

Centralidade de Proximidade por Múltiplos Caminhos Disjuntos: Aplicação em Redes de Longa Distância

Mariana S. Maciel Barbosa¹, Dianne S. V. Medeiros²,
Miguel Elias M. Campista¹

¹GTA/PEE/COPPE – UFRJ – Universidade Federal do Rio de Janeiro

²MídiaCom/TET/PPGEET – UFF – Universidade Federal Fluminense

{maciel,miguel}@gta.ufrj.br, diannescherly@id.uff.br

Abstract. *Traditional centrality metrics consider only shortest paths, neglecting the existence of bit-longer paths between nodes in a network. In dynamic networks, however, these paths can be strategic due to resilience issues. This paper proposes a closeness centrality based on multiple disjoint paths, which extrapolates the traditional closeness, considering a fixed number of multiple shortest disjoint paths. The idea is to identify multiply-connected nodes able to perform tasks that require higher availability. We evaluate the proposed metric through comparisons with other centrality metrics. Our results confirm that the most central nodes according to the proposed metric are more accessible when failures happen in random nodes in the network. Moreover, our results show that the proposed metric is able to identify better-connected nodes and assign them more central importance.*

Resumo. *As métricas tradicionais de centralidade consideram apenas os caminhos mais curtos, ignorando a existência de caminhos um pouco mais longos entre pares de nós da rede. Em redes dinâmicas, porém, esses caminhos podem ser estratégicos por questões de resiliência a falhas. Assim, este artigo propõe a centralidade de proximidade por múltiplos caminhos disjuntos, que extrapola a proximidade tradicional, considerando um número fixo de múltiplos caminhos disjuntos mais curtos. A ideia é identificar nós multiplamente conectados que possam desempenhar tarefas que exijam maior disponibilidade. A métrica proposta é avaliada através da comparação com outras métricas de centralidade. Os resultados confirmam que os nós mais centrais da métrica proposta são mais acessíveis quando ocorrem falhas em nós aleatórios da rede. Além disso, mostram que a métrica é capaz de identificar nós mais bem conectados e atribuí-los uma maior importância central.*

1. Introdução

O emprego das métricas de centralidade herdadas da teoria de grafos permite quantificar a importância dos nós em redes de comunicação. Tais métricas podem revelar o papel desempenhado por cada nó a partir de informações topológicas ou, de forma mais abrangente, do relacionamento entre os próprios nós [Freeman 1978, Bavelas 1948]. Em redes de comunicação, identificar os nós mais importantes ou mais centrais é fundamental pois pode auxiliar a alocação de recursos ou serviços [Bouet et al. 2015],

ou até mesmo contribuir para a convergência da rede após a detecção de uma falha [Maccari and Cigno 2016]. Nessas redes, as métricas de centralidade são tipicamente usadas para identificar nós concentradores de tráfego ou nós com alto poder de disseminação da informação [Stephenson and Zelen 1989]. Nós com alto poder de disseminação possuem elevada centralidade de proximidade (*closeness*), que elege como nós mais centrais aqueles que na média estão mais próximos de todos os outros nós da rede [Freeman 1978], considerando apenas os caminhos mais curtos entre eles. Essa visão negligencia possíveis caminhos alternativos, que mesmo tendo tamanhos um pouco maiores, podem ser usados para a manutenção da conectividade da rede. Dessa forma, o uso apenas dos caminhos mais curtos pode não refletir o verdadeiro potencial dos nós, principalmente em cenários mais dinâmicos, nos quais falhas podem ocorrer com maior frequência [Medeiros et al. 2017a].

O uso de múltiplos caminhos além dos mais curtos para computar a centralidade de um nó já foi investigado na literatura [Stephenson and Zelen 1989, Newman 2005, Medeiros et al. 2017b], atribuindo maior importância aos caminhos mais curtos [Borgatti and Everett 2006]. A ideia é que, apesar de os fluxos seguirem por outros caminhos, eles tendem a se concentrar nos mais curtos, que oferecem menor resistência para transferi-los [Brandes and Fleischer 2005]. A centralidade de informação [Stephenson and Zelen 1989] e por percursos aleatórios são exemplos de métricas que consideram todos os caminhos entre qualquer par origem-destino da rede. Um problema comum a essas propostas é não considerar a aplicabilidade da métrica, não limitando o número máximo de caminhos.

Este artigo propõe a *centralidade de proximidade por caminhos disjuntos*, uma nova métrica de centralidade que extrapola o conceito da centralidade de proximidade tradicional. A métrica proposta considera a existência de múltiplos caminhos vértice-disjuntos, visando eleger como nó mais central aquele que for conectado a todos os outros nós da rede através de um número fixo de múltiplos caminhos disjuntos com o menor custo médio possível. Aplicada a redes de comunicação, a métrica aponta nós mais resilientes a falhas por privilegiar a maior diversidade de caminhos ao invés de apenas considerar o caminho mais curto. O limite máximo de caminhos disjuntos que podem ser considerados é definido pelo fator de conectividade φ . O uso de caminhos disjuntos e do fator de conectividade φ são importantes pois aumentam a aplicabilidade da métrica em redes reais, como as de longa distância mantidas por operadoras de telecomunicações. A avaliação da métrica proposta é feita através da comparação com os dois extremos das métricas de proximidade: a centralidade de proximidade tradicional [Bavelas 1950, Beauchamp 1965], que considera apenas caminhos mais curtos; e a centralidade de informação [Stephenson and Zelen 1989], que considera todos os caminhos existentes. A análise foca na influência do fator φ no ranqueamento dos nós e na correlação da entre as métricas. Os resultados mostram que a proximidade por caminhos disjuntos aponta nós que devem ser reclassificados por estarem mais bem conectados, mesmo mantendo elevada correlação com a proximidade tradicional. Os resultados mostram ainda que a métrica proposta é mais impactante caso a rede possua caminhos disjuntos e, acima de tudo, caso o φ selecionado consiga explorar essa característica. Por fim, observa-se que, mediante falhas na rede, a métrica proposta identifica os nós cujos caminhos até eles são menos afetados.

Este artigo está organizado da seguinte forma. A Seção 2 apresenta as notações e definições usadas no trabalho. A Seção 3 revisa as métricas de proximidade, enquanto a Seção 4 apresenta a métrica de proximidade por caminhos disjuntos proposta. A metodologia de avaliação e os resultados obtidos são discutidos nas Seções 5 e 6, respectivamente. Por fim, a Seção 7 conclui este trabalho e lista os trabalhos futuros.

2. Notações e Definições

Uma rede pode ser representada por um grafo $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W})$, em que \mathcal{V} denota o conjunto de nós, \mathcal{E} o de enlaces e \mathcal{W} o de custos dos enlaces. Um caminho $p_{s,t}$ entre os nós v_s e v_t é formado por uma sequência de vértices adjacentes distintos interligados por enlaces $\varepsilon_{i,j}$ com pesos $\omega_{i,j}$. O custo $\delta_{s,t}$ associado ao caminho $p_{s,t}$ é dado pela soma dos pesos $\omega_{i,j}$ dos enlaces que compõem o caminho. O caminho mais curto $p_{s,t}^*$ possui o menor custo total $\delta_{s,t}^*$. Dois nós podem estar interligados por n caminhos, representados por $p_{s,t}^{(n)}$, cada um com um custo $\delta_{s,t}^{(n)}$. Dois caminhos $p_{s,t}^{(x)}$ e $p_{s,t}^{(y)}$ entre v_s e v_t são ditos totalmente vértice-disjuntos se, e somente se, $p_{s,t}^{(x)} \cap p_{s,t}^{(y)} = \{v_s, v_t\}$. Essa definição continua válida mesmo se o número de nós intermediários ou se o custo dos caminhos forem diferentes.

Neste artigo, para determinar a centralidade de um nó são utilizados os caminhos mais curtos e múltiplos caminhos quase mais curtos [Medeiros et al. 2017b] vértice-disjuntos. Os múltiplos caminhos são limitados em quantidade, sendo que o número máximo adicional de caminhos que *podem* ser utilizados para alcançar um destino v_t é determinado pelo fator de conectividade φ , definido como segue.

Definição 1. Fator de conectividade (φ): indica o número *desejável* de caminhos totalmente disjuntos adicionais entre um par de nós v_s, v_t , com $\varphi \in \mathbb{N}$.

O fator de conectividade φ limita a busca por caminhos totalmente disjuntos entre um par origem-destino v_s, v_t , de forma que enquanto o número de caminhos totalmente disjuntos for menor que φ , busca-se um novo caminho. A busca se encerra quando φ é alcançado ou quando todos os caminhos disjuntos são encontrados, mesmo que o número total seja menor do que φ . Os caminhos são sempre selecionados em ordem crescente de custo. Vale notar que, caso haja caminhos com o mesmo custo, qualquer um pode ser escolhido. A Figura 1 ilustra a ideia por trás da definição do fator de conectividade φ . Quanto mais caminhos entre os mesmos pares origem-destino existirem, mais bem conectados estarão mesmo se os custos dos caminhos forem diferentes. Logo, caso haja uma falha em um nó, v_f , a rede se mantém conectada com maior probabilidade. O nó v_s , por exemplo, continuaria podendo alcançar v_t através dos caminhos quase mais curtos, com 2 e 3 saltos. Quanto mais bem conectados são os nós, maior é a resiliência da rede. Essa percepção justifica a métrica de centralidade proposta neste artigo, a qual atribui maior importância aos nós com mais caminhos totalmente disjuntos até todos os outros nós da rede, priorizando os menores caminhos.

3. Métricas de Centralidade

As métricas de centralidade são utilizadas frequentemente para analisar a relação entre os nós em uma rede, permitindo fundamentar as decisões tomadas sobre os papéis a serem desempenhados por eles. Os principais tipos de centralidade são o grau, a intermediação, a centralidade de autovetor, a proximidade, e as variantes dessas



Figura 1. O nó v_s se torna mais bem conectado à medida em que mais caminhos disjuntos existem para alcançar o destino (v_t), mesmo que os caminhos sejam quase mais curtos. O aumento de φ contribui para a melhora da conectividade de v_s , aumentando a resiliência da comunicação com o destino.

métricas [Borgatti and Everett 2006]. O grau tem relação com a quantidade de redes e sub-redes que um nó se conecta diretamente, e a intermediação, com o controle de um nó sobre os fluxos entre outros nós da rede. A centralidade de autovetor se relaciona com a influência que um nó exerce sobre os outros nós da rede, de forma que a importância do nó depende da importância dos seus vizinhos. Por fim, a proximidade elege como mais central o nó que está mais próximo, na média, de todos os outros nós. Como consequência, esse nó pode acessar ou ser acessado mais rapidamente por todos os outros nós da rede.

Este artigo investiga a acessibilidade dos nós em uma rede, considerando que os nós mais importantes são os múltiplos conectados e que são alcançados, ou alcançam, mais rapidamente os demais nós. Esses nós podem desempenhar papéis importantes, como o de um servidor de conteúdo. A instalação desse servidor no nó mais central segundo o conceito de proximidade reduz, potencialmente, a latência média para acessá-lo. Assim, este artigo foca na proximidade por ser a métrica mais adequada para a análise de acessibilidade. A *centralidade de proximidade tradicional* (C_{trad}) considera apenas os caminhos mais curtos para determinar a importância de um nó v_s , da seguinte forma: $C_{trad}(v_s) = \frac{|V|-1}{\sum_{t=1}^{|V|} \delta_{s,t}^*}$. A utilização apenas dos caminhos mais curtos, porém, pode ser vista como uma desvantagem da abordagem tradicional, uma vez que um fluxo enviado por um nó pode percorrer caminhos que não são tão diretos quanto os mais curtos. Isso pode ocorrer devido ao processo de comunicação na rede ser aleatório, permitindo que os fluxos percorram quaisquer caminhos aleatoriamente, ou devido à canalização intencional através de nós intermediários específicos [Stephenson and Zelen 1989, Newman 2005]. Assim, os outros caminhos existentes entre os pares de nós também devem ser contabilizados para calcular a centralidade. Nesse sentido, Stephenson e Zelen propõem a *centralidade de informação* (I_s) [Stephenson and Zelen 1989], definida formalmente como $I_s(v_s) = \frac{|V|}{\sum_{t=1}^{|V|} \frac{1}{I_{s,t}}}$, onde $I_{s,t} = \sum_n 1/\delta_{s,t}^{(n)}$, define uma medida da informação presente nos caminhos existentes (n) entre v_s e v_t .

Este artigo também considera múltiplos caminhos e prioriza os caminhos mais curtos, para propor uma nova métrica de centralidade de proximidade. Diferente dos trabalhos mencionados, a métrica proposta considera que existe um limite máximo para o número de caminhos considerados e que estes devem ser obrigatoriamente disjuntos entre si. Considerando uma abordagem mais restritiva, neste artigo são usados apenas

caminhos vértice-disjuntos, o que é justificável no projeto de redes tolerantes a falhas.

4. Centralidade de Proximidade por Caminhos Disjuntos

Este artigo propõe a *centralidade de proximidade por caminhos disjuntos*, cujo objetivo é identificar o potencial de conectividade de um nó quando múltiplos caminhos disjuntos são considerados. A quantidade de caminhos é limitada pelo fator de conectividade φ (Seção 2) e são considerados apenas os caminhos totalmente vértice-disjuntos. A utilização desses caminhos diferencia a métrica proposta tanto da proximidade tradicional quanto da centralidade de informação, destacando nós que possuem maior diversidade de menores caminhos totalmente disjuntos para se comunicarem com os demais nós da rede. Mesmo em caso de falha de outros nós, nós multiplamente conectados tendem a ser mais acessíveis e têm maior probabilidade de permanecerem acessíveis se os múltiplos caminhos forem disjuntos entre si. Além disso, quanto menores forem os caminhos disjuntos, menor tende a ser o custo adicional para chegar ao destino quando a falha ocorre. Logo, nós multiplamente conectados através de caminhos disjuntos curtos são boas opções para instalação de funcionalidades críticas em redes que se preocupam com a disponibilidade.

4.1. Formalização

Neste artigo, um nó é tão mais importante quanto mais rápido ele conseguir alcançar os outros nós, ou vice-versa, considerando ainda o quão bem conectado o nó está com cada destino. Um nó é bem conectado ao destino quando existem múltiplos caminhos totalmente vértice-disjuntos entre eles. As contribuições dos múltiplos caminhos disjuntos entre v_s e v_t são inseridas na métrica através do cálculo do custo agregado $\Delta_{s,t}^{(\varphi)}$, considerando ser desejável encontrar φ caminhos disjuntos. Esse custo tem relação com a resistência média sofrida por um fluxo para ser levado de uma origem até um destino utilizando múltiplos caminhos de resistências distintas. Quanto maior for o custo de um caminho, mais resistente ele é ao transporte do fluxo e, portanto, menor deve ser a contribuição desse caminho para a vazão total do fluxo. Dessa forma, o custo do caminho e a sua contribuição para a importância de um nó são inversamente proporcionais. Devido a essa relação inversa, o custo agregado é calculado como uma média harmônica entre os custos dos φ caminhos disjuntos entre v_s e v_t , conforme Equação 1. Considerando as contribuições dos caminhos entre o nó de interesse, v_s , e os demais nós, obtém-se a *centralidade de proximidade por caminhos disjuntos*, conforme Equação 2.

$$\Delta_{s,t}^{(\varphi)} = \frac{1}{\sum_{n=0}^{\varphi} \frac{1}{\delta_{s,t}^{(n)}}}, \quad (1) \quad C_{\varphi}(v_s) = \frac{|\mathcal{V}| - 1}{\sum_{t=1}^{|\mathcal{V}|} \Delta_{s,t}^{(\varphi)}}, \quad (2)$$

onde $|\mathcal{V}|$ é o número de nós na rede e $|\mathcal{V}| - 1$ é um fator de normalização que permite comparar a centralidade de nós em redes com dimensões diferentes. Ressalta-se que para $\varphi = 0$, nenhum caminho disjunto adicional é contabilizado, e a métrica proposta torna-se igual à centralidade de proximidade tradicional.

5. Metodologia de Avaliação

A métrica proposta é avaliada através de uma análise comparativa com a centralidade de proximidade tradicional e a centralidade de informação. O objetivo é observar a correlação entre as métricas e mostrar como a reclassificação dos nós, provocada pela

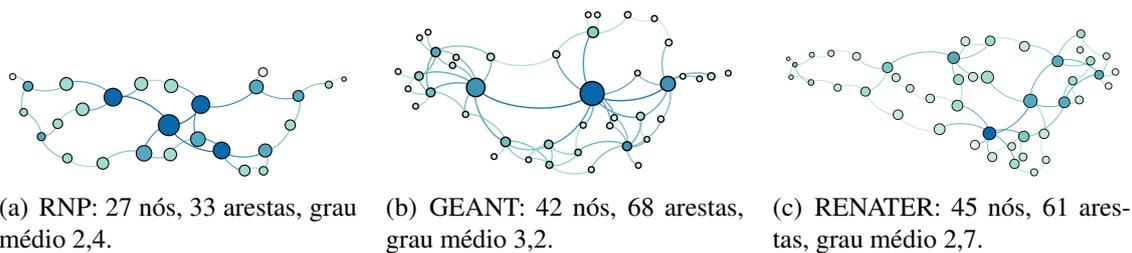


Figura 2. Topologias das redes (a) RNP, (b) GEANT e (c) RENATER. Quanto maior o nó, maior seu grau e quanto mais escuro o nó, maior sua proximidade.

utilização da métrica proposta, pode beneficiar a comunicação na rede, principalmente em caso de falhas de nós. Para tanto, neste artigo considera-se apenas grafos simples, não direcionados, sem laços e não ponderados, com o número de saltos como métrica de custo. Isso não limita a aplicação da métrica a grafos com custos de enlace unitários. São utilizados três conjuntos de dados que representam topologias de redes de longa distância reais de educação e pesquisa: RNP, localizada no Brasil; GEANT, na Europa; e RENATER, na França [Couto et al. 2014], ilustradas na Figura 2.

Considera-se, neste artigo, a maior quantidade possível de caminhos totalmente vértice-disjuntos, respeitando o limite delimitado pelo fator de conectividade φ . A busca por múltiplos caminhos disjuntos pode seguir abordagens distintas. Uma delas tem como objetivo maximizar a quantidade de caminhos disjuntos, desconsiderando se os caminhos mais curtos estão nesse conjunto [Suurballe 1974]. Na outra, a busca por caminhos disjuntos considera a presença do caminho mais curto no conjunto encontrado [Sidhu et al. 1991]. Neste artigo, utiliza-se uma abordagem semelhante à segunda, visando maximizar o número de caminhos totalmente disjuntos de mesmo custo, considerando que o caminho mais curto está obrigatoriamente no conjunto encontrado. Considerando o pior caso, para $\varphi \geq 1$ o algoritmo que busca os caminhos disjuntos tem complexidade $O(n^2)$. Aplicando a busca a cada par de nós, no pior caso o algoritmo que computa a métrica proposta alcança uma complexidade de $O(n^4)$. A seleção dos caminhos segue o algoritmo a seguir.

1. Define-se o parâmetro φ , que determina a quantidade total de caminhos considerados entre cada par de nós v_s e v_t ;
2. Inicia-se a busca pelos caminhos mais curtos entre v_s e v_t ;
3. Compara-se os caminhos encontrados, ordenando-os por grau de disjunção. Desta forma o caminho mais disjunto ao demais é melhor classificado e descarta-se os que não são disjuntos a ele;
4. Se a quantidade de caminhos mais curtos encontrados for maior que φ , apenas φ caminhos são computados;
5. Se a quantidade de caminhos mais curtos computados entre v_s e v_t for menor que φ , os nós intermediários pertencentes aos caminhos já computados são desconsiderados no grafo e retorna-se ao passo 2.
6. A busca por caminhos disjuntos é finalizada se φ caminhos forem encontrados ou se não existirem mais caminhos disjuntos.

A Figura 3 exemplifica a busca pelos caminhos disjuntos no caso em que se deseja utilizar $\varphi = 3$ caminhos disjuntos adicionais entre os pares de nós em uma de-



(a) Total de caminhos considerados menor do que φ . (b) Total de caminhos considerados igual a φ .

Figura 3. O número de caminhos computados nem sempre atinge φ . É possível que (a) nenhum caminho quase mais curto possa ser computado por não existir um disjunto ao mais curto. (b) Caso existam caminhos quase mais curtos disjuntos ao mais curto e entre si, eles são considerados no cálculo da métrica.

terminada rede. Na Figura 3(a) existem 3 caminhos mais curtos entre v_1 e v_8 , com $\delta_{1,8}^* = 3$. Como apenas os caminhos disjuntos interessam, considera-se apenas os caminhos $p_{1,8}^{(0)} = \langle v_1, v_5, v_9, v_8 \rangle$ (vermelho) e $p_{1,8}^{(1)} = \langle v_1, v_0, v_4, v_8 \rangle$ (verde). Não existem caminhos disjuntos aos mais curtos selecionados e apenas $p_{1,8}^{(0)}$ e $p_{1,8}^{(1)}$ são considerados, mesmo com $\varphi = 3$. Diferentemente, entre os nós v_0 e v_3 o número de caminhos disjuntos desejáveis é alcançado. A Figura 3(b) realça o caminho mais curto ($p_{0,3}^{(0)}$) entre v_0 e v_3 , um caminho quase mais curto de 2 saltos ($p_{0,3}^{(1)}$) e dois caminhos quase mais curtos de 3 saltos ($p_{0,3}^{(2)}$, verde, e $p_{0,3}^{(3)}$, amarelo). Os caminhos $p_{0,3}^{(2)}$ e $p_{0,3}^{(3)}$ não são disjuntos entre si, de forma que apenas um deles pode ser considerado. Caso $p_{0,3}^{(2)} = \langle v_0, v_2, v_6, v_3 \rangle$, seja considerado, existirá mais um único caminho disjunto a ele, com 5 saltos. O mesmo ocorre caso $p_{0,3}^{(3)} = \langle v_0, v_1, v_6, v_3 \rangle$ seja considerado. Assim, seleciona-se $p_{0,3}^{(2)}$ ou $p_{0,3}^{(3)}$ e mais o respectivo caminho de 5 saltos.

6. Resultados

A avaliação da métrica proposta é feita em três etapas. Investiga-se primeiramente a (i) influência do aumento do fator de conectividade φ na classificação dos nós. Em seguida, (ii) utiliza-se o coeficiente de correlação W de Kendall para verificar o grau de concordância entre os ranqueamentos produzidos pelas métricas comparadas: proposta, tradicional e centralidade de informação. Por fim, avalia-se a (iii) variação da alcançabilidade dos nós quando ocorrem falhas em nós aleatórios da rede e em nós pertencentes aos caminhos mais curtos.

6.1. Influência do fator de conectividade no ranqueamento

O ranqueamento dos nós pela proximidade de caminhos disjuntos depende do valor do fator de conectividade φ usado, uma vez que φ está diretamente relacionado ao número de caminhos computados na métrica. Logo, é importante investigar como a variação na classificação dos nós é influenciada pelo fator de conectividade. A Figura 4 mostra o número de caminhos total entre todos os nós da rede considerados para computar a métrica, com $\varphi \in [0, 4]$. Note que para $\varphi = 0$ não há busca por caminhos disjuntos e o resultado é o mesmo encontrado para a proximidade tradicional. Observa-se que a maior quantidade de caminhos ocorre para os maiores valores de φ e que o incremento no número de caminhos é cada vez menor à medida que φ cresce. Esse incremento chega a

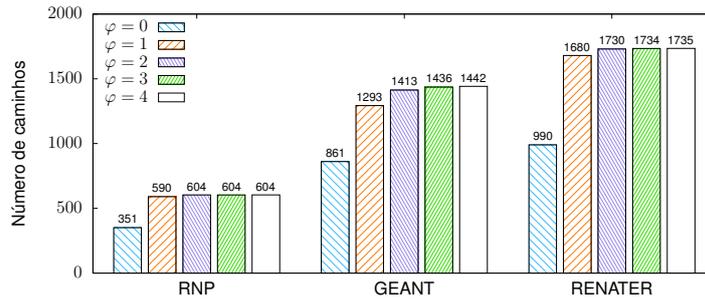


Figura 4. Total de caminhos entre todos os nós da rede usados para computar a métrica proposta. O aumento do fator de conectividade φ resulta no aumento do número total de caminhos considerados, mas com crescimento cada vez menor à medida em que φ aumenta.

ser nulo para $\varphi > 2$ na rede da RNP, que possui o menor número de caminhos alternativos disjuntos entre quaisquer pares de nós, quando comparada às outras redes.

O crescimento no número de caminhos devido ao aumento do fator de conectividade φ pode resultar na reclassificação dos nós nas redes analisada, aumentando ou diminuindo a posição do nó no ranqueamento. Isso ocorre porque o valor da centralidade do nó pode mudar de acordo com o número de caminhos considerados. A Figura 5 apresenta a frequência de variação na reclassificação dos nós devido aos caminhos adicionais contabilizados. Essa métrica é obtida a partir da relação entre o número de nós que trocam de posição ao mudar o valor de φ e o número de nós existentes. Observa-se que a maior quantidade de reclassificações ocorre para a rede RENATER, com 91% dos nós reclassificados, quando se considera a mudança do fator de conectividade de $\varphi = 0$ para $\varphi = 1$. Os nós que ocupam as primeiras posições tendem a coincidir nas métricas de centralidade [Freeman 1978]. Apesar disso, dentre as reclassificações na rede RENATER, a primeira posição, ocupada pelo nó 24 na proximidade tradicional ($\varphi = 0$), passa a ser ocupada pelo nó 37 na métrica proposta. Destaca-se também, que o nó 44 perde 17 posições. Essa perda está diretamente relacionada ao fato desse nó não possuir caminhos disjuntos par os demais nós da rede. Apesar de a rede RNP apresentar a menor frequência de reclassificações quando se começa a considerar os caminhos disjuntos ($\varphi = 1$), essa frequência ainda assim é alta, alcançando aproximadamente 59% dos nós. O mais notório destas reclassificações é o desempate de nós que ocupam a mesma posição na proximidade tradicional. Por exemplo, os nós 22 e 10 são classificados no segundo lugar, mas o nó 10 perde uma posição no ranqueamento obtido para a métrica proposta. Ambas as redes RENATER e RNP atingem estabilidade no ranqueamento dos nós a partir de $\varphi = 2$. Ou seja, não há mais variação na classificação dos nós com o aumento de φ , apesar de haver um aumento no número de caminhos considerados para computar a métrica, como ilustrado na Figura 4. Observa-se, ainda, que a rede GEANT é a que possui o decaimento mais lento na frequência de reclassificações. Isso ocorre porque a proporção entre número de caminhos e número de nós é a maior dentre as topologias estudadas, como pode ser visto na Figura 2.

6.2. Correlação entre as métricas

A fim de investigar a correlação entre a centralidade de proximidade por caminhos disjuntos e as outras métricas de proximidade, utiliza-se o coeficiente de correlação

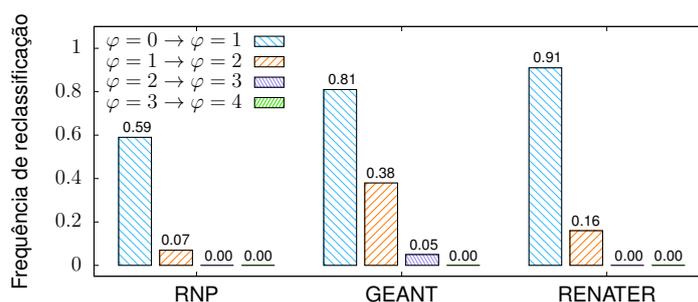


Figura 5. Frequência de reclassificação dos nós de acordo com a variação do fator de conectividade φ . A rede RENATER apresenta a maior frequência de reclassificações quando se começa a considerar os caminhos disjuntos $\varphi = 1$, enquanto a rede RNP apresenta a menor frequência. Ambas estabilizam o ranqueamento a partir de $\varphi = 2$, enquanto a rede GEANT ainda apresenta nós reclassificados para $\varphi = 3$.

W de Kendall. Assim, verifica-se não só a influência de φ na classificação dos nós, como também a influência do uso dos caminhos disjuntos na diferenciação entre os ranqueamentos produzidos pelas métricas. O coeficiente W de Kendall é uma medida de correlação robusta e eficiente quando comparada a outras medidas de correlação, que apresenta bons resultados para correlacionar ranqueamentos [Croux and Dehon 2010]. Através desse coeficiente é possível identificar o grau de concordância entre os ranqueamentos produzidos pelas métricas. A Figura 6 mostra o valor do coeficiente W de Kendall para as métricas de proximidade avaliadas neste artigo. Quanto mais próxima da borda do gráfico de radar estiver a área preenchida, maior é o grau de concordância entre as métricas comparadas. Observa-se na Figura 6, que a centralidade de informação possui a menor correlação com as demais métricas. Isso ocorre porque essa métrica utiliza todos os caminhos existentes entre os pares de nós para determinar o valor da centralidade. É interessante notar na Figura 6(a) que a rede RNP apresenta os maiores valores de correlação entre a proximidade tradicional e a métrica proposta. Além disso, o aumento do fator de conectividade não implica em mudanças significativas na correlação entre as métricas, o que é consequência de a rede RNP possuir poucos caminhos disjuntos. Na Figura 6(b), da rede GEANT, observa-se que a correlação entre a proximidade por caminhos disjuntos e a tradicional é menor do que na rede RNP, porém a correlação ainda é elevada. Também se observa que valores maiores de φ resultam em uma leve redução no grau de concordância com a proximidade tradicional.

Apesar de o grau de concordância entre as métricas de proximidade tradicional e por caminhos disjuntos ser elevado, inúmeros nós são reclassificados ao se utilizar a métrica proposta. A Figura 7 reforça o estudo da correlação entre as métricas, mostrando no eixo- y o número de posições que os nós ganham ou perdem ao se utilizar cada métrica e, através da escala de cor, o percentual de nós que sofreram uma mudança de y posições ao se trocar a métrica classificatória. O eixo- x dessa figura corresponde à transição de uma métrica para outra. Dessa forma, valores positivos em y indicam que os nós foram melhor classificados na transição de uma métrica para a outra e valores negativos indicam o oposto. A Figura 7(a) mostra a variação de posições no ranqueamento para a rede RNP. Observa-se que ao se utilizar a proximidade por caminhos disjuntos com $\varphi = 1$ em vez da tradicional, alguns nós ganham até 4 posições. Ao se aumentar φ de 1 para 2, poucos nós

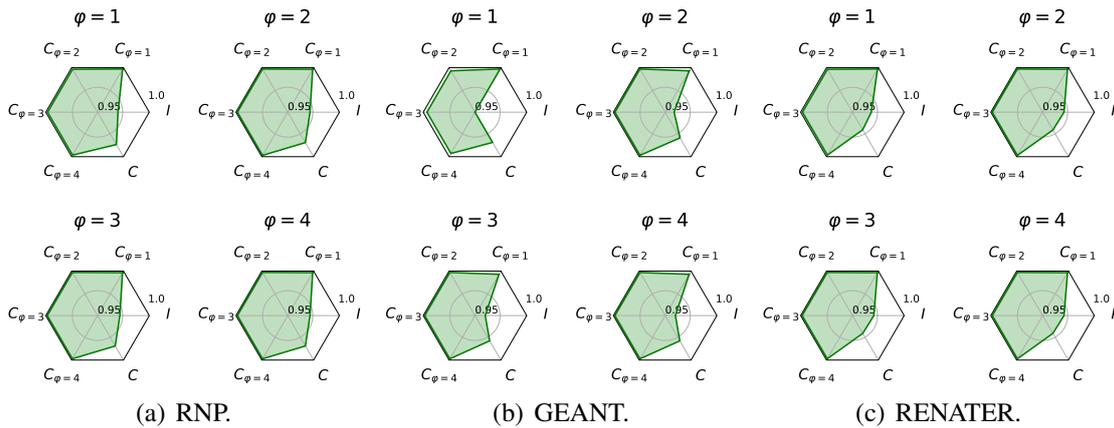


Figura 6. O coeficiente de correlação W de Kendall mostra o grau de concordância entre os ranqueamentos produzidos pelas métricas. A centralidade de informação possui a menor correlação com a métrica proposta em todas as topologias analisadas. A proximidade tradicional e a proposta apresentam elevada correlação entre si. Essa correlação diminui levemente à medida em que o fator de conectividade φ aumenta, em redes com maior quantidade de caminhos disjuntos como RENATER e GEANT.

mudam de posição e o número de posições variado é pequeno. O ranqueamento estabiliza ao se utilizar $\varphi = 2$, de forma que o aumento no valor desse parâmetro não provoca mais mudanças de posição. Isso ocorre porque não existem mais caminhos disjuntos a serem considerados, conforme observado na Figura 4.

Nas Figuras 7(b) e 7(c), que apresentam a variação de posições para as redes GEANT e RENATER, respectivamente, a variação entre a proximidade tradicional e a por caminhos disjuntos com $\varphi = 1$ é maior e menos nós permanecem nas mesmas posições, podendo alguns nós ganharem até 6 posições na rede GEANT e 17 posições na rede RENATER. Novamente, ao se aumentar φ , cada vez menos nós são reclassificados, ocorrendo a estabilização do ranqueamento em $\varphi = 3$ na rede GEANT e em $\varphi = 2$ na rede RENATER. Esses resultados corroboram o que se observa nas Figuras 4 e 5. Em relação aos ranqueamentos produzidos pela proximidade por caminhos disjuntos e pela centralidade de informação, observa-se que na rede RNP, Figura 7(a), aproximadamente 50% dos nós se mantêm na mesma posição quando deixam de ser classificados pela centralidade de informação e passam a ser classificados pela métrica proposta, independente do valor de φ usado. Como para $\varphi > 2$ não são considerados novos caminhos disjuntos, não existe variação de posições entre o ranqueamento produzido pela centralidade de informação e o gerado pela métrica proposta tanto para $\varphi = 2$ ($I \rightarrow \varphi = 2$) quanto para $\varphi = 3$ ($I \rightarrow \varphi = 3$). Esse mesmo comportamento é observado na rede RENATER, na Figura 7(c). Na rede GEANT, Figura 7(b), a amplitude da variação de posições é reduzida quando a centralidade de informação deixa de ser utilizada e se passa a utilizar a métrica proposta com $\varphi = 2$ ($I \rightarrow \varphi = 2$) em vez de $\varphi = 1$ ($I \rightarrow \varphi = 1$). Aumentando o valor de φ para 3 ($I \rightarrow \varphi = 3$), a fração dos nós que ganhavam 11 posições para $\varphi = 2$ ($I \rightarrow \varphi = 2$) passa a ganhar apenas 10. Para valores maiores de φ não são encontrados novos caminhos disjuntos, de forma que a variação no ranqueamento se estabiliza. Caso todos os caminhos existentes entre os nós fossem considerados com o aumento do φ , o ranqueamento produzido pela métrica proposta tenderia ao mesmo ranqueamento gerado

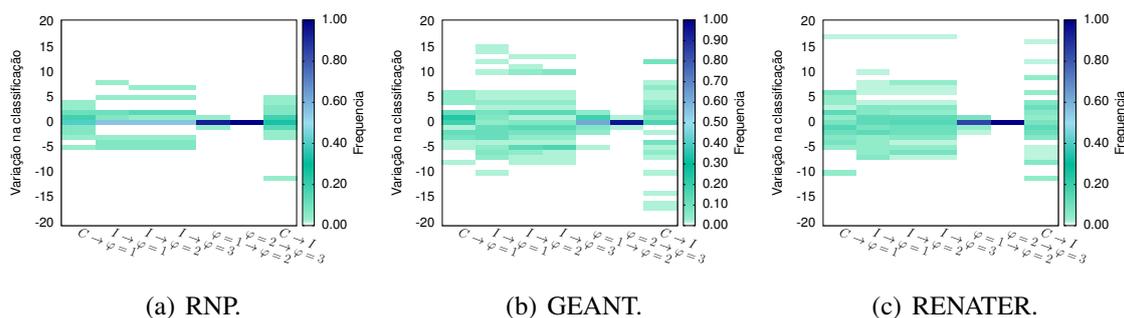


Figura 7. A variação da posição dos nós no ranqueamento depende do número de caminhos usados para computar as métricas. A centralidade de informação e o uso de caminhos disjuntos provocam modificações mais intensas na posição dos nós. À medida em que o fator de conectividade aumenta o ranqueamento dos nós tende a se estabilizar.

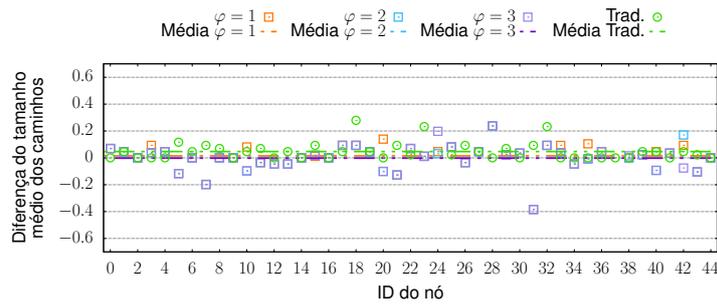
pela centralidade de informação.

Os resultados da Figura 7 mostram que em todas as topologias de rede utilizadas, a classificação dos nós segundo a centralidade de informação, quando comparada à classificação gerada pela proximidade tradicional produz a maior amplitude de reclassificação. Na rede GEANT, Figura 7(b), essa amplitude é maior, chegando a 29, com aproximadamente 2% dos nós perdendo 17 posições e aproximadamente 10% ganhando 12 posições. Essa maior intensidade na reclassificação ao se utilizar a centralidade de informação, quando comparada à proximidade tradicional, é esperada por ser a métrica que mais difere em relação à tradicional, devido ao uso de todos os caminhos pela centralidade de informação para computar a centralidade do nó.

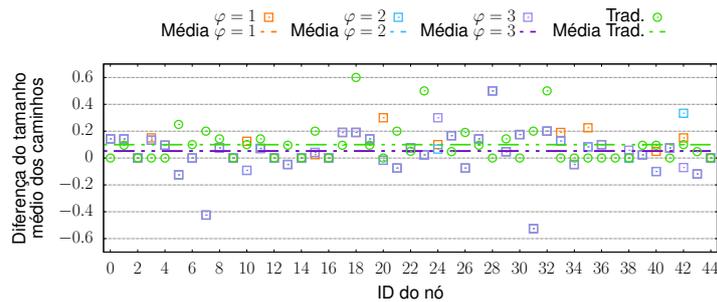
6.3. Alcançabilidade dos nós em presença de falha

Apesar de apresentar elevada correlação com a proximidade tradicional, a métrica proposta visa destacar os nós com maior alcançabilidade na rede quando múltiplos caminhos disjuntos são utilizados. Uma vez comprovado pelos resultados anteriores que a métrica proposta de fato aponta diversos nós que podem ser reclassificados, investiga-se o potencial do uso dos caminhos disjuntos na classificação dos nós em redes propensas a falhas. Para tanto, são feitas duas análises. Na primeira, investiga-se o custo para que os nós alcancem uns aos outros quando ocorrem falhas aleatórias na rede. Já na segunda, investiga-se o custo para alcançar outros nós da rede quando a falha ocorre em um nó qualquer que pertença a um caminho mais curto da rede.

A alcançabilidade entre os nós da rede na presença de falha aleatória é investigada falhando-se cada um dos nós existentes. Calcula-se então, para cada nó, o tamanho médio dos caminhos para alcançar o nó mais central segundo a métrica. Calcula-se também o tamanho médio dos caminhos quando a rede está em pleno funcionamento. Em seguida, verifica-se a diferença entre os tamanhos médios dos caminhos antes e depois das falhas. A Figura 8 mostra o resultado obtido para a rede RENATER. O eixo- x identifica cada nó da rede, enquanto as linhas horizontais apresentam os valores médios para todos os nós antes e depois da falha. Note que, no caso da proximidade tradicional, os caminhos usados são apenas os mais curtos, enquanto para a proximidade por caminhos disjuntos considera-se todos os caminhos disjuntos que atendam o valor determinado para o fator



(a) Falha em nós aleatórios.



(b) Falha em nós de um caminho mais curto.

Figura 8. Após uma falha, todos os caminhos usados pela proximidade tradicional e pela métrica proposta para alcançar seu respectivo nó mais central são considerados. O custo médio é obtido calculando a média das diferenças para todos os nós. (a) Após a falha aleatória, o custo médio para alcançar o nó mais central segundo a métrica proposta usando $\varphi = \{2, 3\}$ é igual a $-0,0023$, enquanto para a proximidade tradicional é igual a $0,047$. (b) Após a falha dos nós pertencentes a um caminho mais curto, para alcançar o nó mais central de métrica proposta o custo médio é de $0,051$ considerando $\varphi = \{2, 3\}$. Já para a proximidade tradicional o custo médio é de $0,099$.

de conectividade φ . Nessa análise, valores positivos da diferença dos tamanhos médios indicam que existe um aumento no custo médio para acessar o nó mais central da rede, de forma que quanto maior for o valor, maior é o aumento no custo médio. Por outro lado, valores negativos são encontrados quando o tamanho médio dos caminhos após a falha é menor do que antes da falha. Isso ocorre quando a falha aleatória provoca a perda de um caminho quase mais curto que anteriormente era utilizado. Assim, quanto mais próximo de $y = 0$, menor é o aumento no custo médio para acessar o nó mais central. Nas redes RNP e GEANT os mesmos nós são identificados como mais centrais pela proximidade tradicional e pela métrica proposta, de forma que os resultados sobre alcançabilidade entre os nós da rede na presença de falha obtidos para essas redes não acrescenta informação.

A Figura 8 mostra que após a ocorrência de uma falha aleatória na rede, o nó classificado como mais central pela métrica proposta consegue oferecer um menor aumento no custo médio dos caminhos usados para acessá-lo quando comparado ao uso dos caminhos mais curtos apenas. O cálculo do custo médio é a média da diferença para todos os nós. Para a métrica proposta obtém-se um valor de $-0,0023$ usando $\varphi = \{2, 3\}$, e para a proximidade tradicional o custo médio é de $0,047$. Quando a falha ocorre apenas nos nós que pertencem a caminhos mais curtos, a média da diferença para cada nó mostra que o custo médio para a métrica proposta é de $0,051$, também usando $\varphi = \{2, 3\}$, e para a pro-

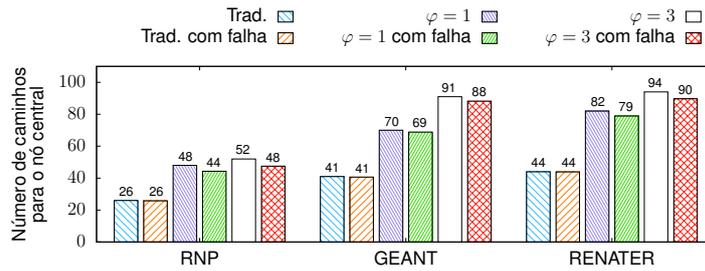


Figura 9. Número de caminhos conhecidos alternativos entre todos os nós da rede e o nó mais central conforme identificado pela métrica proposta com $\varphi = \{1, 3\}$ e pela proximidade tradicional.

ximidade tradicional é de 0,099. Apesar dos valores pequenos, decorrente das pequenas topologias avaliadas, o custo médio para a métrica proposta fica mais próximo de 0 nos dois casos avaliados. Além disso, caso ocorram falhas em nós dos caminhos mais curtos, 4 nós da rede RENATER não conseguem mais se comunicar com o nó mais central identificado pela proximidade tradicional, utilizando os caminhos considerados pela métrica. Por outro lado, apenas 3 nós não conseguem se comunicar com o nó mais central identificado pela métrica proposta na presença desse tipo de falha, ao se utilizar os caminhos considerados pela métrica. A Figura 9 mostra o número de caminhos conhecidos entre todos os nós da rede e o nó mais central identificado pela métrica proposta com $\varphi = \{1, 3\}$ e pela proximidade tradicional. Percebe-se que tanto antes quanto depois da falha aleatória, o número de caminhos considerados até o nó central é maior para a métrica proposta. Os resultados para a centralidade de informação não são plotados, porque a métrica é muito afetada pela redução do número de caminhos.

7. Conclusão

Este artigo propôs uma métrica de centralidade que constrói sobre ideias de proximidade e de múltiplos caminhos. Inspirada em redes reais, a métrica se propõe a responder perguntas sobre qual nó da rede está mais apto a executar um papel central com menor risco de indisponibilidade. Para isso, a métrica calcula qual é o nó mais central da rede considerando o custo médio de um número fixo de caminhos totalmente disjuntos, cuja quantidade é limitada pelo fator de conectividade φ . Os resultados mostraram que, comparada à métrica de proximidade (*closeness*), que considera apenas o caminho mais curto para cálculo, e a de informação, que considera todos os caminhos existentes, a métrica proposta consegue identificar nós mais bem conectados. A análise mostra ainda que o impacto da reclassificação dos nós em relação às outras métricas depende do número de caminhos alternativos nas topologias e do valor de φ . Por fim, os resultados mostraram que, na presença de falhas na rede, a métrica consegue identificar os nós cujos caminhos até eles são menos afetados. Como trabalhos futuros pretende-se avaliar o comportamento da métrica proposta em redes de maior porte, que tenham um maior número de caminhos disjuntos. Também será investigado o que ocorre no grupo de nós mais bem classificados, na presença de falhas única e múltiplas.

Agradecimento

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES), Código de Financiamento 001; e da

Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processos nº 15/24494-8 e 15/24490-2.”

Referências

- Bavelas, A. (1948). A mathematical model for group structures. *Human Organization*, 7(3):16–30.
- Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, 22(6):725–730.
- Beauchamp, M. A. (1965). An improved index of centrality. *Behavioral science*, 10(2):161–163.
- Borgatti, S. P. e Everett, M. G. (2006). A graph-theoretic perspective on centrality. *Social Networks*, 28(4):466 – 484.
- Bouet, M., Leguay, J., Combe, T., e Conan, V. (2015). Cost-based placement of vdpi functions in nfv infrastructures. *International Journal of Network Management*, 25(6):490–506.
- Brandes, U. e Fleischer, D. (2005). Centrality measures based on current flow. Em *STACS*, pp. 533–544.
- Couto, R. S., Secci, S., Campista, M. E. M., e Costa, L. H. M. (2014). Latência versus sobrevivência no projeto de centros de dados geograficamente distribuídos. *XXXII SBRC*, pp. 809–822.
- Croux, C. e Dehon, C. (2010). Influence functions of the spearman and kendall correlation measures. *Statistical methods & applications*, 19(4):497–515.
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239.
- Maccari, L. e Cigno, R. L. (2016). Pop-routing: Centrality-based tuning of control messages for faster route convergence. Em *Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on*, pp. 1–9. IEEE.
- Medeiros, D. S. V., Campista, M. E. M., Marcelo Dias de Amorim, N. M., e Pujolle, G. (2017a). Eficiência dos caminhos quase mais curtos em redes dinâmicas. Em *SBRC*, pp. 544–557.
- Medeiros, D. S. V., Campista, M. E. M., Mitton, N., Amorim, M. D., e Pujolle, G. (2017b). The power of quasi-shortest paths: ρ -geodesic betweenness centrality. *IEEE Transactions on Network Science and Engineering*, 4(3):187–200.
- Newman, M. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54.
- Sidhu, D., Nair, R., e Abdallah, S. (1991). Finding disjoint paths in networks. *ACM SIGCOMM Computer Communication Review*, 21(4):43–51.
- Stephenson, K. e Zelen, M. (1989). Rethinking centrality: Methods and examples. *Social Networks*, 11(1):1–37.
- Suurballe, J. (1974). Disjoint paths in a network. *Networks*, 4(2):125–145.