Computing
https://doi.org/10.1007/s00607-022-01137-7

REGULAR PAPER

Disjoint multipath closeness centrality

Mariana S. M. Barbosa[1] · **Dianne S. V. Medeiros**[2] · **Miguel Elias M. Campista**[1]

Received: 4 March 2022 / Accepted: 18 November 2022

Abstract

Traditional centrality metrics consider only shortest paths, neglecting alternative paths that can be strategic to maintain network connectivity. This paper proposes the disjoint multipath closeness centrality, a new metric to compute node centrality that extrapolates the traditional closeness to consider multiple shortest and disjoint *quasi*-shortest paths. The idea is to identify nodes that are close to all other nodes and are multiply-connected, which is important to perform high availability tasks. We limit the number of multiple disjoint paths using a connectivity factor φ. We comparatively investigate the correlation between our metric, the traditional closeness, and the information centrality using social and communication networks. We also assess the node ranking obtained by each metric and evaluate node reachability when one or multiple network failures occur. The results show that our metric maintains high concordance with the other closeness metrics but it can reclassify at least 59% of nodes in the evaluated networks. Our metric indeed identifies better-connected nodes, which remain more accessible when failures happen.

Keywords Centrality metrics · Closeness · Disjoint multipaths

Mathematics Subject Classification 05C82 · 05C40 · 91D30

Dianne S. V. Medeiros and Miguel Elias M. Campista have contributed equally to this work.

✉ Mariana S. M. Barbosa
maciel@gta.ufrj.br

Dianne S. V. Medeiros
diannescherly@id.uff.br

Miguel Elias M. Campista
miguel@gta.ufrj.br

1 GTA/PEE-COPPE/DEL-Poli, UFRJ, Rio de Janeiro, Brazil

2 MídiaCom/PPGEET, UFF, Niterói, Brazil

Published online: 13 December 2022

1 Introduction

Centrality metrics are commonly used to rank network nodes according to their topological importance [4, 17]. In communication networks, identifying central nodes is useful for resource or service provisioning [10], or even to avoid network partitioning upon failure conditions [22]. Centrality metrics are also used to determine nodes that concentrate traffic or those with considerable spreading power [30]. This last characteristic relates to the speed at which nodes deliver information to every other node in the network, which is a property captured by the closeness centrality [17, 26, 30]. Closeness assumes that information flows exclusively through the shortest paths and use these paths to compute node importance. This assumption neglects alternative paths that are "a little bit longer" and can still be used to transfer flows. Several authors agree that shortest paths are relatively more important but advocate the use of longer paths to compute node centrality [9, 23, 26, 30]. Nevertheless, the shortest paths concentrate more flows, as they offer less resistance to transfer flows [11].

The information centrality [30] follows the multipath approach, but does not limit the maximum number of paths interconnecting pairs of nodes. The lack of limitation may not suit well practical implementations in which flows cannot follow all paths. Furthermore, this metric does not consider the network dynamics, which may eliminate paths between nodes dynamically. For instance, in failure-prone networks, node's importance must consider disjoint paths, such that better-connected nodes must have higher centrality. The issues of metric applicability and network dynamics are not simultaneously considered in multipath centrality metrics, representing a gap in the literature.

This paper proposes the *disjoint multipath closeness centrality* (C_φ), which extrapolates the traditional closeness concept to account for multiple disjoint paths[1]. We aim to identify nodes more prone to remain reachable when links are eliminated. We consider that such elimination is due to a network failure. The flow process tends to follow shortest paths, but is not limited to them. It is desirable to have disjoint paths connecting nodes to maintain nodes' reachability upon failure. Therefore, we assign more importance to nodes multiply connected to every other node, considering the multiple paths' average cost. The metric only accounts for disjoint paths, that can involve a set of shortest and *quasi*-shortest paths [23]. We tune the maximum number of paths between each pair of nodes using a threshold, the *connectivity factor* (φ). Our proposed metric brings a new notion of centrality for networks relying on topologies with multiple disjoint paths. For example, in real-world scenarios prone to failures, nodes playing a central role must have assigned more importance according to their ability to remain available after a failure. A node hosting a middlebox such as a firewall or a proxy cannot become unreachable as it is in charge of analyzing all outgoing traffic of a given network. A node with more disjoint paths in social networks can also ensure information delivery even if other nodes leave the network. Hence, in our paper, we argue that central nodes must be multiply-connected to avoid periods of networking service interruptions. We dive into more details later in the paper, but we

[1] A preliminary version of this work was published in Portuguese at a Brazilian conference, SBRC 2019 [3]. Here, we improve the metric presentation and evaluation, using more networks, and adding the analysis of node reachability upon network failures.

believe that the importance of being quickly achievable and multiply-connected for nodes playing central roles is evident.

We evaluate the proposed metric using five datasets, including social and communication networks. We assess our metric comparatively to two commonly used closeness centrality metrics: traditional closeness [5] and information centrality [30]. We identify the influence of φ on the number of paths used and on the node ranking. We analyze the concordance between the metrics and the reclassification rates. We also investigate node reachability upon network failures. The results show a trend of using all existing disjoint paths between nodes when increasing the φ value. The proposed metric reveals better-connected nodes, reclassifying at least 59% of nodes in the evaluated networks. Finally, our metric also identifies nodes that remain connected to a greater number of nodes when the network undergoes failure conditions.

This paper is organized as follows. We present notation and definitions in Sect. 2. Section 3 reviews related centrality metrics. Section 4 discusses related works. We formally define our metric in Sect. 5. Section 6 describes the datasets used to assess our metric and presents a comparative analysis. Finally, Sect. 7 concludes this work and anticipates research directions.

2 Notations and definitions

We advocate that the more important nodes are the ones multiply-connected to the other nodes in the network via disjoint shortest paths. To quantify their importance, we model the network as a weighted directed graph $\mathsf{G}(\mathsf{V}, \mathsf{E}, \mathsf{W})$, where V, E and W denote the sets of nodes, links, and weights. A link $\varepsilon_{i,j} \in \mathsf{E}$ with cost $\omega_{i,j} \in \mathsf{W}$ exists between nodes $\upsilon_i, \upsilon_j \in \mathsf{V}$. The rightmost part of Fig. 1 shows a graph with 7 nodes and 8 links with unitary weight. A path $p_{s,t}$ between υ_s and υ_t is composed by a sequence of distinct adjacent nodes interconnected by links $\varepsilon_{i,j}$. The cost $\delta_{s,t}$ of path $p_{s,t}$ is obtained from the sum of weights $\omega_{i,j}$ of all links that compose the path. If weights are unitary, the total cost of path $p_{s,t}$ is equal to the number of hops between υ_s and υ_t. A pair of nodes υ_s, υ_t can be interconnected by a set of n paths, each one represented by $p_{s,t}^{(n)}$, with cost $\delta_{s,t}^{(n)}$. Within this set, there are shortest paths, with cost $\delta^*_{s,t}$, and *quasi*-shortest paths, with cost $\delta_{s,t} > \delta^*_{s,t}$ and maximum cost limited to $\delta^*_{s,t} + \rho$, where $\rho \in \mathbb{N}$ is a spreadness factor. This factor is configurable and can be viewed as the additional cost an entity accepts to pay to arrive at a given destination [23]. The cost of the shortest path $p^*_{s,t}$ is $\delta^*_{s,t} = 2$ and $p_{s,t}^{(1)}$ and $p_{s,t}^{(2)}$ are *quasi*-shortest paths with $\delta_{s,t}^{(1)} = \delta_{s,t}^{(2)} = 3$, i.e., $\rho = 1$. In this paper, we consider ρ unlimited without loss of generality.

Finally, a set of paths of G is disjoint if a vertex is internal to only one path of this set [8]. Hence, two paths $p_{s,t}^{(x)}$ and $p_{s,t}^{(y)}$ are vertex-disjoint, if and only if, $p_{s,t}^{(x)} \cap p_{s,t}^{(y)} = \{\upsilon_s, \upsilon_t\}$. As a corollary, partially disjoint paths share at least one intermediate node. These definitions remain valid even if the number of intermediate nodes or the cost of the paths is different. Hereinafter, we use disjoint paths interchangeably to vertex-disjoint paths. In Fig. 1, all paths in the rightmost graph are disjoint.

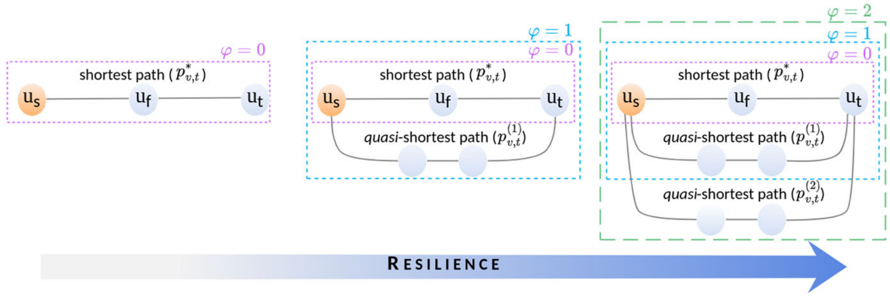


Fig. 1 The communication resilience between nodes υ_s and υ_t increases with the number of paths connecting them, even if *quasi*-shortest paths are used. The pair of nodes is even better connected if these multiple paths are disjoint. Hence, increasing φ contributes to the connectivity improvement between υ_s and υ_t

2.1 Connectivity factor

Multiple paths connecting pairs of nodes are desirable for resilience purposes, even if paths have different costs. Upon a node failure or an overloading condition, the information could follow one or more of the remaining paths. The utilization of multiple paths, however, comes with a tradeoff on the overall cost of the information transfer. Hence, it is beneficial to consider only paths that are "a little bit longer" than the shortest one, so as not to lose performance. As such, *quasi*-shortest paths [23] are considered to accommodate the idea that other paths, besides shortest ones, must also be used to identify central nodes. Besides using shortest and *quasi*-shortest paths, we also advocate in favor of path's disjointness, emphasizing the importance of network connectivity. Disjoint paths help maintain the communication between nodes, as one path's unavailability is not enough to cease the information flow. In practical settings, e.g., core networks, multipath provisioning between node pairs is common for backup or load balancing. In these networks, finding central nodes considering a fixed number of disjoint paths is particularly useful.

The set of multiple paths used to calculate our metric is limited in number by the Connectivity Factor φ. This parameter limits the maximum additional number of paths that *may* be used to reach the target node. The actual number of disjoint paths used depends on the network topology and can be less than φ if all existing disjoint paths are found before reaching φ. Therefore, φ defines an upper bound for the number of disjoint paths between a pair of nodes, with $\varphi \in \mathbb{N}$. Section 5.2 details the search for disjoint paths using φ.

Figure 1 illustrates the idea behind the connectivity factor φ. As illustrated in the leftmost graph, considering a single-path communication between nodes υ_s, υ_t these nodes become disconnected upon failure of node υ_f. If a single-path approach is preferred, only the shortest path is used, which is equivalent to $\varphi = 0$ in our proposal. If we add a disjoint path, υ_s and υ_t can communicate even if the alternative path is a little bit longer than the shortest one. This scenario is possible if $\varphi = 1$, where an increase in the path cost happens if υ_f fails. In the rightmost graph, $\varphi = 2$, and υ_s would still be able to reach υ_t through a 1-hop longer disjoint *quasi*-shortest path. This perception justifies the proposed centrality metric, which assigns higher importance to nodes with more disjoint paths to every other node in the network. We do not leave aside the path cost, as shortest paths are preferable.

3 Centrality metrics review

Centrality metrics are frequently used to analyze the relationship between nodes and the role played by each node in the network. Some examples are betweenness, closeness, and their variants [9]. The betweenness concerns the control a node exerts over flows exchanged between all pairs of nodes. The closeness centrality relates to how quickly a flow spreads or reaches all other nodes in the network [17]. Adequate use of a centrality metric depends on understanding the characteristics of the flow process. The closeness definition fits our purpose well, as we consider networks for which flows from the most important node need to continuously reach all other nodes in the network as fast as possible, and vice-versa. We assume that a central node must be multiply-connected to every other node to improve its reachability and, consequently, the availability of any resource the node provides. This section briefly reviews the main metrics related to our proposed metric: closeness and information. We also discuss the ρ-geodesic betweenness centrality, from which we borrow the concept of *quasi*-shortest paths.

3.1 Closeness and information centrality metrics

The *traditional closeness centrality* (C_{trad}) considers the shortest paths to determine the nodes' centrality. The most central node is the closest one, on average, to every other node in the network [5]. The traditional closeness of node υ_s can be computed as in Eq. 1. Using only shortest paths may not suit well when flows do not exclusively follow shortest paths, such as in networks based on rumors [30]. When the communication process is undefined, other paths besides the shortest ones can be considered. It is a commonplace, however, that the longer the path, the less valuable it is [9, 23, 26]. As such, Stephenson and Zelen [30] measure the information flowing through a path $p_{s,t}^{(n)}$ of cost $\delta_{s,t}^{(n)}$ as $I_{s,t}^{(n)} = 1/\delta_{s,t}^{(n)}$. This value is used to compute the *Information centrality* (I) as formally defined by Eq. 2.

$$C_{trad}(\upsilon_s) = \frac{|\mathsf{V}| - 1}{\sum_{t=1}^{|\mathsf{V}|} \delta_{s,t}^{*}}, \tag{1}$$

$$I(\upsilon_s) = \frac{|\mathsf{V}|}{\sum_{t=1}^{|\mathsf{V}|} \frac{1}{I_{s,t}}}, \tag{2}$$

where, $|\mathsf{V}|$ is the number of nodes, $\delta_{s,t}^{*}$ is the shortest-path cost from υ_s to υ_t, and $I_{s,t} = \sum_n I_{s,t}^{(n)}$ is the sum of the information measured between υ_s and υ_t. Note that the information centrality is a closeness variant that accounts for all paths between pairs of nodes. These metrics can be viewed as two extremes of a spectrum, in which the traditional closeness stands on one extreme considering only the shortest paths, and the information centrality stands on the other extreme considering all paths. Our metric lies in-between, as it includes a limited number of disjoint paths starting from the shortest ones.

3.2 ρ-geodesic betweenness centrality

Medeiros et al. [23] focus on multiple paths applied to betweenness centrality. The authors consider a-little-bit longer paths without significant performance loss and propose the *ρ-Geodesic Betweenness Centrality*, formalized as:

$$B_\rho(\upsilon_k) = \sum_{\substack{i \in \mathsf{V} \quad j \in \mathsf{V} \\ \delta_{i,k}+\delta_{k,j}-\delta^*_{i,j} \leq \rho}} \sum \frac{n^*_{i,j}(\upsilon_k) + n_{i,j}(\upsilon_k)}{n^*_{i,j} + n_{i,j}} \times \frac{\delta^*_{i,j}}{\delta_{i,k} + \delta_{k,j}}, \tag{3}$$

where $n^*_{i,j}(\upsilon_k)$ is the number of shortest paths between υ_i and υ_j going through υ_k, and $n^*_{i,j}$ is the total number of shortest paths between υ_i and υ_j.

The ρ-geodesic betweenness extends the traditional definition of betweenness to consider shortest paths and all *quasi*-shortest paths. The number of paths is limited by their cost, which is, in turn, limited by the spreadness factor ρ (Sect. 2). We borrow the definition of *quasi*-shortest paths to propose our metric but we only use the paths that are disjoint, *i.e.*, disjoint *quasi*-shortest paths, limited according to the connectivity factor φ. In a nutshell, the ρ-geodesic betweenness centrality aims to capture the control exerted by a node considering *quasi*-shortest paths, whereas the disjoint multipath closeness centrality aims to capture the reachability of a node considering *quasi*-shortest disjoint paths. Hence, these metrics differ on the meaning of node importance, and also on the paths taken into account to compute the metrics.

4 Related work

Centrality metrics can be applied to several research areas. When applied in computer networking, they can identify strategically positioned nodes, used to improve the network convergence upon failures or congested conditions. Maccari and Cigno [22] use node centrality to optimize the network convergence after failure detection. Node betweenness is used to adjust the interval between transmission of consecutive messages, with more frequent transmissions from central nodes. While it makes sense to use node betweenness, as highly-central nodes control most flows, the paper does not consider nodes that could quickly reach all other nodes in the network so that messages could be delivered faster. Node centrality is also used to optimize traffic engineering based on Software-Defined Networking (SDN) [1], considering that nodes with high betweenness are more prone to congestion. Amaral et al. develop an algorithm that chooses paths composed of nodes wit lower betweenness to carry the traffic and perform load balancing. Nevertheless, Amaral et al. do not consider that these paths can have shared links, overloading the connected nodes. Therefore, taking disjoint paths into account could be an improvement.

The identification of highly-central nodes can be of great importance to network resource or service provisioning. Bouet et al. [10] evaluate the allocation of virtual Deep Packet Inspection (vDPI) functions in a network infrastructure that uses Network Function Virtualization (NFV). The work minimizes the cost to allocate the vDPI by formulating a greedy algorithm based on traditional betweenness. The main drawback of this approach is that nodes with high betweenness are more prone to become overloaded, as most flows traverse them. Nakarmi et al. [24] develop a novel centrality metric to identify critical transmission lines that could originate a cascade failure process in power grids. The authors use a graph based on interactions, where edges denote the interactions between transmission lines. The approach divides the graph into communities and identifies the interaction among communities according to the closeness centrality. The community-based centrality can highlight the nodes in communities that spread the failure cascade process more quickly from the closeness centrality of each community. The community-based centrality proposed by Nakarmi et al. supposes an epidemic flow process, the cascading failures, which is not the same flow process considered in our work. Therefore, the proposal of Nakarmi et al.'s proposal does not fit our problem. Other approaches seek to identify nodes to spread efficiently the information [7, 28]. Hu et al. [20] evaluate the impact of the distance between the initial infected nodes in regular and small-world networks, concluding that larger distances contribute to the recovery density. As Nakarmi et al., Hu et al. consider an epidemic flow process, which is not suitable for our work. The epidemic flow process considers all possible multiple paths that can be simultaneously used with a certain probability that depends on whether the node is the neighbor of an "infected" node, *i.e.*, the node that holds the flow unit at the moment.

We highlight that the adequate use of a centrality metric depends on the flow process. Unlike the approaches presented in this section, the flow process considered in our paper is not epidemic and tends to prioritize the shortest paths. Moreover, we aim to find nodes that are multiply-connected by disjoint paths and can quickly reach all other nodes in the network. Hence, the goal is to capture the connectivity power of a node, *i.e.*, how quickly it can reach and be reached, and how available it is to the other nodes.

Many studies also focus on disjoint path utilization. They are either concerned with path computation, usage of disjoint paths to compute node centrality, or improvements over network communications. Aouiz et al. [2], for instance, develop a novel path centrality metric to find disjoint paths that improve a multipath routing protocol for mobile *ad hoc* networks. The centrality value of a path is an average based on the channel occupancy ratio on nodes at the same path. The routing protocol chooses the optimum path, which is the disjoint path with the lowest proposed centrality. Aouiz et al. do not consider the number of disjoint paths to compute the path centrality, disregarding the availability of nodes to determine their importance.

5 Disjoint multipath closeness centrality

We propose the *disjoint multipath closeness centrality*, a new centrality metric that aims to identify the importance of a node based on its connectivity power. To this end, we consider other paths in addition to the shortest ones, taking into account a fixed number of multiple disjoint paths. The use of such paths is the differential of our metric when compared to the previously presented metrics (Sect. 3). The use of multiple disjoint paths to transfer flows can potentially reduce the overload on shortest paths and, at the same time, improve network resilience. Multiply-connected nodes are more likely to remain accessible even when a failure happens. Our metric captures the relationship between the number and length of a set of disjoint paths between pairs of nodes. The connectivity factor φ limits the number of disjoint paths. We consider that nodes connected by multiple short disjoint paths are better for running critical functionalities on networks that praise resilience.

5.1 Metric formalization

We begin the formalization of our metric by taking into account the contribution of the multiple disjoint paths between υ_s, υ_t. In this case, to compute the importance of υ_s, we propose the utilization of the aggregated cost, $\Delta_{s,t}^{(\varphi)}$, which simultaneously considers the number of disjoint paths and their average cost. On the one hand, the number of disjoint paths to or from a given node contributes to its importance and then must be considered accordingly. On the other hand, more costly paths must contribute less to the importance of a node. This inverse proportionality leads us to compute the aggregated cost based on the harmonic mean of the cost of all disjoint paths connecting υ_s and υ_t divided by the number of paths $\varphi + 1$. The same approach is seen on the "cistern problem", in which a set of pipes with different flows work together to fill a cistern [12, 31]. In this problem, the time needed to fill a cistern is the harmonic mean of the time spent by each faucet divided by the number of pipes. In our scenario, the aggregated cost represents the mean resistance to traverse all disjoint paths, weighted by the number of disjoint paths. Hence, the aggregated cost is calculated as an harmonic mean weighted by the number of disjoint paths. We highlight that the number of disjoint paths must be less than or equal to φ.

$$\Delta_{s,t}^{(\varphi)} = \frac{\varphi + 1}{\sum_{n=0}^{\varphi} \frac{1}{\delta_{s,t}^{(n)}}} \times \frac{1}{\varphi + 1} = \frac{1}{\sum_{n=0}^{\varphi} \frac{1}{\delta_{s,t}^{(n)}}}. \tag{4}$$

$$C_\varphi(\upsilon_s) = \frac{|\mathsf{V}| - 1}{\sum_{t=1}^{|\mathsf{V}|} \Delta_{s,t}^{(\varphi)}}. \tag{5}$$

The *disjoint multipath closeness centrality* of node υ_s is defined by a generalization of Eq. 4, which is obtained considering all network nodes. Equation 5 presents the proposed metric, C_φ, where $|\mathsf{V}|$ is the number of nodes in the network, and $|\mathsf{V}| - 1$ is a normalization factor that allows comparing the centrality of nodes in networks of different dimensions.

We highlight that for $\varphi = 0$, the proposed metric reduces to the traditional closeness centrality as we only consider the shortest path. The disjoint multipath closeness centrality has the following characteristics: (i) it considers both the cost and the number of paths; (ii) it takes into account shortest paths and the largest possible set of disjoint

Algorithm 1 DISJOINT MULTIPATH CLOSENESS CENTRALITY OF NODE υ_s.

```
Require: φ, υ_s, G(V, E, W)
Ensure: C_φ(υ_s)
DJ ← ∅
for t ← 1, | V | +1 do
    if υ_s ≠ υ_t then
        H ← INITIALIZE(G)
        DJ[t] ← FIND_DISJ_PATHS(H, s, t, φ)
    end if
end for
C_φ(υ_s) ← ACCUMULATE(φ, | V |, DJ)
return C_φ(υ_s)
```

paths, limited by a connectivity factor (φ); (iii) it assigns more importance to paths with lower costs.

5.2 Implementation

The computation of the proposed disjoint multipath closeness centrality requires design decisions described in this section. The first one concerns path selection. It is already clear that we use multiple disjoint paths. Nevertheless, these paths can be different according to the selection approach. The set of paths may influence the centrality value obtained for a node. It is worth mentioning that, in this paper, we use undirected links with unitary weights for the sake of simplicity. As such, the total cost of a path is equal to the number of hops in this path. Nevertheless, the proposed metric is generic enough to be also used in directed and weighted graphs. We start by explaining the main algorithm, Algorithm 1, used to compute the proposed metric for a node υ_s. Then, we continue by presenting two auxiliary algorithms, Algorithms 2 and 3, respectively used to compute the set of φ shortest disjoint paths from υ_s to every other node in the network and the function used to update the metric value.

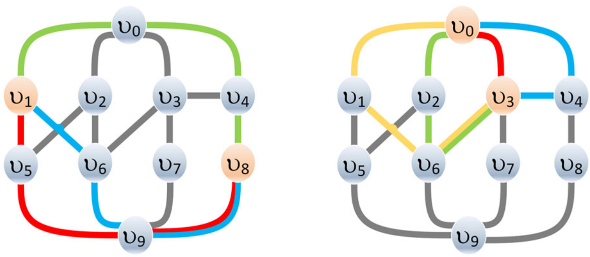
The inputs of Algorithm 1 are the graph G, the node υ_s, and the connectivity factor φ. This algorithm returns the output $C_\varphi(\upsilon_s)$, which is the value of the proposed metric obtained for node υ_s. In the following algorithms, the source (υ_s) and target (υ_t) nodes are represented by their indexes, s and t. In Algorithm 1, we initialize the vector DJ, that will contain the disjoint paths costs considered between υ_s and every other node in the network. We explain in detail this vector in the next steps. The graph H receives a copy of the original graph G at each iteration. The following operations are performed on graph H, without changing the graph G. The algorithm iterates over one node in the network at a time, the *source* υ_s. The shortest paths and the disjoint paths found between υ_s and υ_t by the FIND_DISJ_PATHS function (Algorithm 2) are stored in the vector DJ[t]. Then, $\texttt{DJ[t]} = [\delta_{s,t}^{(0)}, \ldots, \delta_{s,t}^{(\varphi)}]$. Note that there are up to φ disjoint paths between a pair of nodes. This means that the first element of DJ[t], $\delta_{s,t}^{(0)}$, is the cost of the shortest path between υ_s, υ_t; the second element, $\delta_{s,t}^{(1)}$, is the cost of the shortest disjoint path; the third element, $\delta_{s,t}^{(2)}$, is the cost of the second shortest disjoint path; and so on, until the last element, $\delta_{s,t}^{(\varphi)}$, which is the φ-th-shortest disjoint path. Therefore, vector DJ contains t vectors, where each one contains the cost of all considered paths between υ_s and a different node υ_t. The contribution of the paths to the node centrality is computed by function ACCUMULATE, as described in Algorithm 3. The output of this function is the disjoint closeness centrality for the *source* υ_s.

Algorithm 2 finds the set of paths used to compute the node centrality. This set is composed of the φ shortest-disjoint paths between a pair of source-target nodes. The inputs of the algorithm are the network graph H; the corresponding indices of nodes υ_s and υ_t, s, t, respectively; and the connectivity factor φ. In the FIND_DISJ_PATHS function, costVec is a vector that stores the costs of the disjoint paths discovered by SHORTDISJ. Moreover, this function also returns the nodes from the discovered disjoint paths and the number of such paths. The nodes are stored in vector nodesVec, while auxPath stores the number of disjoint paths. The vectors are initially empty and auxPath receives 0 as initial value. Note that vector DJ[t] is the output of this algorithm, and this vector is an element of vector DJ, which contains the cost of all paths considered for computing the metric for υ_s. The search for paths initiates in the while-loop, controlled by φ. In Line 6, we append the elements of costVec to DJ[t] to update the costs of the disjoint paths found at each iteration. While φ is not reached, the intermediate nodes belonging to the paths already found are removed from the graph (Line 7), and the search for paths continues in the remaining graph. The while-loop is terminated if one of two conditions is met: (i) $\varphi + 1$ paths are found, or (ii) there are no new disjoint paths between the source and target nodes.

Algorithm 2 FIND_DISJ_PATHS FUNCTION.

Require: H, s, t, φ
Ensure: DJ[t]
1: nodesVec $\leftarrow \varnothing$; DJ[t] $\leftarrow \varnothing$; costVec $\leftarrow \varnothing$
2: paths $\leftarrow$ 0; auxPath $\leftarrow$ 0
3: **while** paths $\leq \varphi$ **do**
4: (costVec, nodesVec, auxPath) $\leftarrow$ SHORTDISJ(H, s, t, φ, paths)
5: **if** costVec $= \varnothing$ **then break**
6: **end if**
7: DJ[t] $\leftarrow$ APPENDELEMENTSTO(DJ[t], costVec)
8: H $\leftarrow$ REMOVENODESFROM(H, nodesVec)
9: paths $\leftarrow$ paths + auxPath
10: **end whilereturn** DJ[t]

Figure 2 illustrates the search for disjoint paths. In this example, we use $\varphi = 3$ additional disjoint paths between pairs of nodes. In Fig. 2a, there are 3 shortest paths between υ_1 and υ_8, $p^{(0)}_{1,8} = \langle \upsilon_1, \upsilon_5, \upsilon_9, \upsilon_8 \rangle$ (red), $p^{(1)}_{1,8} = \langle \upsilon_1, \upsilon_0, \upsilon_4, \upsilon_8 \rangle$ (green), and $p^{(2)}_{1,8} = \langle \upsilon_1, \upsilon_6, \upsilon_9, \upsilon_8 \rangle$ (blue), all of them with $\delta^*_{1,8} = 3$. Path $p^{(1)}_{1,8}$ (green) is disjoint to the other two and, thus, it is automatically selected. As only disjoint paths matter, we randomly select between paths $p^{(0)}_{1,8}$ (red) and $p^{(2)}_{1,8}$ (blue). If the choice is $p^{(0)}_{1,8}$ (red), then the algorithm stores this path and searches for additional disjoint paths to both $p^{(0)}_{1,8}$ (red) and $p^{(1)}_{1,8}$ (green). There are no disjoint paths to the shortest ones selected and only $p^{(0)}_{1,8}$ and $p^{(1)}_{1,8}$ are considered. The search terminates even if $\varphi = 3$ is not reached, *i.e.*, condition (ii) is met. Differently, Fig. 2b shows an example in which condition (i)



(a) Total number of disjoint paths less than φ.

(b) Total number of disjoint paths equal to φ.

Fig. 2 The number of computed paths does not always reach φ. It is possible that **a** no disjoint *quasi*-shortest path is found, or **b** *quasi*-shortest paths disjoint to the shortest one and disjoint to each other are found and then any of them can be considered to compute the metric

is met. The maximum number of disjoint paths that may be considered between nodes υ_0 and υ_3 is reached. The shortest paths $p^{(0)}_{0,3}$ with cost $\delta^*_{0,3} = 1$ is highlighted in red. This is the only shortest path. There are other three *quasi*-shortest paths disjoint to the shortest one: $p^{(1)}_{0,3}$ (blue) with cost $\delta_{0,3} = 2$; $p^{(2)}_{0,3}$ (green) with cost $\delta_{0,3} = 3$; and $p^{(3)}_{0,3}$ (yellow) also with cost $\delta_{0,3} = 3$. Paths $p^{(2)}_{0,3}$ and $p^{(3)}_{0,3}$ are not disjoint to each other. In case $p^{(2)}_{0,3} = \langle \upsilon_0, \upsilon_2, \upsilon_6, \upsilon_3 \rangle$ is considered, there is one more disjoint path with 5 hops. The same occurs in case $p^{(3)}_{0,3} = \langle \upsilon_0, \upsilon_1, \upsilon_6, \upsilon_3 \rangle$ is considered. Thus, either $p^{(2)}_{0,3}$ or $p^{(3)}_{0,3}$ is selected at random, and the respective 5-hop disjoint path is added to the set of paths used to compute the node centrality.

The ACCUMULATE function updates the value of the disjoint multipath closeness centrality for the *source* υ_s. This function receives as input the value of φ, the number of nodes `numNodes`, and the cost of disjoint paths found for *source* υ_s stored in `DJ`. For each target υ_t, the algorithm calculates the aggregated cost of the paths found, $\Delta_{s,t}$, with the aid of an auxiliary variable auxSum. Then, the variable costAcc accumulates the aggregated cost to achieve each target υ_t from the examined source. Finally, after accumulating all aggregated costs for all targets, the closeness centrality of the source node is calculated. The value found is stored in $C_\varphi(\upsilon_s)$, which is the output of the ACCUMULATE function. This function updates for the *source* υ_s the value of the proposed centrality according to Eqs. 4 and 5. After these iterations, the disjoint multipath closeness centrality for the *source* υ_s, $C_\varphi(\upsilon_s)$, is obtained.

5.3 Time complexity

The disjoint multipath closeness centrality of node υ_s, $C_\varphi(\upsilon_s)$, is given by Algorithm 1. This algorithm's time complexity is governed by the time complexity of functions FIND_DISJ_PATHS (Algorithm 2) and ACCUMULATE (Algorithm 3).

The complexity of computing the $(\varphi + 1)$-disjoint paths between two nodes is $O(|E|)$ [18]. Algorithm 2 uses Dijkstra's algorithm to compute shortest paths and, from these paths, compute the disjoint ones (i.e., *quasi*-shortest paths). As Dijkstra's

Algorithm 3 ACCUMULATE FUNCTION.

Require: φ, $\mid \mathtt{V} \mid$, $\mathtt{DJ}$
Ensure: $\mathrm{C}_{\varphi}(\upsilon_s)$
costAcc $\leftarrow 0$; auxSum $\leftarrow 0$; $\Delta_{s,t} \leftarrow 0$
for $\mathtt{t} \leftarrow 1, \mid \mathtt{V} \mid +1$ **do**
  if $\upsilon_s \neq \upsilon_t$ **then**
    for $\mathtt{n} \leftarrow 0, \varphi + 1$ **do**
      auxSum $\leftarrow$ auxSum $+ \frac{1}{\mathtt{DJ[t][n]}}$
    end for
    $\Delta_{s,t} \leftarrow \frac{1}{\text{auxSum}}$
    costAcc $\leftarrow$ costAcc $+ \Delta_{s,t}$
  end if
end for
$\mathrm{C}_{\varphi}(\upsilon_s) \leftarrow \frac{|\mathtt{V}|-1}{\text{costAcc}}$
return $\mathrm{C}_{\varphi}(\upsilon_s)$

algorithm is $O(\mid E \mid + \mid V \mid \log \mid V \mid)$, Algorithm 2 is $O(2 \mid E \mid + \mid V \mid \log \mid V \mid)$, which is approximately $O(\mid E \mid + \mid V \mid \log \mid V \mid)$. Algorithm 3 is $O(\varphi \mid V \mid)$, where φ is a constant. Therefore, Algorithm 1 has time complexity composed by time complexity of Algorithm 2 and 3 so that it becomes $O(\mid V \mid ((\mid E \mid + \mid V \mid \log(\mid V \mid)) + \varphi \mid V \mid))$, which is equal to $O(\mid V \mid^2 \log(\mid V \mid) + \mid V \mid (\mid E \mid + \varphi))$. For sparse networks, $\mid V \mid \gg \mid E \mid$, Algorithm 1's complexity becomes $O(\mid V \mid^2)$. In turn, for dense networks, where $\mid E \mid \approx \mid V \mid^2$, the complexity becomes $O(\mid V^3 \mid)$.

5.4 Numerical example

We numerically introduce the impact of the disjoint multipath closeness centrality using Fig. 2a. In this figure, there are 10 nodes and 16 links. Table 1 presents the node ranking in ascendant order according to the proposed metric, considering $\varphi = 1$. We also show the node ranking for the traditional closeness ($C_{trad}(\bullet)$) and the information centrality ($I(\bullet)$). The value obtained for each node using each centrality metric appears in parentheses.

We can divide the node ranking in Table 1 into 3 sets, Γ. Each set is composed of the same set of nodes for all metrics, but with a different ordering, $\Gamma_1 = \{\upsilon_0, \upsilon_3, \upsilon_6, \upsilon_9\}$, $\Gamma_2 = \{\upsilon_1, \upsilon_2, \upsilon_4, \upsilon_5\}$, and $\Gamma_3 = \{\upsilon_7, \upsilon_8\}$. Focusing on Γ_1, both the traditional closeness and the information centrality result in the same ranking, with nodes υ_0, υ_3, and υ_9 tied in the second position. Differently, the proposed metric can break ties, classifying the nodes in different positions, even if we use a small connectivity factor such as the one in the example. In this scenario, the tiebreaking occurs due to the disjoint paths from node υ_3 having lower costs to reach the other nodes than the multiple paths from nodes υ_0 and υ_9. Looking at Γ_2, the only nodes for which there is a tie in the disjoint multipath closeness centrality are υ_1 and υ_2. This happens because the number and length of the disjoint paths for these nodes are the same. In addition, both nodes have the same value when we consider the eigenvector centrality. We also observe in Table 1 that some nodes lose more positions than other nodes. For instance, comparing the traditional closeness with the other metrics, υ_4 is the node that loses more positions. This is because of the costs of node υ_4's disjoint paths, which are greater than the

Table 1 Node ranking according to the traditional closeness, information and disjoint multipath closeness metrics, for the network graph depicted in Fig. 2

$C_{trad}(\bullet)$		$I(\bullet)$		$C_{\varphi=1}(\bullet)$	
Rank	Node (Centrality)	Rank	Node (Centrality)	Rank	Node (Centrality)
1	**v_6 (0.64)**	1	**v_6 (0.18)**	1	**v_6 (1.07)**
2	**v_0 (0.6)**	2	**v_0 (0.17)**	2	**v_3 (1.05)**
2	**v_3 (0.6)**	2	**v_3 (0.17)**	3	**v_9 (1.02)**
2	**v_9 (0.6)**	2	**v_9 (0.17)**	4	**v_0 (1.0)**
5	*v_4 (0.56)*	5	*v_1 (0.15)*	5	*v_1 (0.95)*
6	*v_1 (0.53)*	5	*v_2 (0.15)*	5	*v_2 (0.95)*
6	*v_2 (0.53)*	5	*v_5 (0.15)*	7	*v_4 (0.93)*
6	*v_5 (0.53)*	8	*v_4 (0.14)*	8	*v_5 (0.9)*
9	v_7 (0.5)	9	v_7 (0.12)	9	v_7 (0.85)
9	v_8 (0.5)	10	v_8 (0.11)	10	v_8 (0.82)

ones for nodes v_1 and v_2, for instance. In the set Γ_3, it is difficult for v_7 and v_8 to improve their positions. The shortest paths attached to these nodes are already longer than those attached to the other nodes in the network. Moreover, the topology does not favor the presence of multiple low-cost disjoint paths for the nodes in this set. Even in this unfavorable situation, our metric can break the tie obtained for the traditional closeness for nodes v_7 and v_8.

5.5 Discussions

We choose the φ value according to the ranking stability to show the benefit of using the disjoint *quasi*-shortest paths. In practical scenarios, the value of φ is typically fixed and defined by the network administrator. For instance, according to an internal policy, a telecommunications operator sets the number of alternative paths towards a client network. Alternative paths can be used for backup reasons or because a single path may not minimize at the same time different Quality of Service (QoS) metrics (e.g., latency and bandwidth). If we do not use a fixed value, we could find the optimal φ value. As characteristics such as node degree and network density influence the number of additional disjoint paths, they probably affect the optimal φ. Our results confirm that the network density impacts on the performance of the proposed metric.

The proposed centrality can benefit several scenarios where the network topology allows multiple disjoint paths, improving the performance of applications deployed in real networks. In this work, we use both computer and social networks, and we focus on the applicability of our metric in such networks. Computer networks are well established and keep growing with the increase of connected devices. Adaptation to new applications and demands originates new paradigms, such as edge computing and Network Function Virtualization (NFV), for instance. These paradigms enable more distributed applications and services closer to users. In NFV applications, it is paramount that a network function, such as a firewall running atop a virtual machine, be reachable by the network nodes, even in case of a link failure. Hence, the network topology must be designed to provide connectivity. Moreover, reducing the distance between all nodes and the Virtualized Network Function (VNF) can improve the overall network latency. Thus, reducing the distance and simultaneously improving connectivity is a dual goal that must be taken into account. Our metric ranks the most central node as the one that can reach all other nodes using multiple shortest disjoint paths. This node can host the VNFs, and such a deployment contributes to achieving the dual goal. In multihop wireless scenarios, a wireless node playing the role of a gateway interconnecting different networks cannot undergo moments of lack of connectivity if neighboring nodes fail. Again, locating a gateway is a challenge that cannot exclusively consider only one path from the gateway to every other node. In this sense, the most central node of the proposed metric can assume a relevant role in the network because this node is closest to all other nodes by multiple disjoint paths. In social networking, a node should not stop spreading content if a neighbor node leaves the network. In this case, a node with more alternative paths to reach farther participants could be preferable than another that strongly relies on a smaller set of alternative paths, even if they are comparably shorter. For instance, in the Train Bombing network, multiple disjoint paths enable the information to reach the target by different intermediaries, avoiding communication with the same people. Moreover, in the event of communication failure in one path, alternative paths would ensure information delivery.

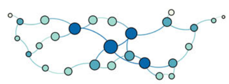
The previous examples are not exhaustive. Our metric fits well in scenarios where the tradeoff between quick and resilient reachability exists. This tradeoff is not considered by the traditional closeness, which focuses on the fastest possible reachability, *i.e.*, it considers only shortest paths. In turn, the information centrality considers all existing paths, which may not be suitable for some applications. For instance, considering geographic distance as a limitation for the application, if a route exists between a server in the United States and a client in Brazil via the west coast of the Americas, it is better to deploy a backup route between these countries via the east coast than via Europe. Hence, focusing on practical limitations, alternative paths should not be much longer than the shortest ones. For last, we borrow the concept of *quasi*-shortest paths from the ρ-geodesic betweenness, even though this metric measures other node characteristics. As it is based on the betweenness centrality, the goal is to find nodes that intermediate flows, *i.e.*, the most important nodes have more control over all flows that travel through the network. Our proposal, in turn, is based on the closeness centrality, whose goal is to find nodes that can quickly reach all other nodes. Our metric adds to this goal the requirement of being quickly and, at the same time, resiliently reached.

6 Results and discussions

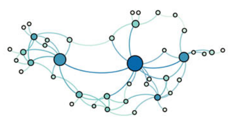
We compare our metric with others considering similar flow processes and capturing similar roles played by nodes. A typical evaluation is the impact over the node ranking, as seen in the literature [7, 11, 23, 26, 28, 30]. The goal is to show that the proposed centrality metric ranks the nodes differently from other similar centrality metrics, electing different nodes as the most important ones. We assess the impact of our metric in three steps. First, (i) we investigate how the increase in the connectivity factor influences the number of paths, and how the additional paths considered in computation for each φ impact the node ranking. Then, (ii) we use Kendall's concordance coefficient (Kendall's W) to verify the level of agreement between the rankings of our proposal, the traditional closeness, and the information metric. We also evaluate Kendall's W for our metric, varying the connectivity factor. Finally, (iii) the reachability of nodes is evaluated when a failure occurs.

6.1 Datasets

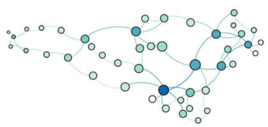
We evaluate our proposal using five datasets with different characteristics, representing real long-distance educational and research network topologies and social network topologies. Figure 3 shows a representation of these datasets in which bigger nodes have higher degree, while darker nodes have higher traditional closeness. The long-distance network topologies are the following. **RNP** describes the Ipê network [13], which interconnects Brazilian universities, research, and cultural institutes. We use the 2014 topology [14] with 33 links and 27 nodes. The average node degree is 2.4, the diameter is 9, and the density is 0.094. **Renater** represents the topology of the network infrastructure that serves educational and research purposes in France [29]. We use the 2013 topology, with 45 nodes connected by 61 links. This network has an average node degree of 2.7, diameter equal to 11, and a density of 0.062. **Géant** interconnects national research networks in Europe [16]. We use the 2013 topology, with 42 nodes interconnected by 68 links. The average node degree is 3.2, diameter is 7, and density is 0.079.



(a) RNP: 27 nodes, 33 edges, average degree 2.4.

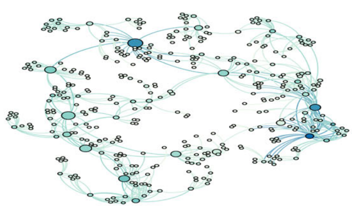


(b) Géant: 42 nodes, 68 edges, average degree 3.2.



(c) Renater: 45 nodes, 61 edges, average degree 2.7.

(d) Train Bombing: 64 nodes, 243 edges, average degree 7.6.



(e) Netscience: 379 nodes, 914 edges, average degree 4.8.

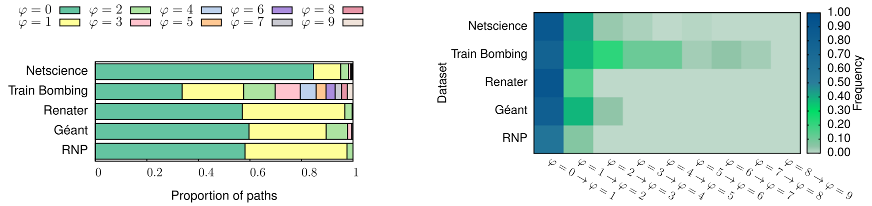
Fig. 3 Network topologies **a** RNP, **b** Géant, **c** Renater, **d** Train Bombing e **e** Netscience. Bigger nodes have higher degree, while darker nodes have higher traditional closeness

The social networks are the following. **Train Bombing** describes the relationship between 64 suspected terrorists involved in the train bombing of Madrid on March, 2004 [19]. Nodes are connected by 243 directed edges, which represent the contact between suspects. Without loss of generality, we consider that edges are undirected and have unitary weights. The network has an average node degree of 7.6, a diameter of 6, and density of 0.121. **Netscience** models the relationship of 379 scientists in 2006 [25, 27]. There is an edge between two scientists if they are co-authors. The complete network has 1, 589 scientists, but we use the largest connected component, which has 379 scientists. There are 914 edges connecting these scientists. The network has an average node degree of 4.8, a diameter of 17, and a density of 0.013.

6.2 Influence of φ on the number of paths and on node ranking

The disjoint multipath closeness centrality depends on the connectivity factor φ, as φ is directly related to the number of paths considered to compute the metric. Hence, it is important to know how the φ influences in reclassification to find nodes once neglected by the other metrics. Figure 4 presents how φ influences the increase of the number of paths and, consequently, the node ranking. Firstly, we investigate the number of disjoint paths added as we increase φ. Figure 4a shows the ratio between the number of paths added by each φ increment and the total number of paths found, which is the number of paths for $\varphi = 9$ for each network. The total number of paths for RNP, Géant, Renater, Train Bombing, and Netscience are 604, 1443, 1735, 6032, and 84613, respectively. The number of paths refers to the paths between all pairs of nodes. We highlight that, for $\varphi = 0$, no additional disjoint path is accounted, and our metric is equivalent to the traditional closeness. We stop our evaluation at $\varphi = 9$ because we do not find new disjoint paths for all selected networks for values greater than 9.

We observe that when increasing φ, more disjoint paths are included. Nevertheless, the number of additional paths becomes smaller as φ grows until no more paths are included. We note that the greatest number of additional paths is found when we use $\varphi = 1$ and that each network reaches the maximum number of disjoint paths with a different value of φ. These values are 2, 5, 4, 9, and 8 for RNP, Géant, Renater, Train Bombing, and Netscience, respectively. This result indicates that the additional disjoint paths and the value of φ stabilizing the node ranking depend on the network



(a) The proportion of disjoint paths added is smaller for higher φ values.

(b) The total number of additional paths affect the node reclassification differently for each φ.

Fig. 4 Influence of φ on **a** the number of paths and **b** the node ranking topology. Rank stabilization in this context means that node reclassification is no longer possible because their disjoint multipath closeness does not change, as there are no additional paths when φ varies. Hence, we note that our metric is influenced by the network density, as the higher the number of nodes and links, the higher the number of alternative paths. In these conditions, it is more likely to find additional disjoint paths.

There are many paths added as we increase φ. Nevertheless, in Netscience, this proportion is lower, which is a consequence of its lower density. When $\varphi = 0$, we already use 85% of the total number of paths, and thus the number of disjoint paths considered by our metric represents only 15% of the total number. This is quite similar to the traditional closeness. On the other hand, in the Train Bombing network, $\varphi = 0$ concentrates only 34% of the paths, as shown in Fig. 4a. Higher density networks likely provide more disjoint paths, benefiting our proposed metric. Hence, the higher density observed at the Train Bombing network increases the impact of our metric.

We also investigate the occurrence changes in the node ranking when we increase φ. Figure 4b shows the frequency of node reclassification for each network. We compute the reclassification frequency as the ratio between the number of nodes that change ranking position for each φ increment and the total number of nodes in the network. The highest reclassification frequency happens for the Renater network, reaching 91% when we increase the connectivity factor from $\varphi = 0$ to $\varphi = 1$. Although Renater and RNP have the largest proportion of additional paths when increasing φ to 1, as seen in Fig. 4a, the effect of these additional paths on the node ranking is different for these networks. In RNP, 59% of the nodes are reclassified, which is smaller than the result found for Renater. We observe that the reclassification frequency is more prominent when we change φ to 1, given the higher proportion of additional paths considered. The obtained results also suggest that the total number of paths added influences the reclassification frequency.

Observe that the high reclassification frequency found in all networks results from the tiebreaking between nodes classified in the same position according to the traditional closeness ($\varphi = 0$). For instance, nodes 22 and 10 from RNP are in the second position according to the traditional closeness, but node 10 loses one position when ranked by our metric. We note that the top-ranked node tends to coincide, independently of the centrality metric used [17]. This is true, except for Renater, where the first position of the traditional closeness ranking, occupied by node 24, is occupied by node 37 in our metric's ranking.

Figure 4b also shows the maximum value of φ for which each ranking becomes stable, *i.e.*, nodes do not change position anymore. Ranking stability is achieved in Renater and RNP networks for $\varphi_{max} = 2$. Géant, Train Bombing, and Netscience, on the other hand, achieve stability for φ_{max} equal to 3, 8, and 6, respectively. This means that even though new paths are added, this may not be enough to reflect on ranking position changes. Hence, for practical reasons, we call the value of φ for which no more ranking changes occur as the φ_{max} of each network.

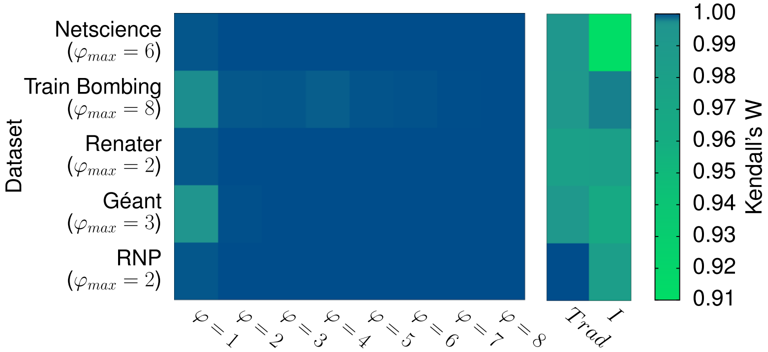


Fig. 5 Kendall's W concordance coefficient shows the level of agreement between the rankings produced by $C_{\varphi_{max}}$

6.3 Metrics concordance

We compute Kendall's W coefficient [15] to investigate the concordance level between the evaluated metrics. This coefficient indicates the degree of association of two or more rankings for the same samples, each ranking ordered by a rater. Hence, the coefficient measures the agreement among these raters when they assess the same samples [6, 21]. In this paper, the raters are the metrics, which produce different rankings. For our analysis, a high Kendall's W coefficient means that the centrality metrics apply the same standard to assess the dataset and the same general properties of the dataset are measured, which would not change the metric interpretation. In this sense, it is important to vary φ to assess whether the increase in the number of disjoint paths changes the node ranking without changing the metric interpretation.

Figure 5 shows the Kendall's W coefficient value for the φ_{max} of each network compared with different values of φ and also with the traditional closeness (C_{trad}) and information centrality (I). We observe that the increase in the connectivity factor φ does not imply significant changes in Kendall's W concordance coefficient for any network. In the evaluated datasets, this means that there is a high level of agreement between the C_φ for the various φ, generating slightly different rankings. We observe in Fig. 5 that our metric and the traditional closeness (C_{trad}) have the highest concordance for the RNP network, among all networks. In addition, the information centrality has the lowest concordance with our metric for all networks, excluding the Train Bombing network. This network is denser and has a higher average degree than all the other networks, which impacts the number of disjoint paths. The higher number of paths for the Train Bombing network can be seen in Fig. 4a, then higher values of φ for this network, most of the existing paths are included in the computation of our metric, bringing the result closer to the information centrality. In the Netscience network, Kendall's W shows the opposite behavior compared to the information centrality as this network has the lowest density. We draw two conclusions from the analysis of Kendall's W coefficient results. On the one hand, in dense networks with high values for φ, the proposed metric tends to the information centrality. On the other hand, in sparse networks, the proposed metric tends to the traditional closeness centrality.

This is a direct consequence of the number of additional disjoint paths considered for computing the proposed metric.

6.4 Node reachability

The proposed disjoint multipath closeness centrality aims at assigning more importance to nodes with greater reachability using the connectivity factor φ. As shown in our previous results, this design decision changes the ranking generated by the proposed metric and the one generated by the traditional closeness and information centrality. As a consequence, we investigate in this section the node ranking behavior in failure-prone networks. The goal is to understand the reachability changes upon network failures. We conduct two different evaluations considering that nodes can access other nodes through shortest or disjoint paths. The first evaluation investigates the cost to reach the top-ranked node according to both the traditional closeness and our proposed metric when a single failure occurs. We only use the Renater network because it is the only one where the top-ranked node changes. Following, We consider the set of 20% most central nodes according to the traditional closeness, information centrality, and the top-ranked one according to our proposed metric when up to 3 nodes failure occurs.

6.4.1 Single-node failure

We investigate the reachability of the most central node in Renater when a single failure occurs. Firstly, we consider a failure in a random network node, and secondly, the failure occurs in a node from any shortest path. To this end, we remove the selected node from the network, and after each removal, we evaluate if the remaining nodes are still able to reach the most central node. We also compute the difference between the average path cost to reach the most central node before and after the failure for each remaining node, as shown in Fig. 6. The x-axis shows the remaining node identification, while the y-axis shows the difference between the average path cost before and after the failure. The horizontal line is the average difference considering all nodes, before and after the failure. For the traditional closeness, we only consider the shortest path cost. In contrast, for the disjoint multipath closeness, we consider the average cost of the multiple disjoint paths that satisfy the corresponding connectivity factor φ. In this analysis, $y > 0$ indicates an increase in the average cost for accessing the most central node. Hence, the higher the y-value, the greater the increase in the average cost. In turn, negative values indicate that the average path cost becomes smaller after the failure. Negative values occur when the single failure eliminates a *quasi*-shortest path previously used, but the node does not lose access to the most central node. We fail all possible nodes. Thus, we repeat the same procedure, one-node removal, and evaluation, for each node in the network (random-node failure) or each node in any shortest path (shortest-path-node failure).

Figure 6a shows that, after a random node failure, the remaining nodes reach the most central node of our metric with the smallest increase in the average path cost. The average path cost difference is equal to -0.0023 for $\varphi = \{2, 3\}$ and 0.015 for $\varphi = 1$,

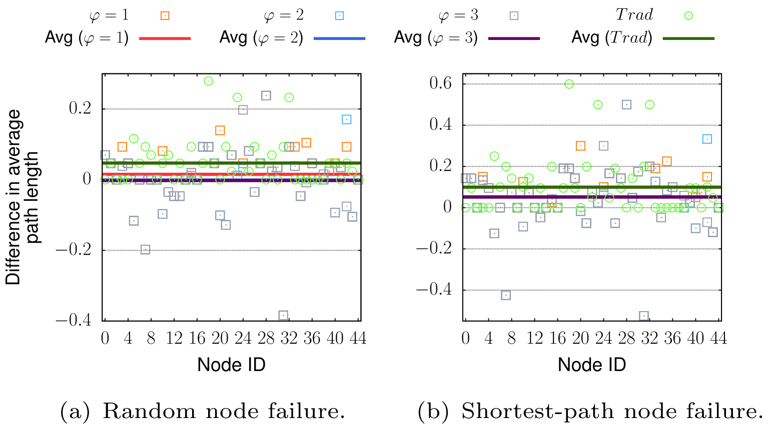


Fig. 6 Average path increase/decrease after a single node failure considering all paths used in the traditional closeness and our metric for both failure **a** on a random node and **b** on a shortest-path node

whereas for the traditional closeness, the average difference is 0.047. The negative average for $\varphi = \{2, 3\}$ is due to the loss of *quasi*-shortest paths. For instance, the difference in the average path length for node v_{32} to reach node v_{37} is -0.3937. The negative value appears because, before the failure, the average length of the shortest and the *quasi*-shortest paths is 7; whereas after the failure, it becomes 6.6163. Still, the proposed metric has the lowest difference in the average path length. When a failure occurs in a node from the shortest path, the average of differences is equal to 0.052 for $\varphi = \{1, 2, 3\}$, while for the traditional closeness, the average is 0.099, as shown in Fig. 6b. Despite the small values, the average path cost according to the proposed metric is closer to 0, in both cases. We also verify the number of nodes that lose communication with the most central node identified by the traditional closeness and our metric when the failure happens on a node from the shortest path. In the worst case, 4 nodes of Renater lose communication with the top-ranked node identified by the traditional closeness. In turn, the worst scenario results in 3 nodes losing communication with the top-ranked node of our metric.

6.4.2 Multiple-node failure

We investigate the impact on communications when multiple failures happen on a set composed of the 20% most central nodes of each metric. Failures happen on shortest paths only, and we evaluate the results for $\varphi \leq \varphi_{max}$, for each network. We fail up to three nodes and verify the number of nodes that lose communication with the set of most central nodes identified by each metric. For a single failure, we repeat the same methodology employed in Sect. 6.4.1 for failures in shortest-path nodes. Thus, we fail a different node at each run, where the number of runs is equal to the number of nodes in at least one shortest path. For two and three failures, we randomly choose 200 different node combinations, being 100 for two and 100 for three failures. Then, we fail one combination at each run, totaling 200 runs. Figure 7 presents the results for each network. The y-axis is the average number of nodes that lose communication

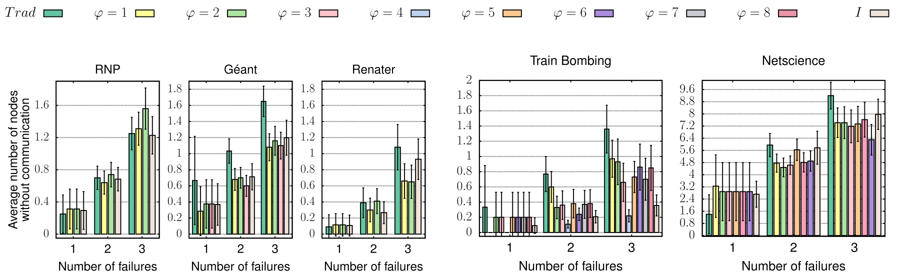


Fig. 7 The average number of nodes that lose communication with the set of 20% most central nodes according to the traditional closeness, the disjoint multipath closeness centrality and Information metrics. We consider $\varphi \in [1, \varphi_{max}]$

with at least one of the nodes from the set of 20% most central nodes. The x-axis indicates the number of nodes that simultaneously fail.

Figure 7 shows that, in all scenarios, the average number of nodes that lose communication with at least one of the 20% most central nodes increases with the number of failed nodes. Considering all networks, RNP is the only one where inversions occur for multiple failures given its size and consequent low number of additional disjoint paths. Whereas a single failure eliminates only the shortest path, multiple failures can eliminate disjoint paths in addition to the shortest paths. Therefore, the traditional closeness demonstrates better performance in this network. The top 20% set corresponds to the 5 most central nodes. Hence, if the average number of nodes losing communication is greater than 1, at least 1 node loses communication with one of the 5 most central nodes. In the Géant network, the 20%-set corresponds to the 8 most central nodes. The average number of nodes without communication with one node from the top 20% set is always smaller when classification follows our metric or the information centrality. On average, our metric performs better. Hence, when using multiple disjoint paths, including both shortest and *quasi*-shortest paths, fewer nodes lose communication with the 20% most central nodes. It is noteworthy that, for two or three failures, there is a slight increase in the average when φ changes from 1 to 2, or 3. We believe that this variation is related to the set of selected failure samples.

For the Renater network, the top 20% set corresponds to the 9 most central nodes. For a single failure, the difference between the metrics is negligible. For two failures, the average number of nodes without communication is still very similar for all metrics. Nevertheless, fewer nodes lose communication with the top 20% set as classified by the disjoint multipath closeness for three failures, even compared with the results of a metric that also considers multiple paths, such as the information centrality. In the Train Bombing network, the 20%-set is composed of 13 nodes. For this network, the communication with the top 20% set classified by information is less impacted by a single failure. For some values of φ, such as $\varphi = \{1, 4\}$ the results are even better than for the information centrality. Nevertheless, other values of φ perform worse than this metric. For multiple failures, $\varphi = 4$ shows the best results, followed by the information centrality. We believe this happens due to the high density of the Train Bombing network. For the Netscience network, the set of 20% most central nodes corresponds to 76 nodes. For a single failure, the average number of nodes without communication is better for the traditional closeness. Nevertheless, when we consider 2 and 3 failures, our metric presents better results, on average, even compared to the other metrics, for all values of φ. The low density of Netscience reduces the availability of alternative paths. Therefore, when a failure occurs, there is a greater probability of precluding the communication between a pair of nodes. Nevertheless, the disjoint multipath closeness presents the best result, reaching approximately 2 nodes more compared with the traditional closeness, and 1 node more compared with the information centrality.

The results show that for the studied networks, the disjoint multipath closeness elects nodes that tend to remain connected when the network undergoes node failure conditions, considering that failures happen on shortest paths and that information flows are not restricted to these paths.

7 Conclusion

We proposed a new centrality metric, the disjoint multipath closeness centrality, that considers shortest and disjoint *quasi*-shortest paths to compute node importance. The proposed metric weights the importance of a path inversely proportional to its cost, as information flows prefer shorter paths. The idea is to highlight nodes that are multiply-connected to all other network nodes, being more suitable to act in high-availability roles. The connectivity factor φ is the maximum number of additional disjoint *quasi*-shortest paths to be accounted between a pair of nodes. We evaluate our metric through a comparative analysis with two other centrality metrics, the traditional closeness and the information centrality. We applied the three metrics in five network topologies with distinct characteristics. Our results showed that, when considering multiple disjoint paths, the proposed metric can reclassify from 59% to 91% of nodes using a small connectivity factor, $\varphi = 1$. Furthermore, we observed that reclassification stops after a particular φ value, which is directly related to the network topology. Our metric is similar to the traditional closeness because both assign importance regarding node proximity. Despite this similarity, our metric reclassifies a large number of nodes, breaking ties and consequently highlighting nodes that are better connected. When node failures occur, the most central node classified by our metric remains reachable if a single failure occurs on a shortest-path node. Extending the analysis to a set of most central nodes, a smaller number of nodes lose access to at least one node of the set of most central nodes classified by our metric in most cases.

As future work, we intend to analyze more efficient algorithms to find multiple disjoint paths and propose a time-based metric to identify more accessible nodes. We also envision to automate the computation of the optimal φ value for each network topology considering the mean degree and network density. Finally, we intend to develop an application that uses the disjoint multipath closeness centrality for network function placement.

References

1. Amaral P, Pinto PF, Bernardo L, et al (2017) SDN based traffic engineering without optimization: a centrality based approach. In: Communications (ICC), 2017 IEEE international conference on, IEEE, p 1–7
2. Aouiz AA, Hacene SB, Lorenz P (2019) Channel busyness based multipath load balancing routing protocol for ad hoc networks. IEEE Netw 33(5):118–125
3. Barbosa MSM, Medeiros DSV, Campista MEM (2019) Centralidade de proximidade por múltiplos caminhos disjuntos: Aplicação em redes de longa distância. In: SBRC, p 88–101
4. Bavelas A (1948) A mathematical model for group structures. Human Organiz 7(3):16–30
5. Bavelas A (1950) Communication patterns in task-oriented groups. J Acoust Soc Am 22(6):725–730
6. Bentley TG, Cohen JT, Elkin EB et al (2017) Validity and reliability of value assessment frameworks for new cancer drugs. Val Health 20(2):200–205
7. Berahmand K, Bouyer A, Samadi N (2019) A new local and multidimensional ranking measure to detect spreaders in social networks. Computing 101(11):1711–1733
8. Bondy JA, Murty USR, et al (1976) Graph theory with applications, vol 290. Citeseer
9. Borgatti SP, Everett MG (2006) A graph-theoretic perspective on centrality. Social Net 28(4):466–484
10. Bouet M, Leguay J, Combe T et al (2015) Cost-based placement of VDPI functions in NFV infrastructures. Int J Netw Manag 25(6):490–506
11. Brandes U, Fleischer D (2005) Centrality measures based on current flow. In: STACS, p 533–544
12. Buckingham FM (1969) The harmonic mean in forest mensuration. The Forest Chron 45(2):104–106
13. Comitê Gestor RNP (2007) Rede ipê: Política de uso. In: Report, Rede Nacional de Pesquisa
14. Couto RS, Secci S, Campista MEM, et al (2014) Latency versus survivability in geo-distributed data center design. In: 2014 IEEE global communications conference, IEEE, p 1102–1107
15. Croux C, Dehon C (2010) Influence functions of the spearman and kendall correlation measures. Stat Meth Appl 19(4):497–515
16. DANTE (2007) Géant: Transforming the way researchers collaborate. In: Report, DANTE
17. Freeman LC (1978) Centrality in social networks conceptual clarification. Social Netw 1(3):215–239
18. Grossi R, Marino A, Versari L (2018) Efficient algorithms for listing k disjoint st-paths in graphs. In: Latin American Symposium on Theoretical Informatics, Springer, p 544–557
19. Hayes B (2006) Connecting the dots. Am Scient 94(5):400–404
20. Hu ZL, Liu JG, Yang GY et al (2014) Effects of the distance among multiple spreaders on the spreading. EPL (Europhy Lett) 106(1):18002
21. Kendall MG, Smith BB (1939) The problem of m rankings. Annals Math Stat 10(3):275–287
22. Maccari L, Cigno RL (2016) Pop-routing: Centrality-based tuning of control messages for faster route convergence. In: Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE international conference on, IEEE, p 1–9
23. Medeiros DSV, Campista MEM, Mitton N et al (2017) The power of quasi-shortest paths: ρ-geodesic betweenness centrality. IEEE Trans Netw Sci Eng 4(3):187–200. https://doi.org/10.1109/TNSE.2017.2708705
24. Nakarmi U, Rahnamay-Naeini M, Khamfroush H (2019) Critical component analysis in cascading failures for power grids using community structures in interaction graphs. IEEE Trans Netw Sci Eng. https://doi.org/10.1109/TNSE.2019.2904008
25. Newman MEJ (2006) Finding community structure in networks using the eigenvectors of matrices. Phys Rev E, 74:036104. https://doi.org/10.1103/PhysRevE.74.036104, https://link.aps.org/doi/10.1103/PhysRevE.74.036104
26. Newman MJ (2005) A measure of betweenness centrality based on random walks. Social Netw 27(1):39–54
27. Rossi RA, Ahmed NK (2015) The network data repository with interactive graph analytics and visualization. In: AAAI, http://networkrepository.com
28. Samadi N, Bouyer A (2019) Identifying influential spreaders based on edge ratio and neighborhood diversity measures in complex networks. Computing 101(8):1147–1175
29. Schafer V (2015) Part of a whole: RENATER, a twenty-year-old network within the internet. Inf & Cult 50(2):217–235
30. Stephenson K, Zelen M (1989) Rethinking centrality: methods and examples. Social Netw 11(1):1–37
31. Taylor WH (1933) The meaning of an average. J Educat Psychol 24(9):703