. 99–131, 1979.

[12] MathWorks, “Matlab 2013,” *http://www.mathworks.com/support/sysreq/sv-r2013a/*.

[13] W. Asif, H. K. Qureshi, and M. Rajarajan, “Variable rate adaptive modulation (vram) for introducing small-world model into wsns,” in *Information Sciences and Systems (CISS), 2013 47th Annual Conference on*, pp. 1–6, IEEE, 2013.

[14] B. M. Waxman, “Routing of multipoint connections,” *Selected Areas in Communications, IEEE Journal on*, vol. 6, no. 9, pp. 1617–1622, 1988.

[15] D. J. Watts and S. H. Strogatz, “Collective dynamics of small-worldnetworks,” *nature*, vol. 393, no. 6684, pp. 440–442, 1998.