Weighted Betweenness for Multipath Networks

Dianne S. V. Medeiros, Miguel Elias M. Campista
Universidade Federal do Rio de Janeiro
Rio de Janeiro, RJ – Brazil
{dianne,miguel}@gta.ufrj.br

Nathalie Mitton
INRIA Lille-Nord Europe
Villeneuve d’Ascq, France
nathalie.miton@inria.fr

Marcelo Dias de Amorim, Guy Pujolle
Université Pierre et Marie Curie
Paris, France
{marcelo.amorim, guy.pujolle}@lip6.fr

Abstract—Typical betweenness centrality metrics neglect the potential contribution of nodes that are near but not exactly on shortest paths. The idea of this paper is to give more value to these nodes. We propose a weighted betweenness centrality, a novel metric that assigns weights to nodes based on the stretch of the paths they intermediate against the shortest paths. We compare the proposed metric with the traditional and the distance-scaled betweenness metrics using four different network datasets. Results show that the weighted betweenness centrality pinpoints and promotes nodes that are underestimated by typical metrics, which can help to avoid network disconnections and better exploit multipath protocols.

Keywords—Centrality metrics, graph, static and dynamic networks.

I. INTRODUCTION

Metrics from graph theory provide means to quantify the importance of a node in a system and, thus, help to identify nodes playing central roles. The importance of a node is often computed through centrality metrics [1]–[4], which classify nodes according to their topological position in the network. It is very common to employ the betweenness centrality, a metric that considers the proportion of shortest paths a node falls on. The idea is that the more shortest paths a node participates in, the more the node is central [1].

Centrality is important to many networked applications, such as election mechanisms and routing protocols. These applications often rely on shortest paths between source-destination pairs. Assigning importance to nodes based only on their participation in shortest paths, however, may lead to biased classifications. This could happen, for instance, when a node that falls on a number of shortest paths is classified as more important than another that falls on slightly fewer shortest paths, but on a multitude of quasi-shortest paths.

Several authors have already questioned the use of only shortest paths to quantify nodes’ importance [5]–[7]. We also believe that such a definition limits the metric applicability and propose a novel weighted betweenness centrality. This proposal extends the definition of the traditional betweenness to also include nodes that fall on quasi-shortest paths. In a nutshell, the computation of the weighted betweenness of a node \( \nu_k \) is based on the ratio between the length of the shortest path connecting a given pair of nodes and the length of the quasi-shortest path passing through \( \nu_k \) between the same pair of nodes. The idea is to give more importance to shortest paths and paths slightly longer than the shortest one. In addition, we scale the contribution of each path according to the ratio between the number of shortest paths and quasi-shortest paths between the pair of nodes. In order to bound the computation complexity, each path is only accounted if the difference between its length and the length of the shortest path is less or equal than a given parameter \( \gamma \).

The impact of our metric is investigated using four network datasets. To this end, we compare the weighted betweenness with both the traditional [8] and the distance-scaled betweenness [7] metrics. We examine the node ranking for each metric and we investigate its behavior over time to analyze the impact on network stability. Our main findings are that the proposed metric (i) identifies nodes misclassified by the traditional and distance-scaled betweenness centralities, (ii) solves ties between nodes classified in the same position but that do not present the same importance to the network and (iii) potentially keeps more nodes with the ability of intermediating flows.

This paper is organized as follows. Section II presents all the notations and definitions used herein. Section III discusses the related work and overviews betweenness centrality. In Section IV, the problem is stated, while in Section V the proposed betweenness is introduced. The selected datasets are described in Section VI and the obtained results are shown in Section VII. Finally, Section VIII concludes this work and presents our future plans.

II. NOTATION AND DEFINITIONS

We model a network as an weighted graph \( \mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega) \), where \( \mathcal{V} \) and \( \mathcal{E} \) are the sets of vertices and edges, respectively, and \( \omega : \mathcal{V} \rightarrow \mathbb{R}_+ \) is a weight function defined on its vertices. Each vertex \( \nu_i \in \mathcal{V} \) and each edge \( \epsilon_{i,j} \in \mathcal{E} \) represent, respectively, a node in the network and a link between a pair of nodes \( [\nu_i, \nu_j] \). Each edge has a weight \( \omega_{i,j} \in \mathbb{R}_+ \), that is the cost of the link. The edges \( \epsilon_{i,j} \) and \( \epsilon_{j,i} \) exist simultaneously if the network is symmetric. Following, we present important definitions to understand this work.

**Definition 1 Path:** A path \( p_{(\nu_i, \nu_j)} \) from a source \( \nu_i \) to a destination \( \nu_j \) is an ordered sequence of distinct nodes in which any consecutive pair is connected by a link.
A path \( p_{1,L} \) between the source, \( v_1 \), and the destination, \( v_L \), has a total length of \( \Delta L = L - 1 \) hops, with \( L \in \mathbb{N}^* \), and is given by \( p_{1,L} = (v_1, v_2, ..., v_{L-1}, v_L) \), where \( \{v_1, v_2, ..., v_{L-1}, v_L\} \subseteq V \) and \( \{\varepsilon_{1,2}, ..., \varepsilon_{L-1,L}\} \subseteq E \). The shortest path, or geodesic, between these nodes is the one where \( \Delta L \) is the smallest possible value, \( \Delta L^* \).

**Definition 2 Quasi-shortest path:** The path is quasi-
shortest if \( \Delta L - \Delta L^* \leq \gamma \).

Parameter \( \gamma \) serves to limit the stretch of quasi-shortest
paths and avoid the explosion of the number of possibilities.
Furthermore, given \( p_{1,L}^* \), it is not reasonable to consider as
useful those paths for which \( \Delta L \gg \Delta L^* \).

More than one path may exist between a pair of nodes, either
shortest or quasi-shortest. Therefore, the number of
goedescs from \( v_1 \) to \( v_L \) is denoted by \( n_{1,L}^* \), and the number
of quasi-shortest paths is \( n_{1,L} \). Since we rely on the number
of hops for path computation, a quasi-shortest path has
always a greater number of intermediate nodes.

**Definition 3 Path cost:** The cost of a path \( p_{1,L}^* \) between
a pair of nodes is obtained from the number of hops
connecting them.

We use the number of hops as the cost metric, such that the
shortest path is the geodesic, represented by \( p_{1,L}^* \). The
cost of \( p_{1,L}^* \) is \( \delta_{1,L}^* \), while the cost of the quasi-
shortest path between the same pair of nodes is \( \delta_{1,L} \geq \delta_{1,L}^* \). Note that if we
consider an intermediary node \( v_k \), the path cost can be
always computed as the sum \( \delta_{1,k}^* + \delta_{k,L}^* \), independently of whether \( v_k \) falls on the geodesic.

### III. Revisiting the Betweenness Centrality

The betweenness centrality expresses the influence that a
specific node could have on other nodes in the network. It is based on the computation of the number of geodesics a	node is part of, considering all possible pairs of nodes in the
network. Nevertheless, accounting only shortest paths can
limit the connectivity of the greater component. Freeman et al. [5] and Opsahl et al. [3] tackles
this issue by considering the link weight in the computation.

Newman [6] uses another approach to handle the problem
of accounting only shortest paths. The author advocates that
information does not travel only through shortest paths, either because the shortest path is not known in advance or
it does not exist, as a specific destination is not determined.
Newman proposes then the random walk betweenness, in
which both shortest paths and non-shortest paths are
accounted for. Borgatti et al. [7], in turn, argue that information
tends to concentrate on shorter paths and propose the
distance scaled betweenness, in which the contribution of each
path is inversely proportional to its length.

![Figure 1. Example of network in which the traditional betweenness centrality elects \( v_b \) as more important than \( v_p \), even considering the connectivity of the greater component.](image)

This work also shares the same rationale of previous works –
information does not necessarily travel through shortest paths. Nevertheless, our approach differs by proposing a closed expression for a weighted betweenness, which does not only consider shortest paths, but also paths longer than the shortest one up to a \( \gamma \) value. This would avoid the metric to take into account all possible existing paths in the network. The following subsections review the formal
definition of traditional and distance-scaled betweenness,
selected for comparison with our proposal.

#### A. Traditional betweenness centrality

Traditionally, the betweenness is defined for unweighted
graphs, and it can be used in symmetric or asymmetric
graphs. Its formal definition is given by:

\[
b(v_k) = \sum_{i,j} n_{i,j}^*(v_k) / n_{i,j}^*, \tag{1}
\]

where \( n_{i,j}^*(v_k) \) is the number of geodesics between \( v_i \)
and \( v_j \) which have \( v_k \) as intermediary node. Note that to be considered as geodesic, a path \( p_{1,L} \) must cost \( \delta_{1,L}^* \).
Therefore, any path with cost higher than \( \delta_{1,L}^* \) is ignored
and \( n_{i,j}^*(v_k) = 0 \), nullifying the term in the summation.

#### B. Distance-scaled betweenness centrality

The idea here is to make the contribution of each path
inversely proportional to its length, assigning different impor-
tance to each path considered in Equation 1. Its formal
definition is given by Equation 2.

\[
b_{\lambda_{i,j}}(v_k) = \sum_{i=1}^{\vert V \vert} \sum_{j=1}^{\vert V \vert} \frac{1}{\Delta L} n_{i,j}^*(v_k) / n_{i,j}^*, \tag{2}
\]

### IV. Motivating Example and Problem Statement

The traditional betweenness frequently ignores nodes ap-
aparently important to the connectivity of the network. This
happens because the metric traditionally uses only shortest
paths to define the importance of nodes. This problem is
illustrated in Figure 1, in which \( v_c \) is one of the main nodes responsible for maintaining the components \( N1 \) and \( N2 \) connected. If it fails, \( v_p \) should replace the role of \( v_c \), since it would be then the responsible for the connectivity of a greater component. In the traditional betweenness, however, the centrality of \( v_b \) is greater than the one of \( v_p \), because it connects \( v_f \) to all the other nodes in the network, despite \( v_b \) not being able to keep both components connected. Considering network connectivity, the presence of alternative paths not taken into account can represent, depending on the application, under-utilization of available resources.

In the Freeman’s graph (Section VI), several paths exist in the network, but they can be underutilized when only geodesics are considered. We can have an idea of the number of paths neglected due to the traditional definition of betweenness if we compute the difference between the geodesic length, with cost \( \delta_{i,j}^* \), and the quasi-shortest path intermediated by \( v_k \), \( p_i(v_k) \), with cost \( \delta_{i,j} = \delta_{i,k}^* + \delta_{k,j}^* \). If \( (\delta_{i,k}^* + \delta_{k,j}^*) - \delta_{i,j}^* = 0 \), then \( v_k \) belongs to a geodesic. This difference is referred to as \( \Delta = (\delta_{i,k}^* + \delta_{k,j}^*) - \delta_{i,j}^* \), which is upper bounded by \( \gamma \).

Figure 2 shows the distribution for the parameter \( \Delta \) for Freeman’s graph. The \( X \)-, \( Y \)-axes represent, respectively, the intermediary node and all possible values of \( \Delta \) for this node. The color scale indicates the relative frequency of occurrence for a given \( \Delta \), such that higher frequencies are darker. We observe in Figure 2 that \( \Delta > 0 \) frequently happens and, furthermore, less than 20\% (0 \( \leq \) ID \( \leq \) 5) of nodes often fall on shortest paths. We can conclude that a significant number of nodes may never intermediate a communication due to the traditional definition of betweenness, assuming that only shortest paths are used.

Figure IV shows how frequently we can find a given distance \( \Delta \) in Freeman’s graph. Figure IV, in turn, shows the probability of finding a value \( \Delta \) smaller or equal to \( \Delta = x \). In general, we note that \( \Delta = 1 \) and \( \Delta = 2 \) happen more often than \( \Delta = 0 \), while the maximum value \( \Delta = 4 \) is negligible. This indicates that nodes in quasi-shortest paths should be considered when computing betweenness, but only those falling in paths slightly longer than the geodesic.

V. WEIGHTED BETWEENNESS CENTRALITY

Our goal is to capture the potential of intermediate nodes, even when they do not belong to a geodesic, possibly revealing nodes that are ignored by the traditional vision of betweenness. To this end, we propose the weighted betweenness centrality, which considers both geodesics and longer paths upper bounded by a parameter \( \gamma \). The contribution for the metric is proportional to the ratio between the number of shortest and quasi-shortest paths connecting \([v_i, v_j]\). In addition, the metric considers the length of the paths, unlike the traditional betweenness.

While the traditional betweenness considers the ratio \( n_{i,j}(v_k)/n_{i,j}^* \), based on the number of geodesics, the weighted betweenness considers the ratio between the lengths of the geodesic and the quasi-shortest path between \([v_i, v_j]\) when passing through \( v_k \). Hence, the ratio between the aforementioned costs is defined as \( \delta_{i,j}^*/(\delta_{i,k}^*+\delta_{k,j}^*) \) and if it is greater than \( \gamma \), the quasi-shortest path is ignored. Note that using the number of hops does not limit the metric, which can be generalized to use other types of link weight.

Figure 4 shows the difference between all three approaches using the number of hops to compute the path cost. In Figure 4(b), node \( v_b \) belongs to one of two geodesics between \([v_i, v_j]\). Thus, for the traditional and distance-scaled
betweenness, we have \( n_{i,j}^*(v_k)/n_{i,j}^* = 1/2 \), and for the latter metric we still multiply this ratio by \( 1/\Delta L^* = 1/4 \). In Figure 4(c), \( v_t \) falls in a quasi-shortest path between the same nodes. While both traditional and distance-scaled approaches ignore the potential of this node for the pair \([v_i, v_j]\), our metric considers the ratio \( \delta_{i,j}^*/(\delta_{i,k}^* + \delta_{k,j}^*) = 4/5 \). To account for multiple paths with equal lengths intermediated by \( v_t \), the ratio between the costs is scaled by the number of such paths. The idea is to assign more importance to quasi-shortest paths if they are proportionally less numerous than shortest paths. In opposition, the importance is reduced if they are proportionally more numerous than the number of shortest paths. Table I compares the betweenness centralities computed for nodes \( v_c, v_b, \) and \( v_p \) in Figure 1, according to the traditional, distance scaled, and the proposed weighted betweenness, considering only 5 nodes totally connected in both networks \( \Pi \) and \( \Pi \). The weighted betweenness can capture better the centrality of the nodes, even if they do not belong to several geodesics. This result could also be used, e.g., to determine which node could be used for load balancing or to aggregate data from sensor networks.

We formally define the proposed metric for node \( v_k \), \( w(v_k) \), as:

\[
w(v_k) = \sum_{i=1}^{\mid V \mid} \sum_{j=1 \neq k}^{\mid V \mid} \frac{n_{i,j}^*(v_k)}{n_{i,j}(v_k)} \times \frac{\delta_{i,j}^*}{\delta_{i,k}^* + \delta_{k,j}^*}, \tag{3}
\]

where \( n_{i,j}^* \) is the number of geodesics between \([v_i, v_j]\), and \( n_{i,j}(v_k) \) is the number of paths with the same length of the quasi-shortest path \( p_{i,j}(v_k) \). The weighted betweenness as defined by Equation 3 is computed for pairs of nodes in the same component. In case of distinct components, the contribution is null.

Both superior and inferior limits for the proposed metric depend on the relation between the number of paths, and between the cost of the paths. The inferior limit further depends on the parameter \( \gamma \). In the worst case for the inferior limit, if \( n_{i,j}(v_k) \gg n_{i,j}^* \) or if \( \delta_{i,k}^* + \delta_{k,j}^* \gg \delta_{i,j}^* \), the corresponding term will tend to zero. The superior limit, in turn, can be equal to 1 when the quasi-shortest path is, in fact, a geodesic, or equal to \( n_{i,j}^* \times \delta_{i,j}^*/(\delta_{i,k}^* + \delta_{k,j}^*) \), if only one quasi-shortest path exists. As the weighted betweenness considers paths slightly longer than the geodesic, accounting all quasi-shortest paths between a pair of nodes intermediated or not by \( v_k \) can become unfeasible. Therefore, the parameter \( \gamma \) limits the Depth-First Search (DFS) algorithm used in this work, reducing its complexity.

Finally, the weighted betweenness requires the previous knowledge of the path costs, which can be infinite for disconnected components. As a consequence, the metric is computed only for nodes from the same component to avoid problems related to infinite costs. In this case, the contribution to the betweenness is considered null.

### VI. Datasets

In order to maintain the generality of the metric, we use four datasets with distinct characteristics to evaluate the weighted betweenness centrality proposed in this work.

- **Freeman’s EIES**: presents the relationships in a group of 32 academics [9]. A directed edge between two nodes \([v_i, v_j]\) exists only if \( v_i \) has sent a message to \( v_j \), totaling 460 links.
- **Dolphins**: provides the association relationship between 62 dolphins in Doubtful Sound, New Zealand [10]. Each node corresponds to a dolphin and the interaction between them is represented by an undirected edge \( \varepsilon_{i,j} \), totaling 159 links.
- **PhD Students**: it is a network of relationships between 1,025 PhD students and supervisors [11]. A directed link exists from \( v_i \) to \( v_j \) if \( v_i \) is the supervisor of \( v_j \), totaling 1,043 links.
- **TAPASCologne Dataset**: it models the vehicular traffic in the city of Cologne, Germany [12]. We use 10 samples of the original subset, containing 1,584 to 1,916 nodes and 1,573 to 2,044 undirected links. Each node is a vehicle and an edge exists between them if they are less than 50 meters away from each other.

### VII. Results

In this section, we present the impact of our metric on the selected datasets. We use \( \gamma = 3 \), since for the Freeman’s graph a \( \Delta \geq 4 \) is not frequent. Thus, we account quasi-shortest paths for which \( \Delta L \leq \Delta L^* + 3 \) or \( \delta_{i,L} \leq \Delta L^* + 3 \). Our first goal is to analyze the behavior of the ranking obtained for each metric. Following, we investigate the impact of the proposed metric on the stability of network. Note that in our results we consider that nodes with the same value of betweenness are tied in the same position.

#### A. Impact on the node ranking

We investigate the ranking variation for the distance scaled betweenness and our metric, through analysis of the node position gain for each metric in relation to the traditional betweenness. A positive gain implies centrality increase, whereas a negative gain indicates the opposite. Figure 5 shows that both metrics are able to modify the classification of nodes. In Figure 5(a), the X-axis is organized according to the traditional betweenness classification for the Freeman dataset, with “Lin Freeman” as the most central
Figure 5. Position gain in relation to the ranking provided by the traditional betweenness compared with the distance-scaled betweenness.

Figure 6. Cumulative distribution of normalized betweenness computed for the four datasets using all three metrics.

node. We observe that our metric modifies the classification of some nodes, specially those that are in lower positions, indicating that the traditional betweenness can underestimate the centrality of nodes that do not participate in many geodesics. For instance, “John Boyd” gains 10 positions according to the proposed metric. In addition, “Gary Coombs”, “Brian Foster”, and “Nick Poukhinsky” which are tied with null traditional and distance-scaled betweenness are, in turn, reclassified in new positions according to the weighted betweenness. We emphasize these nodes with vertical arrows at the plot. The reclassification is also observed for the Dolphins dataset, as shown in Figure 5(b). Several nodes are demoted and promoted using both metrics. Once again, a significant number of nodes classified in lower positions are reclassified by our metric, while the distance scaled betweenness barely modifies the ranking. Further, nodes that were once tied with null betweenness (from node “Cross” downwards) are redistributed in the ranking.

Figure 6 shows the cumulative distribution functions for all three metrics. The betweenness value was normalized by the maximum value found for the dataset, allowing all metrics to be plotted using the same range in the X-axis. In Figure 6(a), the curves for the traditional and distance-scaled betweenness are coincident, corroborating the results illustrated in Figure 5(a), where the position gain related to the distance-scaled betweenness is null. Further, we observe that approximately 40% of the nodes have low or no importance to the network, whereas our metric assigns low or no importance to only less than 5% of nodes. We observe the same behavior in Figures 6(b) and 6(c).

Figure 6(d) shows a singular behavior for the cumulative distribution of betweenness for the PhD Students dataset. All three curves are practically coincident. This happens because the PhD Students network is a directed graphs with many vertices that never, or almost never, would intermediate communications, since most edges connect supervisors to their respective students and only a few nodes represent both roles, originating many leaf nodes with null betweenness. In addition, as multipaths barely exist in this network, all three metrics equivalently capture the centrality for most nodes.

B. Impact on the number of network disconnections

From our results so far, we could note that the weighted betweenness is able to reclassify nodes that apparently should be given more importance. The main idea of our metric, however, is to reduce the number of disconnections in networks with dynamic topology. In this direction, we claim that the number of disconnections can potentially increase if the number of nodes with zero betweenness...
The proposed weighted metric is able to reduce the number of disconnections in the network compared with the other metrics. As future work, we plan to verify the influence of the parameter $\gamma$ and extend the metric to weighted networks.

**REFERENCES**


