



CENTRALIDADE DE PROXIMIDADE POR MÚLTIPLOS CAMINHOS DISJUNTOS

Mariana de Souza Maciel Barbosa

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientadores: Miguel Elias Mitre Campista
Dianne Scherly Varela de
Medeiros

Rio de Janeiro
Julho de 2019

CENTRALIDADE DE PROXIMIDADE POR MÚLTIPLOS
CAMINHOS DISJUNTOS

Mariana de Souza Maciel Barbosa

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Miguel Elias Mitre Campista, D.Sc.

Prof^a. Dianne Scherly Varela de Medeiros, D.Sc.

Prof. Marcelo Gonçalves Rubinstein, D.Sc.

Prof. Rodrigo de Souza Couto, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
JULHO DE 2019

Barbosa, Mariana de Souza Maciel

Centralidade de Proximidade por Múltiplos
Caminhos Disjuntos/Mariana de Souza Maciel Barbosa. –
Rio de Janeiro: UFRJ/COPPE, 2019.

XIII, 52 p.: il.; 29, 7cm.

Orientadores: Miguel Elias Mitre Campista

Dianne Scherly Varela de Medeiros

Dissertação (mestrado) – UFRJ/COPPE/Programa de
Engenharia Elétrica, 2019.

Referências Bibliográficas: p. 50 – 52.

1. Centralidade. 2. Proximidade. 3. Múltiplos
Caminhos Disjuntos. 4. Fator de Conectividade. I.
Campista, Miguel Elias Mitre *et al.* II. Universidade
Federal do Rio de Janeiro, COPPE, Programa de
Engenharia Elétrica. III. Título.

*À minha família e a minha
amada mãe Mírian Pinto
(in memoriam).*

Agradecimentos

Agradeço a Deus por me permitir este momento e por me dar forças para prosseguir. Agradeço a minha amada mãe Mírian Pinto que me apoiou incondicionalmente durante toda a minha vida, principalmente por minha formação e nesta jornada. Agradeço a minha irmã Andreia C. Maciel Barbosa por todo apoio e incentivo, principalmente para iniciar esta jornada. Agradeço ao meu filho Gabriel Maciel S. Thomaz, a fonte de inspiração de minha vida e Alessandro S. Thomaz por todo o apoio.

Agradeço aos meus orientadores, Miguel Elias Mitre Campista e Dianne Scherly Varela de Medeiros por todo o aprendizado e por todo o apoio e ajuda em todos os momentos, profissional e pessoal, por serem exemplos a seguir, e me inspirarem profissionalmente. Agradeço aos professores Marcelo Gonçalves Rubinstein e Rodrigo de Souza Couto por fazerem parte da banca examinadora.

Agradeço aos demais professores do Grupo de Teleinformática e Automação pela contribuição na minha formação, aos colegas do laboratório pelo companheirismo e amizade, e a toda a equipe de funcionários e colaboradores do Programa de Engenharia Elétrica - PEE/COPPE/UFRJ.

Cada um que faz parte da minha vida me ensina a ser melhor a cada dia. Cada um da sua forma, promovendo sempre o aprendizado. Agradeço a Deus por tê-los em minha vida, mesmo que às vezes por um curto momento.

Por fim, agradeço ao apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES), Código de Financiamento 001; do CNPq; da FAPERJ; e da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processos nº 15/24494-8 e 15/24490-2.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

CENTRALIDADE DE PROXIMIDADE POR MÚLTIPLOS CAMINHOS DISJUNTOS

Mariana de Souza Maciel Barbosa

Julho/2019

Orientadores: Miguel Elias Mitre Campista

Dianne Scherly Varela de Medeiros

Programa: Engenharia Elétrica

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, que podem ser estratégicos para manter a conectividade da rede. Assim, esta dissertação propõe a centralidade de proximidade por múltiplos caminhos disjuntos, que extrapola a proximidade tradicional, considerando múltiplos caminhos disjuntos mais curtos, e quase mais curtos. O fator de conectividade φ limita a quantidade desejável de múltiplos caminhos disjuntos. A ideia é identificar nós que estejam simultaneamente mais próximos de todos os demais nós e sejam multiplemente conectados. Esses nós são importantes para desempenhar tarefas que exijam maior disponibilidade. A métrica proposta é avaliada através da comparação com outras métricas de centralidade. Os resultados confirmam que o nó mais central da métrica proposta é mais acessível quando ocorrem falhas em nós aleatórios da rede. Além disso, mostram que o uso dos múltiplos caminhos pode reclassificar pelo menos 59% dos nós da rede nos conjuntos de dados utilizados, que permite identificar nós mais bem conectados e atribuí-los uma maior importância.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

DISJOINT MULTIPATH CLOSENESS CENTRALITY

Mariana de Souza Maciel Barbosa

July/2019

Advisors: Miguel Elias Mitre Campista

Dianne Scherly Varela de Medeiros

Department: Electrical Engineering

Traditional centrality metrics consider only shortest paths, neglecting the existence of bit-longer paths between nodes in a network, which can be strategic to maintain network connectivity. This work proposes the disjoint multipath closeness centrality, which extrapolates the traditional closeness by considering multiple shortest and *quasi*-shortest disjoint paths. The connectivity factor φ limits the desirable number of multiple disjoint paths. The idea is to identify nodes that are simultaneously multiply-connected and close to all the remaining nodes. Such nodes are important 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 use of multiple disjoint paths can reclassify at least 59% of nodes in the evaluated datasets. Hence, our proposed metric is able to identify better-connected nodes and assign them more importance.

Sumário

Lista de Figuras	x
Lista de Tabelas	xiii
1 Introdução	1
1.1 Objetivos	3
1.2 Contribuições	3
1.3 Organização do texto	4
2 Notações e Definições	6
2.1 Modelo da rede	6
2.2 Incorporação de caminhos disjuntos	7
3 Métricas Relacionadas e Aplicações	10
3.1 Centralidade de proximidade tradicional	11
3.2 Centralidade de informação	11
3.3 Centralidade de intermediação ρ -geodésica	12
3.4 Aplicações das métricas de centralidade	12
3.5 Aplicação dos múltiplos caminhos disjuntos	13
4 Proximidade baseada em Múltiplos Caminhos Disjuntos	15
4.1 Abordagem conceitual	15
4.2 Formalização	16
4.3 Implementação	17
4.4 Comparação entre os ranqueamentos das centralidades de proximidade	20
5 Conjuntos de Dados	23
5.1 RNP	23
5.2 Renater	24
5.3 Géant	24
5.4 Train Bombing Network	24
5.5 Netscience	25

5.6	Sumário	25
6	Resultados e Discussões	27
6.1	Metodologia de avaliação	27
6.2	Influência do fator de conectividade no ranqueamento	28
6.3	Correlação entre as métricas	34
6.4	Alcançabilidade dos nós em presença de falha	41
6.4.1	Cenário 1: acessibilidade do nó mais central da rede RENA- TER em presença de falha única	42
6.4.2	Cenário 2: acessibilidade dos nós mais centrais da rede RE- NATER em presença de múltiplas falhas	44
7	Conclusões	47
	Referências Bibliográficas	50

Lista de Figuras

2.1	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 entre v_s e v_t , aumentando a resiliência da comunicação.	8
4.1	O número de caminhos computados nem sempre atinge φ . É possível que (a) nenhum caminho quase mais curto possa ser computado. (b) Caso existam caminhos quase mais curtos disjuntos ao mais curto e entre si, eles são considerados no cálculo da métrica.	20
5.1	Topologias das redes (a) RNP, (b) GEANT, (c) RENATER, (d) Train Bomb e (e) Netscience. Quanto maior o nó, maior seu grau e quanto mais escuro o nó, maior sua proximidade.	26
6.1	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. O valor máximo de φ para o qual caminhos não são mais encontrados é maior para redes com grau médio maior.	29
6.2	Total de caminhos normalizados entre todos os nós da rede usados para computar a métrica proposta. A normalização é em relação à maior quantidade de caminho computadas para cada topologia. A rede Netscience possui menor incremento de caminhos disjuntos, enquanto a rede Train Bombing possui a maior.	30

6.3	Frequência de reclassificação dos nós de acordo com a variação do fator de conectividade φ . A rede RENATER apresenta a maior frequência de reclassificação quando se começa a considerar os caminhos disjuntos ($\varphi = 0 \rightarrow \varphi = 1$), enquanto a rede RNP apresenta a menor frequência. A rede Train Bombing possui o decaimento da frequência de reclassificação mais suave do que o das demais redes, e estabiliza em $\varphi = 8$	31
6.4	Variação da Reclassificação de $\varphi = 0$ para $\varphi = 1$ por nó de cada topologia. As maiores variações ocorrem para nós que ocupam as posições mais centrais da classificação. As variações para nós mais bem classificados são menores mas bastante significativas. A RENATER é a única topologia analisada a variar a 1 ^a posição da classificação.	33
6.5	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 para a maioria das topologias analisadas.	35
6.6	RNP. As métricas apresentam alta correlação, ainda assim, nota-se que ocorre reclassificação dos nós em maior ou menor amplitude. A rede da RNP é a que apresenta maior correlação entre a proximidade por múltiplos caminhos disjuntos e a tradicional, mesmo assim nós chegam a ganhar 5 posições.	36
6.7	RENATER. As métricas apresentam alta correlação, ainda assim, nota-se que ocorre reclassificação nos nós em maior ou menor amplitude. A correlação entre a proximidade por múltiplos caminhos disjuntos e a tradicional ou de informação são próximas. Em ambos os casos nós perdem até 17 posições.	38
6.8	GEANT. As métricas apresentam alta correlação, ainda assim, nota-se que ocorre reclassificação dos nós em maior ou menor amplitude. É possível notar a redução da correlação entre a métrica tradicional e a proposta ao aumentar φ de 1 para 3.	39
6.9	Train Bombing. As métricas apresentam alta correlação, ainda assim, nota-se que ocorre reclassificação so nós em maior ou menor amplitude. A rede GEANT maior correlação entre a métrica proposta e de informação, mesmos assim nós chegam a ganhar 17 posições.	40

6.10	Netscience. As métricas apresentam alta correlação, ainda assim, nota-se que ocorre reclassificação os nós em maior ou menor amplitude. Esta rede possui menor correlação entre a proximidade por múltiplos caminhos disjuntos e de informação. a variação de classificação entre as métricas possui maior dispersão e uma quantidade menor de nós permanecem na mesma posição.	41
6.11	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. Tanto para (a) falhas aleatórias, quanto para (b) falhas em nós dos caminhos mais curtos, valores próximos a $y = 0$ pelo lado positivo indicam menor aumento do custo médio para alcançar o nó mais central.	43
6.12	Média de nós que perdem a comunicação com o grupo formado pelos 20% nós mais bem classificados segundo a métrica tradicional e a proximidade de caminhos disjuntos. Considera-se φ de 1 até o valor máximo que mantém o ranqueamento estável para cada conjunto de dados. A métrica proposta apresenta melhores resultados para a maioria das topologia em quaisquer cenários de falha avaliados. . . .	45

Lista de Tabelas

2.1	Sumário da notação utilizada.	9
4.1	Valores das métricas de centralidade para o cenário da Figura 4.1 . . .	21
5.1	Principais características dos conjunto de dados.	25
6.1	Valores máximos de φ quando ocorre estabilização do ranqueamento, para cada conjunto de dados.	32

Capítulo 1

Introdução

A importância da ciência de redes é cada vez mais crescente devido à sua abrangência multidisciplinar. Uma das abordagens frequentemente utilizada nessa área de pesquisa é a utilização de grafos para estudo de aplicações cotidianas nas mais diversas áreas do conhecimento, incluindo química, biologia, economia, análise de redes sociais e de comunicação, dentre outras [1]. Esta dissertação foca nas redes sociais e de comunicação. Na análise de redes sociais, avalia-se as relações entre os nós e como as informações, sejam elas notícias, rumores, doenças, etc., podem ser disseminadas em (ou por) um grupo social. Na análise de redes de comunicação, também se avalia a disseminação de informação. Nesse contexto, no entanto, a informação passa a ser interpretada como fluxos de dados trocados entre pares de nós quaisquer da rede. Em ambas as redes, as métricas de centralidades são de grande importância para o estudo da disseminação de informação. A centralidade de um nó permite caracterizá-lo de forma a revelar como ele se relaciona com outros nós da rede. A partir dessa caracterização, é possível definir que papel o nó deve desempenhar na rede.

A caracterização de nós utilizando métricas de centralidade exige o conhecimento de informações topológicas da rede [2, 3]. A partir dessas informações, essas métricas quantificam a importância de cada nó. É importante destacar que o significado de *importância* varia de acordo com a métrica utilizada. Em redes de comunicação, é fundamental identificar os nós mais importantes, ou seja, mais centrais. Essa identificação auxilia no desenvolvimento de estratégias para alocação de recursos ou serviços [4], ou até mesmo na elaboração de contramedidas que permitam uma convergência mais rápida e eficiente da rede após a detecção de uma falha [5]. Em redes de comunicação, as métricas de centralidade são tipicamente usadas para identificar nós concentradores de tráfego ou nós com elevado poder de disseminação de informação [6]. Isso é possível porque nós com essas características frequentemente apresentam elevada centralidade. O poder de disseminação de informação remete à rapidez com que um nó dissemina informação para todos os outros nós da rede. Nós

que são bons disseminadores também podem ser alcançados rapidamente por todos os outros nós, caso a informação possa fluir no sentido inverso. Sendo assim estes nós podem exercer funções como conter (ou ser) um servidor de dados ou até mesmo possuir uma maior capacidade de armazenamento para facilitar o acesso a determinadas informações ou aplicações. Nesse contexto, a métrica mais adequada para avaliação da importância do nó é a *centralidade de proximidade (closeness)*. Para determinar essa importância, a proximidade considera a distância para alcançar todos os demais nós da rede [2]. Dessa forma, os nós mais centrais são aqueles que, na média, estão mais próximos de todos os outros nós [2]. O conceito de proximidade é importante também para as redes sociais. Nessas redes, nós com elevada centralidade de proximidade são vetores de disseminação capazes, por exemplo, de espalhar rapidamente, notícias, rumores e doenças. Dessa forma, políticas eficazes para neutralização (notícias falsas, rumores e doenças) ou potencialização (notícias urgentes) dos nós podem ser propostas.

Tradicionalmente, a centralidade de proximidade utiliza para computar a importância de um nó, apenas as distâncias dos caminhos mais curtos. Essa visão negligencia possíveis caminhos alternativos, que mesmo tendo tamanhos um pouco maiores, podem ser usados para aumentar o poder de disseminação ou para aumentar a chance de a informação ser de fato disseminada. Considerando que a informação também possa fluir no sentido inverso, os nós disseminadores alcançáveis por múltiplos caminhos, mesmo que mais longos, são potencialmente mais acessíveis do que aqueles alcançados por poucos caminhos ou apenas por caminhos mais curtos. Isso é válido porque a comunicação entre dois nós conectados por poucos caminhos pode ser prejudicada devido a falhas de nós ou enlaces nos caminhos, podendo chegar ao caso extremo no qual a comunicação torna-se inviável. Por isso, diversos autores defendem o uso de outros caminhos além dos mais curtos para computar a centralidade de um nó, além de defenderem que os caminhos mais curtos devem ser priorizados [6–10]. 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 [11]. A centralidade de informação [6] e a 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 e, assim, não limitam o número máximo de caminhos. Além disso, essas métricas não consideram que em redes dinâmicas, falhas podem ocorrer com frequência. Caso uma falha ocorra, múltiplos caminhos podem deixar de existir. Dessa forma, dentre os múltiplos caminhos que conectam dois nós, aqueles que apresentam maior grau de disjunção devem ter maior importância do que aqueles que compartilham diversos nós ou enlaces entre si. Assim, nesta dissertação, também se defende a necessidade de utilizar múltiplos caminhos, privilegiando aque-

les que forem mais curtos. No entanto, mais do que isso, esta dissertação prioriza também o uso de caminhos disjuntos.

1.1 Objetivos

Esta dissertação tem como objetivo geral identificar o potencial de acessibilidade de um nó quando múltiplos caminhos disjuntos são considerados. Para tanto propõe-se uma nova métrica de proximidade, a *centralidade de proximidade por múltiplos caminhos disjuntos*, que extrapola o conceito da centralidade de proximidade tradicional. A métrica proposta considera a existência de múltiplos caminhos vértice-disjuntos para computar a importância de um nó. O nó mais central é aquele que for conectado a todos os outros nós da rede através de múltiplos caminhos disjuntos com o menor custo médio possível. O número máximo de caminhos vértice-disjuntos é limitado por um *fator de conectividade* - φ , reduzindo a complexidade do algoritmo desenvolvido para computar a métrica. Para alcançar o objetivo estabelecido, a métrica proposta é avaliada frente a outras métricas de centralidade de proximidade comuns, investigando-se a influência do fator de conectividade na classificação, a correlação entre as métricas e a acessibilidade dos nós mais centrais em situação normal de funcionamento da rede, em presença de falhas únicas aleatórias e em presença de falhas múltiplas também aleatórias.

1.2 Contribuições

Esta dissertação tem como principal contribuição o desenvolvimento de uma nova métrica de centralidade de proximidade. Essa métrica utiliza múltiplos caminhos disjuntos, atribuindo maior importância para aqueles de menor tamanho, para computar a importância de um nó. A métrica proposta auxilia na identificação do potencial de conectividade de um nó quando múltiplos caminhos vértice-disjuntos são considerados. 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. Já aplicada as redes sociais, a métrica identifica nós com rápida difusão da informação ao destino, passando por nós intermediários distintos. Além de propor uma nova métrica, esta dissertação também contribui para evoluir o estado-da-arte ao aplicar o conceito de caminhos quase mais curtos para a utilização de múltiplos caminhos disjuntos na rede.

A avaliação da métrica proposta é feita através de uma análise comparativa com duas outras métricas de centralidade de proximidade comuns: a centralidade de proximidade tradicional [12, 13]; e a centralidade de informação [6]. Essas métricas foram escolhidas por representarem os dois extremos de um “espectro de proximida-

des”, no qual de um lado são utilizados apenas os caminhos mais curtos (proximidade tradicional), e do outro, todos os caminhos (centralidade de informação). Na avaliação comparativa, são utilizados conjuntos de dados que representam topologias reais de redes de comunicação de longa distância de educação e pesquisa. Também são utilizadas redes sociais com características distintas entre elas e que possuem maior número de nós em relação às redes de longa distância usadas. A avaliação da métrica proposta tem como foco (i) identificar a influência do fator φ no ranqueamento dos nós, (ii) analisar a correlação entre a proposta e as demais métricas e (iii) investigar a variação da alcançabilidade dos nós após a ocorrência de falhas em diferentes nós da rede. O parâmetro φ define a quantidade de múltiplos caminhos disjuntos desejáveis, com base nos caminhos quase mais curtos definido por Medeiros et al. [8, 9]. Nesta dissertação, mostra-se que o parâmetro φ é capaz de capturar os múltiplos caminhos de forma que a métrica identifica os nós que são mais rapidamente acessíveis através dos múltiplos caminhos disjuntos. Desse modo, uma das possíveis funções da métrica proposta nesta dissertação é identificar nós mais resilientes em caso de falhas na rede. Os resultados mostram que a proximidade por múltiplos caminhos disjuntos é capaz de apontar nós que devem ser reclassificados por estarem mais bem conectados, chegando a reclassificar 91% dos nós de uma das redes. Apesar disso, a métrica proposta ainda mantém 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.

Os resultados e contribuições desta dissertação estão reportados no artigo intitulado “*Centralidade de Proximidade por Múltiplos Caminhos Disjuntos: Aplicação em Redes de Longa Distância*”, publicado nos Anais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC '19) [14].

1.3 Organização do texto

Esta dissertação está organizada em sete capítulos. O Capítulo 2 apresenta as notações e definições para melhor compreensão desta dissertação. O Capítulo 3 revisa as métricas de proximidade e discute os trabalhos relacionados à proposta desta dissertação. O Capítulo 4 define e formaliza a métrica de proximidade por múltiplos caminhos disjuntos proposta nesta dissertação. Os conjuntos de dados utilizados para avaliar a métrica proposta são descritos no Capítulo 5. O Capítulo 6 apresenta a avaliação comparativa da métrica proposta e discute os resultados obtidos. Por fim, o Capítulo 7, apresenta a conclusão desta dissertação e lista possibilidades de

trabalhos futuros.

Capítulo 2

Notações e Definições

Este capítulo descreve o modelo de rede e formaliza os conceitos e definições utilizadas nesta dissertação para facilitar a compreensão da proposta desenvolvida. Os Capítulos 3 e 4 apresentam respectivamente algumas métricas existentes de grande importância para esta dissertação e a métrica proposta, para as quais os conceitos e definições descritos nesta seção são necessários.

2.1 Modelo da rede

Uma rede pode ser representada por um grafo $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W})$, em que \mathcal{V} , \mathcal{E} e \mathcal{W} denotam respectivamente os conjuntos de nós, de enlases e de custos dos enlases. Dessa forma, um enlace $\varepsilon_{i,j} \in \mathcal{E}$, de custo $\omega_{i,j} \in \mathcal{W}$, existe entre dois nós v_i e $v_j \in \mathcal{V}$ se eles forem vizinhos. Em grafos orientados o conceito de vizinhança é assimétrico. Dessa forma, dois nós vizinho v_i e v_j possui um enlace $\varepsilon_{i,j}$, mas o o enlace $\varepsilon_{j,i} \in \mathcal{E}$ só existe se v_j também for vizinho de v_i . O grafo \mathcal{G} que representa a rede é dito conexo se existir pelo menos um caminho que conecte todos os pares de nós em \mathcal{V} , caso contrário é dito não conexo. Nesta dissertação, utilizam-se redes representadas por grafos não-orientados e com peso de aresta unitário, para possibilitar a utilização dos algoritmos com menor complexidade temporal. No entanto, essa simplificação não prejudica nem inviabiliza a aplicação da métrica proposta em grafos orientados e em grafos ponderados.

Um caminho $p_{s,t}$ entre os nós v_s e v_t é formado por uma sequência de vértices adjacentes distintos interligados por enlases $\varepsilon_{i,j}$ com seus respectivos pesos $\omega_{i,j}$. O custo $\delta_{s,t}$ associado ao caminho $p_{s,t}$ é obtido a partir da soma dos pesos $\omega_{i,j}$ dos enlases que compõem o caminho. Se os pesos forem unitários, o custo total de um caminho $p_{s,t}$ torna-se igual ao número de saltos entre v_s e v_t . Onde o número de saltos é a quantidade de nós intermediários pelo quais a informação de um nó v_s passa até alcançar um nó v_t , inclusive. Um par de nós v_s, v_t pode estar interligado por n caminhos, cada um representado por $p_{s,t}^{(n)}$, com custo $\delta_{s,t}^{(n)}$. Dentro desse conjunto de

caminhos existem os caminhos mais curtos e os quase mais curtos. O caminho que apresenta o menor custo total $\delta_{s,t}^*$ é chamado de caminho mais curto $p_{s,t}^*$. O caminho quase mais curto, por sua vez, tem custo um pouco maior do que o do caminho mais curto. O custo máximo do caminho quase mais curto é limitado a $\delta_{s,t}^* + \rho$, onde ρ é um fator de espalhamento pré-definido [8]. O fator de espalhamento ρ representa o custo adicional que uma entidade está disposta a pagar para chegar ao destino.

2.2 Incorporação de caminhos disjuntos

A informação em uma rede pode seguir por caminhos de formas distintas, nas redes de comunicação, por exemplo, é comum utilizar os caminhos mais curtos. No entanto o uso destes caminhos é limitante, uma vez que mesmo nessas redes a disseminação de informação pode seguir um processo que utiliza outros caminhos além dos mais curtos. Além disso, em redes que prezam a resiliência, diversos caminhos conectando os pares de nós devem existir, mesmo que possuam custos diferentes. Dessa forma, em caso de falha ou sobrecarga de nós que pertencem a algum dos caminhos, a informação pode ser transferida pelos caminhos restantes. Nesse sentido, o uso de múltiplos caminhos ajuda a manter a comunicação entre os nós. No entanto, para que os múltiplos caminhos não aumentem significativamente o custo da transferência da informação, feita antes da falha ou sobrecarga pelos caminhos mais curtos, é benéfico considerar adicionalmente apenas caminhos que sejam um poucos mais longos que os mais curtos, i.e., os caminhos quase mais curtos [8, 9]. Logo, os caminhos quase mais curtos são considerados para acomodar a ideia de que utilizar apenas os caminhos mais curtos não é suficiente para identificar nós importantes em algumas redes.

Defende-se a ideia de que não é suficiente considerar apenas todos os caminhos mais curtos e quase mais curtos quando se deseja enfatizar a importância da conectividade da rede. Isso porque não é suficiente que existam múltiplos caminhos entre dois nós para que eles sejam considerados como bem conectados. Para que isso seja verdade, os múltiplos caminhos precisam ser disjuntos entre si. Dessa forma, a exclusão de nós em um caminho inviabiliza a comunicação por esse caminho, mas tem menor probabilidade de inviabilizar os caminhos restantes. Essa probabilidade é tão menor quanto maior for o grau de disjunção entre os caminhos. Logo, quanto maior o grau de disjunção, mais bem conectados estão os nós. No caso extremo, apenas a origem e o destino são comuns aos múltiplos caminhos. Assim, dentre os caminhos entre v_s e v_t , dois caminhos $p_{s,t}^{(x)}$ e $p_{s,t}^{(y)}$ são 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.

Nesta dissertação, para determinar a centralidade de um nó são utilizados os

caminhos mais curtos e múltiplos caminhos quase mais curtos 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.

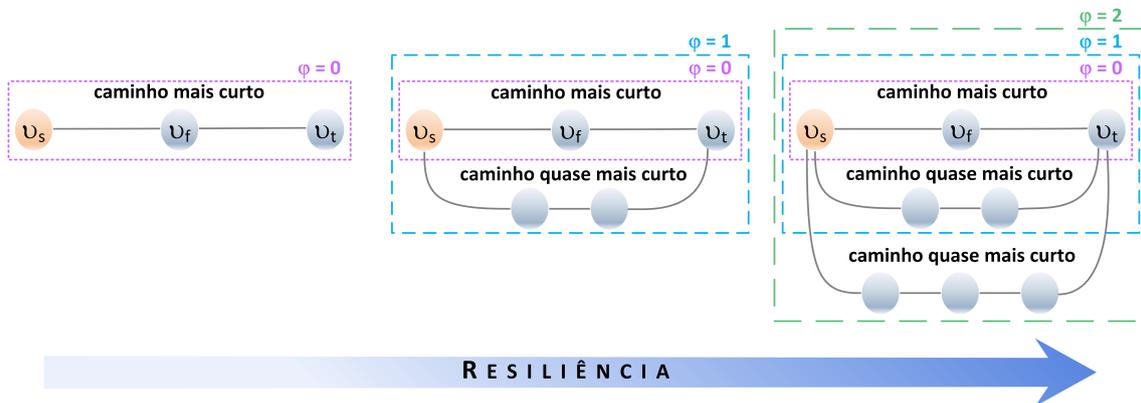


Figura 2.1: O nó v_s se torna mais bem conectado à medida em que mais caminhos disjuntos existem para alcançar o destino (v_t), mesmo que os caminhos sejam quase mais curtos. O aumento de φ contribui para a melhora da conectividade entre v_s e v_t , aumentando a resiliência da comunicação.

A Figura 2.1 ilustra a ideia por trás da definição do fator de conectividade φ . Quando existe apenas um caminho entre o par de nós v_s, v_t como ilustrado na imagem da esquerda, uma falha no nó v_f desconecta o par de nós origem-destino. Caso existisse outro caminho, totalmente vértice-disjunto, mesmo que um pouco mais longo, a falha em v_f resultaria apenas no aumento do custo para chegar ao destino v_t . Quanto mais caminhos entre o par de nós origem-destino existirem, mais bem conectados estarão os nós, mesmo se os custos dos caminhos forem diferentes. Logo, caso haja uma falha no nó v_f , a rede se mantém conectada com maior probabilidade devido à existência de outros caminhos entre o par de nós. O ônus da utilização dos caminhos restantes é a possibilidade de aumentar o custo para chegar ao destino,

caso não sejam caminhos mais curtos. O nó v_s , por exemplo, continuaria podendo alcançar v_t através dos caminhos quase mais curtos, com 2 e 3 saltos, como ilustrado na imagem da direita. Quanto mais bem conectados são os nós, maior é a resiliência da rede. Essa percepção justifica a métrica de centralidade proposta nesta dissertação, a qual atribui maior importância aos nós com mais caminhos totalmente disjuntos até todos os outros nós da rede, priorizando sempre os menores caminhos.

A Tabela 2.1 sumariza a notação utilizada nesta dissertação.

Tabela 2.1: Sumário da notação utilizada.

Notação	Significado
\mathcal{V}	Conjunto de nós
\mathcal{E}	Conjunto de enlaces
\mathcal{W}	Conjunto de custos
$ \mathcal{V} $	Quantidade de nós do conjunto \mathcal{V}
v_i	Nó i
$\varepsilon_{i,j}$	Enlace entre os nós i e j
$\omega_{i,j}$	Custo do enlace entre os nós i e j
$p_{i,j}$	Caminho entre os nós i e j
$\delta_{i,j}$	Custo do caminho entre os nós i e j
$p_{i,j}^{(n)}$	Caminho n entre os nós i e j
$\delta_{i,j}^{(n)}$	Custo do caminho n entre os nós i e j
$p_{i,j}^*$	Caminho mais curto entre os nós i e j
$\delta_{i,j}^*$	Custo do caminho mais curto entre os nós i e j
ρ	Fator de espalhamento
φ	Fator de conectividade
$\Delta_{i,j}^{(n)}$	Custo agregado dos n caminhos entre os nós i e j

Capítulo 3

Métricas Relacionadas e Aplicações

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 pelos nós. Os principais tipos de centralidade são o grau, a intermediação, a proximidade e as variantes dessas métricas [10].

O grau tem relação com a quantidade de nós adjacentes de um nó; em outras palavras, está associado com a popularidade de um nó. Essa métrica é muito simples de ser computada, pois não exige conhecimento de toda a rede, mas sim apenas conhecimento local. A intermediação tem relação com o controle de um nó sobre os fluxos entre outros nós da rede. Dessa forma, nós que participam mais dos caminhos entre outros nós possuem uma capacidade maior para controlar e influenciar o fluxo entre os demais nós. 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.

Esta dissertação investiga a acessibilidade dos nós em uma rede, considerando que os nós mais importantes são os multiplamente conectados e que são alcançados, ou alcançam, os demais nós mais rapidamente. 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, ou como a identificação de um determinado nó que possa distribuir rapidamente a informação pela rede sem que a informação precise passar pelos mesmo intermediários. Assim, esta dissertação foca na proximidade por ser a métrica mais adequada para a análise de acessibilidade, devido à sua relação com a rapidez com que um fluxo é difundido para todos os nós [2] ou com que fluxos partindo de todos os outros nós da rede são concentrados em um único nó.

3.1 Centralidade de proximidade tradicional

A *centralidade de proximidade tradicional* (C_{trad}) inicialmente proposta por Bavelas et al. [12] e aprimorada por Beauchamp et al. [13], que a torna independente do número de nós da rede, considera como nó mais central os nós que estão, em média, mais próximos dos demais nós da rede através dos caminhos mais curtos. Sendo assim, a Equação 3.1 define formalmente a proximidade tradicional para determinar a importância de um nó v_s :

$$C_{trad}(v_s) = \frac{|\mathcal{V}| - 1}{\sum_{t=1}^{|\mathcal{V}|} \delta_{s,t}^*}, \quad (3.1)$$

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.

Métricas de proximidade são aplicadas quando se deseja identificar a independência ou eficiência dos nós [2]. Porém a utilização apenas dos caminhos mais curtos pode ser vista como uma desvantagem da abordagem tradicional, uma vez que podem existir redes em que um fluxo enviado por um nó percorre caminhos que não são tão diretos quanto os mais curtos.

3.2 Centralidade de informação

A necessidade de considerar os demais caminhos, além dos mais curtos pode ocorrer devido ao processo de comunicação na rede ser indefinido, permitindo que os fluxos percorram quaisquer caminhos aleatoriamente, ou devido à canalização intencional através de nós intermediários específicos [6, 7]. Assim, o fluxo não percorre apenas os caminhos mais curtos, e os outros caminhos existentes entre os pares de nós também devem ser contabilizados para calcular a centralidade. Nesse sentido, Stephenson e Zelen defendem que a medida de informação presente em um caminho $p_{s,t}^{(n)}$, de custo $\delta_{s,t}^{(n)}$, existente entre v_s e v_t é definida por $I_{s,t}^{(n)} = 1/\delta_{s,t}^{(n)}$, e propõem a *centralidade de informação* (I_s) [6], definida formalmente pela Equação 3.2:

$$I_s(v_s) = \frac{|\mathcal{V}|}{\sum_{t=1}^{|\mathcal{V}|} \frac{1}{I_{s,t}^{(n)}}}, \quad (3.2)$$

onde $I_{s,t} = \sum_n 1/I_{s,t}^{(n)}$ é a soma das medidas de informação existentes entre v_s e v_t . No entanto, os caminhos mais longos são menos valiosos para serem controlados nas aplicações estudadas até hoje [7].

3.3 Centralidade de intermediação ρ -geodésica

Em uma rede de computadores tradicional, ou em uma rede de transporte, o fluxo, em sua maioria, tentará utilizar o caminho mais curto. No entanto, este caminho não será necessariamente o mais adequado, pois ele pode já estar congestionado ou estar na iminência desse acontecimento. Dessa forma, é interessante considerar caminhos um pouco mais longos, além dos mais curtos. Por essa razão, Medeiros et al. [8] propuseram a *centralidade de intermediação ρ -geodésica*, formalizada pela Equação 3.3:

$$B_{\rho}(v_k) = \sum_{\substack{i \in \mathcal{V} \\ \delta_{i,k} + \delta_{k,j} - \delta_{i,j}^* \leq \rho}} \sum_{j \in \mathcal{V}} \frac{n_{i,j}^*(v_k) + n_{i,j}(v_k)}{n_{i,j}^* + n_{i,j}} \times \frac{\delta_{i,j}^*}{\delta_{i,k} + \delta_{k,j}}. \quad (3.3)$$

A intermediação ρ -geodésica considera todos os caminhos menores ou iguais a um limite determinado pelo parâmetro ρ . Esse limite é baseado no custo do caminho mais curto entre os pares de nós, e todos os caminhos que são maiores do que os mais curtos e menores ou iguais ao limite definido são chamados *caminhos quase mais curtos*. Esse conceito é explorado nesta dissertação. Medeiros et al. [8] afirmam que esses caminhos são importantes para alterar a importância dos nós em redes nas quais é necessário considerar múltiplos caminhos.

A intermediação ρ -geodésica considera todos os caminhos dentro do limite predefinido, não discriminando caminhos com vértices em comum (caminhos não disjuntos), que conectam um mesmo par de nós. Contudo, caminhos disjuntos são essenciais para reduzir uma possível sobrecarga em um nó que faça parte dos múltiplos caminhos e para aumentar a resiliência da rede. Assim como Medeiros et al. esta dissertação também considera os caminhos quase mais curtos, limitado por um fator de conectividade, definido no Capítulo 2, para propor uma nova métrica de centralidade de proximidade. Dessa forma, diferente dos trabalhos mencionados, a métrica proposta considera que existe um limite máximo para o número de caminhos considerados e que eles devem ser obrigatoriamente disjuntos entre si. Considerando uma abordagem mais restritiva, nesta dissertação são usados apenas caminhos vértice-disjuntos, o que é justificável no projeto de redes tolerantes a falhas.

3.4 Aplicações das métricas de centralidade

As métricas de centralidade são aplicadas em diferentes áreas de estudos. Quando aplicada a redes de computadores, as centralidades podem identificar nós estratégicos na rede, bem como buscar soluções de convergência da rede para resolver falhas ou congestionamentos.

A centralidade é usada por Maccari [5], para otimizar a convergência da rede após uma detecção de falha. A intermediação é utilizada para identificar os nós mais centrais e então reduzir o tempo de mensagens periódicas enviadas por estes nós e aumentar o tempo de mensagens dos demais nós da rede, e desta forma reduz a quantidade das mensagens periódicas e reestabelece com maior rapidez as rotas da rede. A centralidade é também usada para otimizar a engenharia de tráfego baseada em SDN (*Software Defined Networking*) [15]. O trabalho utiliza a intermediação tradicional para classificar a importância dos enlaces das redes, a partir de algoritmos com a utilização da SDN, e então realizar a distribuição de tráfego de acordo com esse algoritmo.

A identificação dos nós com alto valor de centralidade na rede pode ser de grande importância para a alocação de recursos ou serviços na rede. Essa abordagem é feita por Bouet [4], que avalia a alocação de funções vDPI (*virtual Deep Packet Inspection*) em uma infraestrutura de rede que utiliza NFV (*Network Functions Virtualization*). O DPI realiza o monitoramento da rede em tempo real, possibilitando a avaliação dos fluxos das redes para realizar gerenciamento de tráfego e serviços de segurança. O trabalho tem como objetivo avaliar a melhor alocação das vDPIS em uma rede visando minimizar o custo de sua alocação, através da formulação de um algoritmo guloso baseado na centralidade de intermediação tradicional. Já Hui propõe um algoritmo para detecção de comunidades (*cluster*) através do uso da métrica de proximidade tradicional e o sinal transmitido pelos nós [16]. A partir da classificação dos nós, calcula-se similaridades entre esses nós, de acordo com o sinal enviado, e então as comunidades são estabelecidas de forma iterativa, até que os grupos formados estejam estáveis.

3.5 Aplicação dos múltiplos caminhos disjuntos

Muitos estudos na literatura abordam a utilização de múltiplos caminhos disjuntos para aprimorar a comunicação de uma rede; no entanto, para cada abordagem a busca e seleção destes caminhos diferem. Para desenvolver um protocolo de roteamento para redes Ad Hoc, existem trabalhos que abordam o uso de caminhos disjuntos por arestas [17]. Um dos objetivos desse trabalho é reduzir a restrição de caminhos que ocorre quando se utilizam caminhos disjuntos pelos vértices. Essa abordagem é diferente da utilizada nesta dissertação, que considera a indisponibilidade de um nó e adota o uso de caminhos vértice-disjuntos. Na abordagem adotada por Sidhuand *et al.* [18], através de uma árvore geradora do grafo encontram-se os caminhos disjuntos a partir de um único caminho mais curto, de forma que adiciona-se um menor caminho disjunto por vez aos caminhos já computados. Essa proposta difere da utilizada nesta dissertação, que visa garantir o maior conjunto de caminhos

mais curtos disjuntos.

Com o objetivo de encontrar uma quantidade pré-estabelecida de caminhos disjuntos entre os nós, Suurballe utiliza o algoritmo de Dijkstra para encontrar o menor caminho e, através de transformações canônicas do grafo, encontra um maior conjunto de caminhos disjuntos possível [19]. Caso não seja possível atingir o número de caminhos predefinido, a procura é interrompida e são computados apenas os caminhos encontrados. Vale ressaltar que, esse método busca pelo conjunto de caminhos de menor custo, sendo assim, existe a possibilidade do caminho mais curto não estar dentre os caminhos computados. Essa proposta difere da desta dissertação por não privilegiar os caminhos mais curtos. Para maximizar o número de caminhos e otimizar a qualidade de serviço de roteamento, caso o algoritmo não encontre mais caminhos disjuntos que atendam os parâmetros, caminhos parcialmente disjuntos são adicionados aos caminhos mais curtos [20]. O algoritmo utiliza como parâmetros de escolha de caminhos, além dos tamanhos, a banda e o atraso dos caminhos. A proposta se mostra interessante por adicionar caminhos parcialmente disjuntos caso não exista caminhos disjuntos possíveis, o que difere bastante da abordagem proposta nesta dissertação. A computação dos caminhos disjuntos usados na métrica proposta nesta dissertação objetiva maximizar a quantidade de caminhos disjuntos de mesmo custo. A Seção 4.3 explica com detalhes a forma como esses caminhos são encontrados na rede.

Capítulo 4

Proximidade baseada em Múltiplos Caminhos Disjuntos

Neste capítulo propõe-se a *centralidade de proximidade por múltiplos caminhos disjuntos*, cujo objetivo é identificar o potencial de conectividade de um nó quando outros caminhos além dos mais curtos são usados; mais especificamente, quando múltiplos caminhos disjuntos são considerados para computar a métrica.

4.1 Abordagem conceitual

No Capítulo 3 apresentam-se métricas de centralidade de proximidade conhecidas, que delimitam um “espectro de proximidades”. Em um extremo do espectro, a proximidade tradicional utiliza apenas caminhos mais curtos. No outro extremo, a centralidade de informação utiliza todos os caminhos existentes, atribuindo maior contribuição para os mais curtos. A utilização de múltiplos caminhos é reconhecidamente necessária para quantificar a importância de nós em alguns tipos de redes [7, 11]. O uso desses caminhos pode surgir da necessidade de aumentar a resiliência ou a vazão da rede ou de reduzir os impactos negativos da concentração de fluxos sobre os caminhos mais curtos. No Capítulo 3 também se apresenta a centralidade de intermediação ρ -geodésica, que é conceitualmente diferente das métricas de proximidade, já que usa como base a intermediação tradicional. Apesar da diferença conceitual, essa métrica é apresentada porque ao introduzir a ideia de caminhos quase mais curtos, ela se torna um meio termo entre o uso de caminhos mais curtos apenas e o uso de todos os caminhos.

Existem redes nas quais a resiliência na comunicação é importante. Um exemplo simples é uma rede que utiliza um servidor que controla o acesso de usuários. Esse servidor deve estar sempre disponível para que os usuários possam acessar um determinado tipo de serviço. Além disso, é interessante que, na média, esse nó seja

alcançado pelos usuários o mais rápido possível. Logo, um nó com elevada proximidade é um bom candidato para desempenhar o papel de servidor de controle de acesso. No entanto, a disponibilidade desse servidor está associada à quantidade de caminhos que o alcançam. Mais especificamente, a quantidade de caminhos disjuntos que o alcançam. A disjunção é importante para reduzir a probabilidade de inviabilizar a comunicação com o servidor em caso de falha de nós em um dos caminhos até ele. Dessa forma, defende-se a ideia de que um nó deve ser tão mais importante quanto mais caminhos vértice-disjuntos chegam ou partem dele.

Com base na ideia discutida, nesta dissertação propõe-se a *Centralidade de Proximidade por Múltiplos Caminhos Disjuntos*. O objetivo é quantificar a importância dos nós levando em conta seu potencial de conectividade quando múltiplos caminhos disjuntos são contabilizados. A proximidade por múltiplos caminhos disjuntos de um determinado nó é capturada através da relação entre a quantidade e o tamanho dos caminhos disjuntos existentes entre o nó e os demais nós da rede, além dos caminhos mais curtos. A quantidade de caminhos é limitada pelo fator de conectividade φ 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, os nós múltiplemente conectados tendem a ser mais acessíveis por apresentarem alternativas ao caminho principal (mais curto). Ademais, esses nós têm maior probabilidade de permanecerem acessíveis se os múltiplos caminhos forem disjuntos entre si. Ainda, quanto menores forem os caminhos disjuntos, menor tende a ser o custo adicional para chegar ao destino quando a falha ocorre. Logo, nós múltiplemente conectados através de caminhos disjuntos curtos são boas opções para instalação de funcionalidades críticas em redes que se preocupam com a disponibilidade.

4.2 Formalização

Nesta dissertação, considera-se que um nó é tão mais importante quanto mais rapidamente ele for alcançável pelos demais nós, ou vice-versa, levando em conta, ainda, quão bem conectado o nó está com cada destino. Um nó é bem conectado ao destino se existem múltiplos caminhos totalmente vértice-disjuntos entre eles. As contribuições dos múltiplos caminhos disjuntos entre um par de nós v_s, v_t são inseridas na métrica proposta através do cálculo do custo agregado $\Delta_{s,t}^{(\varphi)}$, considerando ser desejável encontrar φ caminhos disjuntos. O custo agregado 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 aproximação média harmônica entre os custos dos φ caminhos disjuntos entre v_s e v_t , conforme a Equação 4.1:

$$\Delta_{s,t}^{(\varphi)} = \frac{1}{\sum_{n=0}^{\varphi} \frac{1}{\delta_{s,t}^{(n)}}}. \quad (4.1)$$

Considerando as contribuições dos caminhos entre o nó de interesse, v_s , e os demais nós, obtém-se a *centralidade de proximidade por múltiplos caminhos disjuntos*, conforme a Equação 4.2:

$$C_{\varphi}(v_s) = \frac{|\mathcal{V}| - 1}{\sum_{t=1}^{|\mathcal{V}|} \Delta_{s,t}^{(\varphi)}}, \quad (4.2)$$

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.

Dentre as propriedades da centralidade de proximidade por múltiplos caminhos disjuntos vale destacar que ela:

- considera não somente o tamanho dos caminhos, mas também a quantidade de caminhos;
- leva em conta caminhos mais curtos e o maior conjunto possível de caminhos vértice-disjuntos, dado um limite (φ);
- prioriza os caminhos de menor custo, que adicionam maior contribuição ao cálculo da métrica do que os caminhos de maior custo.

4.3 Implementação

A centralidade de proximidade por múltiplos caminhos disjuntos proposta nesta dissertação considera múltiplos caminhos disjuntos, limitados pelo fator de conectividade, tal que, $\varphi \in \mathbb{N}$. O cálculo da métrica é feito através de um algoritmo, cuja implementação requer algumas decisões de projeto, como a forma de escolha dos caminhos disjuntos. Os Algoritmos 1, 2 e 3 descrevem como a métrica é computada. Em uma visão macro, o Algoritmo 1 é responsável por retornar o valor da métrica proposta obtido para cada nó. Para tanto, é necessário implementar uma função

que encontra os caminhos mais curtos e os caminhos disjuntos. Essa funcionalidade é implementada no Algoritmo 2. Como o Algoritmo 1 é centrado em um nó raiz, a cada iteração calcula-se a contribuição dos caminhos encontrados entre o nó raiz e os demais. Logo, é necessário acumular essas contribuições, a medida que o nó raiz muda. Essa é a função do Algoritmo 3.

As entradas do Algoritmo 1 são o fator de conectividade φ , representando o número de caminhos disjuntos desejáveis e o conjunto de dados modelado como um grafo G através da biblioteca NetworkX 1.10 do Python. O algoritmo retorna como saída o vetor C_φ que contém a centralidade de proximidade por múltiplos caminhos disjuntos para todos os nós. Os custos dos caminhos disjuntos são armazenados a matriz DJ. A função INITIALIZE inicializa o vetor C_φ e a matriz DJ com os valores apropriados. O algoritmo chama recursivamente os Algoritmos 2 e 3. Os caminhos mais curtos e os disjuntos encontrados pelo Algoritmo 2 são armazenados no vetor $DJ[s][t]$. Esse vetor é, na verdade, um elemento da matriz DJ, localizado na posição s, t . A cada chamada, a função implementada no Algoritmo 3 atualiza o valor da centralidade de proximidade por caminhos disjuntos para cada nó da rede. Após executar todas as iterações, obtém-se o valor da centralidade proposta para todos os nós.

Algoritmo 1 PROXIMIDADE POR MÚLTIPLOS CAMINHOS DISJUNTOS

Entrada: φ, G

Saída: C_φ

```

1:  $C_\varphi \leftarrow \text{INITIALIZE}(\text{numNodes})$ 
2:  $DJ \leftarrow \text{INITIALIZE\_DJ}(\text{numNodes})$ 
3: para  $s \leftarrow 1, \text{numNodes}$  faça
4:   para  $t \leftarrow 1, \text{numNodes}$  faça
5:     se  $v_s \neq v_t$  então
6:        $DJ[s][t] \leftarrow \text{FIND\_DISJ\_PATHS}(G, s, t, \varphi)$ 
7:    $C_\varphi[s] \leftarrow \text{ACCUMULATE}(\varphi, \text{numNodes}, DJ[s])$ 

```

Nesta dissertação, considera-se que o custo de um enlace é unitário, de forma que o custo total de um caminho é igual ao número de saltos. Assim, o vetor $DJ_{s,t}$ contém os custos dos caminhos disjuntos entre v_s e v_t . Existem até φ caminhos entre esse par de nós, logo, o vetor de custos apresenta a seguinte formação: $DJ[s][t] = [\delta_{s,t}^{(0)}, \dots, \delta_{s,t}^{(\varphi)}]$. Isso significa que o primeiro elemento, $\delta_{s,t}^{(0)}$, é o custo dos caminhos mais curtos entre v_s, v_t . O segundo elemento, $\delta_{s,t}^{(1)}$, é o custo do primeiro menor caminho disjunto adicional encontrado. O terceiro elemento, $\delta_{s,t}^{(2)}$, é o custo do segundo menor caminho disjunto adicional encontrado. E assim por diante, até que o último elemento, $\delta_{s,t}^{(\varphi)}$, armazenará o custo do φ -ésimo caminho disjunto adicional encontrado. Os caminhos são encontrados pelo Algoritmo 2, através da função $\text{SHORT_DISJ_PATHS}_{s,t}$ que busca os caminhos mais curtos disjuntos entre o par origem-

destino de acordo com a limitação φ . Essa função retorna o tamanho dos caminhos, armazenados no vetor `auxPath`, e os nós que fazem parte dos caminhos encontrados, armazenados no vetor `nodesDJ`. Ambos os vetores estão inicialmente vazios. Os caminhos a serem considerados para computar a métrica são armazenados no vetor `DJ[s][t]`, retornado como saída do algoritmo. As entradas do algoritmo são o grafo da rede \mathbb{G} , os nós de origem e destino v_s, v_t , e o fator de conectividade φ . O vetor `DJ[s][t]` é inicializado com zeros através da função `INITIALIZE_DJ`. Em seguida inicia-se a busca pelos caminhos. Caso o valor de φ não seja alcançado na primeira busca, os nós intermediários pertencentes aos caminhos já computados são retirados do grafo (linha 9), e a partir do grafo restante novos caminhos são computados. A recursividade é finalizada (i) quando se encontra os φ caminhos ou (ii) quando não existem mais novos caminhos entre v_s e v_t .

Algoritmo 2 FUNÇÃO FIND_DISJ_PATHS ENTRE s E t

Entrada: G, s, t, φ

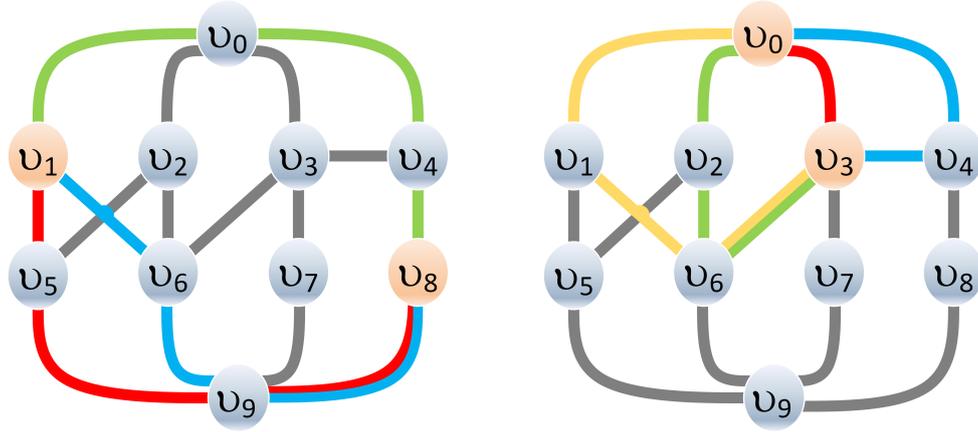
Saída: `DJ[s][t]`

```

1: nodesDJ, auxPath  $\leftarrow [\emptyset]$ 
2: paths, numAuxPath  $\leftarrow 0$ 
3: enquanto paths  $< \varphi$  faça
4:   auxPath, nodesDJ, numAuxPath  $\leftarrow$  SHORT_DISJ_PATHSs,t( $G, \varphi, paths$ )
5:   se auxPath =  $[\emptyset]$  então
6:     sair
7:   DJ[s][t]  $\leftarrow$  DJ[s][t] + auxPath
8:    $G.nodes()$   $\leftarrow$   $G.nodes()$  - nodesDJ
9:   paths  $\leftarrow$  paths + numAuxPath

```

A Figura 4.1 exemplifica a busca pelos caminhos disjuntos, no caso em que se deseja utilizar $\varphi = 3$ caminhos disjuntos adicionais entre os pares de nós. Na Figura 4.1(a) existem 3 caminhos mais curtos entre v_1 e v_8 , com $\delta_{1,8}^* = 3$. No entanto, como apenas os caminhos disjuntos interessam, ocorre uma escolha aleatória entre os caminhos $p_{1,8}^{(0)} = \langle v_1, v_5, v_9, v_8 \rangle$ (vermelho) e $p_{1,8}^{(2)} = \langle v_1, v_6, v_9, v_8 \rangle$ (azul). Escolhendo-se o caminho $p_{1,8}^{(0)}$, considera-se então 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), que são disjuntos. 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, na Figura 4.1(b) estão realçados o caminho mais curto ($p_{0,3}^{(0)}$, vermelho) entre v_0 e v_3 , um caminho quase mais curto de 2 saltos ($p_{0,3}^{(1)}$, azul) 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 de forma aleatória $p_{0,3}^{(2)}$ ou $p_{0,3}^{(3)}$, e mais o respectivo



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

(b) Total de caminhos considerados igual a φ .

Figura 4.1: O número de caminhos computados nem sempre atinge φ . É possível que (a) nenhum caminho quase mais curto possa ser computado. (b) Caso existam caminhos quase mais curtos disjuntos ao mais curto e entre si, eles são considerados no cálculo da métrica.

caminho de 5 saltos.

Por fim, após computar todos caminhos disjuntos possíveis de v_s para os demais nós da rede, a função **ACCUMULATE**, mostrada no Algoritmo 3, atualiza gradativamente o valor da centralidade proposta nesta dissertação conforme as Equações 4.1 e 4.2. Como a função é chamada de forma recursiva pelo Algoritmo 1, ao fim de todas as iterações obtém-se o valor final da centralidade de proximidade por múltiplos caminhos disjuntos para cada nó, $C[s]$. A função **ACCUMULATE** recebe como entrada o valor de φ , o número de nós **numNodes** e os caminhos disjuntos encontrados para v_s armazenados em **DJ[s]**. Para cada destino v_t , calcula-se o custo agregado dos caminhos encontrados, $\Delta_{s,t}$, com o auxílio da variável **aux**. Em seguida, acumula-se os custos agregados para alcançar cada destino v_t com o auxílio da variável **acc**. Por fim, após acumular todos os custos agregados para todos os nós destino, calcula-se a centralidade de intermediação para nó de origem v_s , armazenando o valor em $C[s]$.

4.4 Comparação entre os ranqueamentos das centralidades de proximidade

A Figura 4.1 apresenta um cenário com 10 nós e 16 enlaces para ilustrar como os caminhos são escolhidos. Considerando ainda esse cenário, os algoritmos da Seção 4.3 são utilizados para determinar o valor da centralidade proposta para cada um dos nós dessa topologia. A Tabela 4.1 apresenta os nós em ordem de classificação para a métrica proposta ($C_{\varphi=1}$), para a proximidade tradicional (C) e para

Algoritmo 3 FUNÇÃO ACCUMULATE DE s PARA TODOS OS NÓS

Entrada: φ , numNodes, DJ[s]**Saída:** C[s]

```
1: acc, aux,  $\Delta_{s,t} \leftarrow 0$ 
2: para  $t \leftarrow 1, \text{numNodes}$  faça
3:   se  $v_s \neq v_t$  então
4:     para  $n \leftarrow 0, \varphi$  faça
5:       aux  $\leftarrow$  aux +  $\frac{1}{\text{DJ}[s][t][n]}$ 
6:        $\Delta_{s,t} \leftarrow \frac{1}{\text{aux}}$ 
7:     acc  $\leftarrow$  acc +  $\Delta_{s,t}$ 
8: C[ $s$ ]  $\leftarrow \frac{\text{numNodes}-1}{\text{acc}}$ 
```

a centralidade de informação (I). Os valores obtidos para cada nó utilizando cada uma das centralidades aparecem entre parênteses.

Tabela 4.1: Valores das métricas de centralidade para o cenário da Figura 4.1

Posição	Nó (Centralidade)		
	C	I	$C_{\varphi=1}$
1	v_6 (0,64)	v_6 (0,18)	v_6 (1,07)
2	v_0 (0,6)	v_0 (0,17)	v_3 (1,05)
3	v_3 (0,6)	v_3 (0,17)	v_9 (1,02)
4	v_9 (0,6)	v_9 (0,17)	v_0 (1,0)
5	v_4 (0,56)	v_1 (0,15)	v_1 (0,95)
6	v_1 (0,53)	v_2 (0,15)	v_2 (0,95)
7	v_2 (0,53)	v_5 (0,15)	v_4 (0,93)
8	v_5 (0,53)	v_4 (0,14)	v_5 (0,9)
9	v_7 (0,5)	v_7 (0,12)	v_7 (0,85)
10	v_8 (0,5)	v_8 (0,11)	v_8 (0,82)

Observa-se na Tabela 4.1 que as posições podem ser divididas em 3 grupos, cada um destes grupos são compostos pelo mesmo conjunto de nós para todas as métricas da tabela. O primeiro grupo é composto pelos nós $\langle v_0, v_3, v_6, v_9 \rangle$, o segundo, pelos nós $\langle v_1, v_2, v_4, v_5 \rangle$, e o último, pelos nós $\langle v_7, v_8 \rangle$. Ao focar no primeiro grupo de nós, diferentemente da proximidade tradicional e da centralidade de informação, a métrica proposta é capaz de classificar os nós sem que ocorra empate entre eles, mesmo usando um fator de conectividade baixo ($\varphi = 1$). Isso é consequência de os caminhos disjuntos do nó v_3 para alcançar os demais nós terem menor custo que os caminhos disjuntos dos outros nós antes empatados com o mesmo valor de centralidade. Os únicos nós que empatam para a centralidade de proximidade por múltiplos caminhos disjuntos são os nós v_1 e v_2 , isso ocorre devido à simetria dos nós. O nó que mais perde posição quando se compara a métrica tradicional às demais métricas é o nó v_4 , devido ao acréscimo dos custos de seus caminhos disjuntos que são maiores que o acréscimo dos nós v_1 e v_2 , por exemplo. Em relação ao último grupo,

difícilmente o nó que esteja na última posição irá melhorar sua classificação, visto que, seus caminhos mais curtos já são maiores que os dos demais nós da rede. Então, a possibilidade de seus caminhos disjuntos serem menores em relação aos caminhos disjuntos dos outros nós da rede é pequena. Ainda assim, a métrica proposta é capaz de desempatar os nós v_7 e v_8 , empatados no ranqueamento da proximidade tradicional.

Capítulo 5

Conjuntos de Dados

Esta dissertação utiliza cinco conjuntos de dados com características distintas. Todos eles são estáticos, sendo que três deles representam topologias de redes de longa distância reais de educação e pesquisa [21] (RNP, Renater e Géant), enquanto dois deles representam topologias de redes sociais (Train Bombing Network e Netscience). Este capítulo descreve os conjuntos de dados e destaca as características mais importantes de cada um para a análise realizada nesta dissertação. A Figura 5.1 mostra uma representação destes conjuntos de dados, onde as cores dos nós está relacionada a centralidade de proximidade tradicional do nó, desta forma, quanto maior sua centralidade maior a intensidade da cor. O tamanho dos nós está relacionado ao seu grau, sendo assim, quanto maior o tamanho do nó maior o grau dele.

5.1 RNP

A Rede Ipê da RNP [22] está localizada no Brasil e foi inaugurada em 2005. A rede é uma infraestrutura avançada de serviços de redes que interconecta universidades, institutos de pesquisa e instituições culturais no Brasil. O objetivo dessa infraestrutura é dar suporte à pesquisa, à educação, ao teste e ao desenvolvimento de aplicações avançadas de redes, e ao desenvolvimento social e regional. A rede é um segmento da Internet que interliga atualmente 27 Pontos de Presença (PoP) nas capitais dos Estados e no Distrito Federal. A infraestrutura dessa rede vem evoluindo desde a sua criação e a versão da topologia utilizada [21] possui 33 enlaces que interconectam os 27 PoPs, conforme ilustrado na Figura 5.1(a). A rede da RNP é a que possui menor quantidade de nós dentre as redes usadas nesta dissertação. Além disso, é a que possui o menor grau médio dentre elas (2,4).

5.2 Renater

A rede Renater [23] é uma infraestrutura de referência para a comunidade de ensino e pesquisa na França. Foi inaugurada em 1993 com o objetivo de prover uma infraestrutura de rede base e serviços inovadores, de alta qualidade, interoperáveis e de redes móveis para a comunidade em qualquer ponto da França. Atualmente opera com 150 comprimentos de onda, oferecendo capacidade de 10 a 200 Gb/s e conecta 72 PoPs. A versão da topologia utilizada nesta dissertação é composta por 45 nós interligados por 61 enlaces, conforme ilustra a Figura 5.1(c). A rede Renater possui grau médio 2,7 e apresenta resultados interessantes devido à sua topologia. Esses resultados são discutidos no Capítulo 6.

5.3 Géant

A rede Géant [24] localiza-se na Europa e foi criada em 2000 com o intuito de fornecer uma infraestrutura de rede comum que permitisse a interconexão das diversas redes de pesquisa nacionais existentes nos países da Europa. Atualmente a rede Géant interconecta 45 redes nacionais de pesquisa europeias, incluindo a rede Renater. A versão de topologia utilizada nesta dissertação é composta por 42 nós e 68 enlaces. A rede Géant é a maior e mais complexa rede de pesquisa de longa distância existente [24]. As versões das topologias das redes Géant e Renater usadas nesta dissertação possuem uma quantidade de nós semelhante. No entanto, a rede Géant possui maior grau médio (3,2) do que a Renater e é a rede real de longa distância mais densa analisada nesta dissertação.

5.4 Train Bombing Network

A rede social Train Bombing Network [25] reproduz os contatos entre os suspeitos de terrorismo envolvidos no ataque à bomba coordenados ao sistema de trens de Madri em março de 2004. Nesse conjunto de dados, cada nó representa um terrorista envolvido direta ou indiretamente no ataque. A existência de uma aresta entre dois terroristas mostra a ocorrência de um contato entre eles. O peso de cada aresta está relacionado com a quantidade de contato entre os terroristas. No entanto, nesta dissertação consideram-se arestas com pesos unitários. Um total de 64 terroristas existe nessa rede e estão interligados através de 243 enlaces. O grau médio apresentado pela rede é igual a 7,6 e essa é a rede mais densa dentre todas dos conjuntos de dados utilizados. A Figura 5.1(d) mostra a topologia dessa rede.

5.5 Netscience

A rede social Netscience [26, 27] representa as relações de coautoria entre pesquisadores na área de redes com publicações até o início do ano de 2006. Nessa rede, cada nó é um cientista da área e os enlaces entre dois pesquisadores mostra que ocorreu coautoria entre os pesquisadores. A rede completa possui 1589 cientistas, mas apenas 379 participam da maior componente conexa. Nesta dissertação, utiliza-se apenas essa componente para avaliar a métrica proposta, pois neste momento o foco está apenas em componentes conexas. A Figura 5.1(e) apresenta esta componente, na qual os cientistas estão interligados através de 914 arestas e a rede apresenta grau médio igual a 4,8. Dentre todas as redes analisadas, essa é a que apresenta a menor densidade.

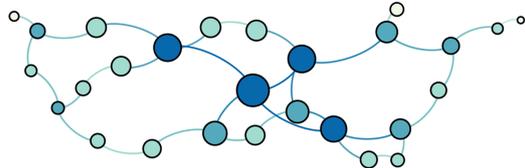
5.6 Sumário

A Tabela 5.1 sumariza as principais características dos conjunto de dados usados. São estas: o grau médio, o diâmetro e a densidade da rede. O grau de está relacionado ao número de enlaces incidentes em um nó, desta forma, o grau médio é a média do grau de cada nó do conjunto de dados. O diâmetro é a maior distância entre os menores caminhos de qualquer par de nós do conjunto de dados, já a densidade é a razão entre a quantidade de enlaces existentes e a quantidade de enlaces de um grafo completo com conjunto de dados. A rede Netscience possui a maior quantidade de nós e enlaces, porém a Train Bombing é mais densa e possui maior grau médio. A afirmação continua verdadeira quando essas redes são comparadas com as redes reais de longa distância, RNP, Renater e Géant. Dentre as redes de longa distância, a Renater possui a maior quantidade de nós, mas a Géant possui maior quantidade de enlaces e o maior grau médio. Apesar disso, a rede RNP é a que apresenta maior densidade dentre as redes de longa distância.

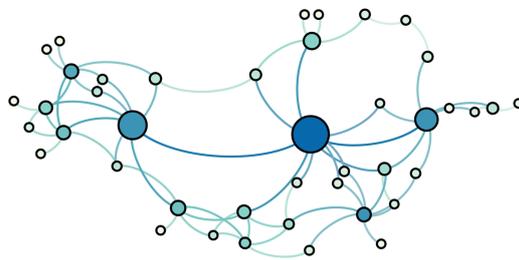
Tabela 5.1: Principais características dos conjunto de dados.

Topologias	\mathcal{V}	\mathcal{E}	Grau médio	Diâmetro	Densidade
RNP (Rede Ipê)	27	33	2,4	9	0,094
GEANT (Géant)	42	68	3,2	7	0,079
RENATER (Renater)	45	61	2,7	11	0,062
Train Bombing	64	243	7,6	6	0,121
Netscience	379	914	4,8	17	0,013

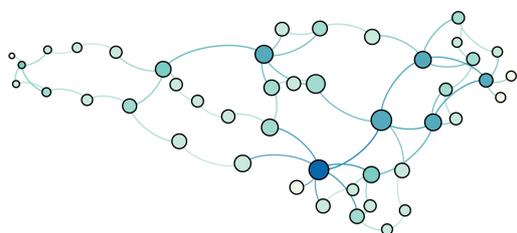
Utiliza-se os cinco conjuntos de dados apresentados neste capítulo para avaliar a métrica proposta desta dissertação (Capítulo 4). Os resultados obtidos são apresentados no Capítulo 6.



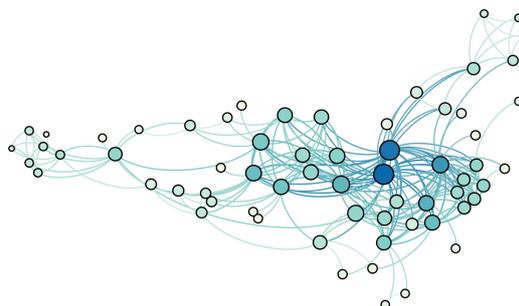
(a) RNP: 27 nós, 33 arestas, grau médio 2,4.



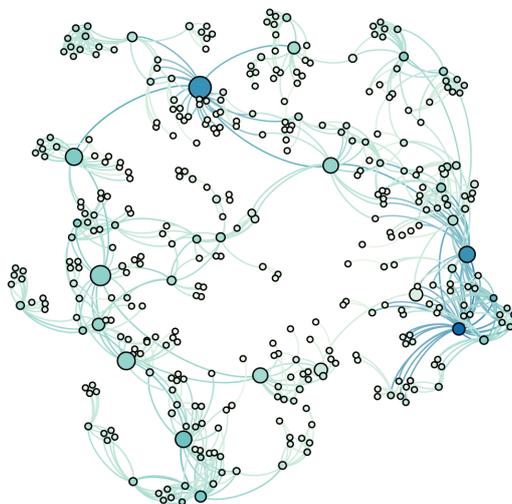
(b) GEANT: 42 nós, 68 arestas, grau médio 3,2.



(c) RENATER: 45 nós, 61 arestas, grau médio 2,7.



(d) Train Bomb: 64 nós, 243 arestas, grau médio 7,6.



(e) Netscience: 379 nós, 914 arestas, grau médio 4,8.

Figura 5.1: Topologias das redes (a) RNP, (b) GEANT, (c) RENATER, (d) Train Bomb e (e) Netscience. Quanto maior o nó, maior seu grau e quanto mais escuro o nó, maior sua proximidade.

Capítulo 6

Resultados e Discussões

Este capítulo tem como foco apresentar e discutir os resultados da avaliação da métrica proposta. O objetivo dessa avaliação é investigar a influência da utilização dos múltiplos caminhos disjuntos na classificação dos nós e na acessibilidade aos nós mais bem classificados em cada conjunto de dados utilizado.

6.1 Metodologia de avaliação

A métrica proposta utiliza tanto caminhos mais curtos quanto múltiplos caminhos disjuntos para determinar a importância de um nó. Os efeitos dessas modificações são avaliados de forma comparativa, utilizando a proximidade tradicional como base de comparação tanto para a métrica proposta, quanto para a centralidade de informação. As três métricas são aplicadas aos conjuntos de dados descritos no Capítulo 5 e analisa-se (i) a influência do fator de conectividade no ranqueamento dos nós, (ii) a correlação entre as métricas e (iii) a alcançabilidade dos nós em presença de falha. Os objetivos de cada análise são os seguintes:

- **Influência do fator de conectividade no ranqueamento:** investiga-se como o aumento do fator de conectividade φ afeta a quantidade de caminhos obtidos para cada conjunto de dados. Além disso, verifica-se como a adição desses caminhos modifica a classificação dos nós. Dessa forma, é possível determinar o valor máximo de φ que estabiliza o ranqueamento dos nós, ou seja, encontra-se qual é o maior valor de φ em cada cenário para o qual a classificação de todos os nós permanece inalterada;
- **Correlação entre as métricas:** utiliza-se o coeficiente de correlação W de Kendall para verificar o grau de concordância entre a classificação dos nós produzidas pelas três métricas avaliadas. Para complementar a análise, verifica-se também a frequência com que nós são reclassificados, tomando como base a proximidade tradicional. Essas avaliações permitem identificar se as

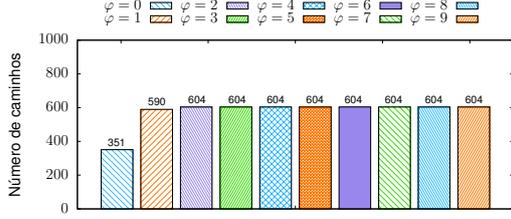
métricas julgam características semelhantes dos nós, ao mesmo tempo em que apontam nós que podem ser reclassificados de acordo com as suas próprias definições de importância para um nó;

- **Alcançabilidade dos nós em presença de falha:** avalia-se como a ocorrência de falhas influencia na alcançabilidade dos nós mais bem classificados segundo cada métrica. Essa avaliação é importante para identificar com quais nós a comunicação é potencialmente mais resiliente.

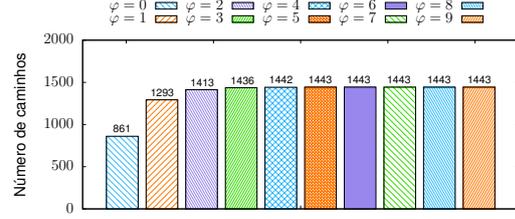
6.2 Influência do fator de conectividade no ranqueamento

O ranqueamento dos nós pela centralidade de proximidade por múltiplos caminhos disjuntos depende do valor do fator de conectividade φ usado, uma vez que φ está diretamente relacionado ao número de caminhos contabilizados pela métrica. Logo, é importante investigar como a variação na classificação dos nós é influenciada pelo fator de conectividade φ . A Figura 6.1 mostra o número total de caminhos entre todos os nós da rede contabilizados no cálculo da métrica proposta, para valores de φ no intervalo $[0, 9]$. É importante destacar que para $\varphi = 0$ não há busca por caminhos disjuntos e a métrica proposta torna-se equivalente à proximidade tradicional. Não se investiga valores para φ maiores do que 9 porque esse é o valor para o qual nenhum caminho disjunto adicional é encontrado no conjunto de dados Train Bombing, que apresenta a maior faixa de φ dentre todos os conjuntos de dados. Observa-se que quanto maior φ mais caminhos adicionais são encontrados. No entanto, o incremento no número de caminhos adicionais encontrados é cada vez menor à medida que φ cresce. Na rede da RNP, não existem caminhos adicionais a partir de $\varphi = 2$, conforme ilustrado na Figura 6.1(a). Na rede Train Bombing, por sua vez, é possível encontrar caminhos adicionais até $\varphi = 9$, como mostra a Figura 6.1(d). Para as demais redes, GEANT, RENATER e Netscience, os valores máximos de φ são, respectivamente, 5, 4 e 8. Na rede Netscience apesar de existir uma grande quantidade de caminhos alternativos encontrados, existem poucos caminhos a mais em $\varphi = 1$ do que em $\varphi = 0$, conforme ilustra a Figura 6.1(e). Esse comportamento é devido à baixa densidade da rede. Nota-se que o valor de φ para o qual novos caminhos não são mais encontrados é fortemente dependente da topologia da rede analisada. Além disso, o valor máximo de φ não parece ter relação com a densidade das redes, mas sim com o grau médio de cada uma.

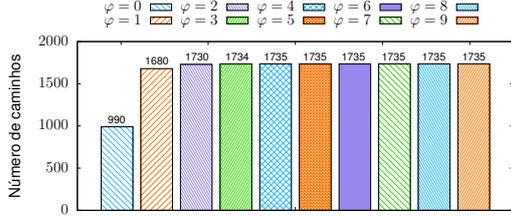
A Figura 6.2 mostra o mesmo aumento do número de caminhos ao se variar φ , mas de forma normalizada. A normalização é feita calculando a razão entre a quantidade de caminhos para cada φ e a maior quantidade de caminhos para cada



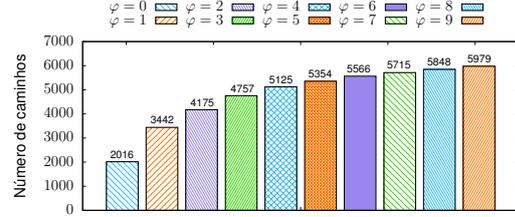
(a) RNP, grau médio 2,4 e densidade 0,094, $\varphi_{max} = 2$



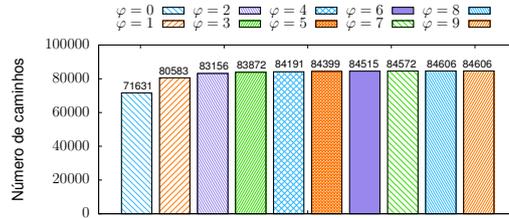
(b) GEANT, grau médio 3,2 e densidade 0,079, $\varphi_{max} = 5$



(c) RENATER, grau médio 2,7 e densidade 0,062, $\varphi_{max} = 4$.



(d) Train Bombing, grau médio 7,6 e densidade 0,121, $\varphi_{max} = 9$



(e) Netscience, grau médio 4,8 e densidade 0,013, $\varphi_{max} = 8$

Figura 6.1: 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. O valor máximo de φ para o qual caminhos não são mais encontrados é maior para redes com grau médio maior.

conjunto de dados. Na Figura 6.2(e) torna-se mais claro que a rede Netscience possui um incremento menor de caminhos quando comparada às outras redes, pois, número de caminhos disjuntos da rede é menor do quando comparado as outras redes, em $\varphi = 0$ já encontra-se 85% dos caminhos encontrados para este conjunto de dados. Em contrapartida, a Figura 6.2(d) mostra que a rede Train Bombing possui um maior incremento com o aumento de φ e possui apenas 34% dos caminhos para $\varphi = 0$. Os maiores incrementos são obtidos ao considerar o primeiro caminho disjunto ($\varphi = 1$) nas redes da RNP e RENATER, conforme mostram as Figuras 6.2(a) e 6.2(c), respectivamente.

O crescimento no número de caminhos devido ao aumento do fator de conectividade φ , pode resultar na reclassificação dos nós nas redes analisadas, promovendo ou rebaixando os nós no ranqueamento. Isso ocorre porque o valor da centralidade do

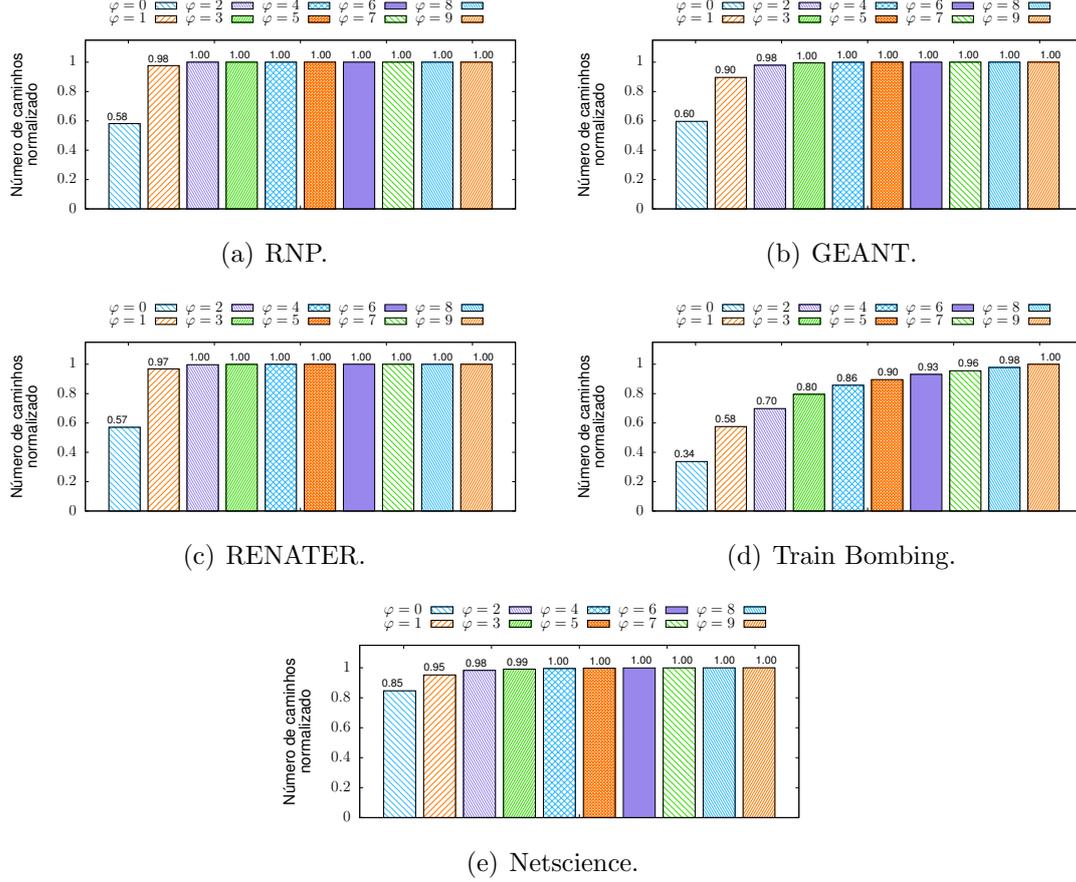
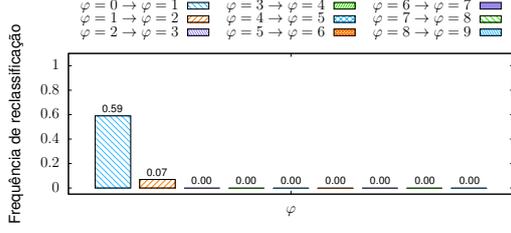
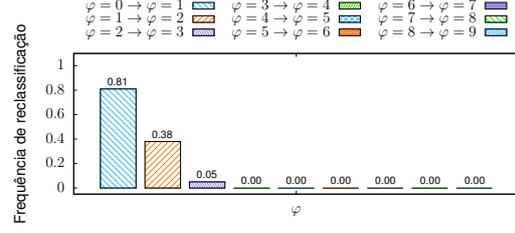


Figura 6.2: Total de caminhos normalizados entre todos os nós da rede usados para computar a métrica proposta. A normalização é em relação à maior quantidade de caminho computadas para cada topologia. A rede Netscience possui menor incremento de caminhos disjuntos, enquanto a rede Train Bombing possui a maior.

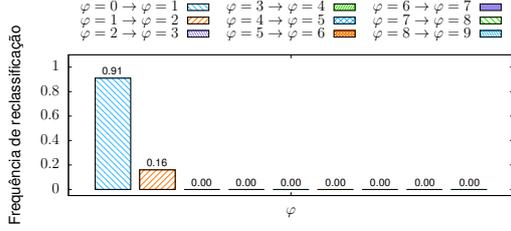
nó pode mudar de acordo com o número de caminhos contabilizados. A Figura 6.3 apresenta a frequência de reclassificação dos nós devido aos caminhos adicionais contabilizados. Essa frequência é obtida a partir da relação entre o número de nós que trocam de posição ao mudar o valor de φ e o número de nós existentes em cada cenário. Assim, obtém-se informação sobre a porcentagem de nós que são reclassificados quando o valor de φ é modificado. Observa-se na Figura 6.3(c) 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 $\varphi = 1$. A rede RENATER e a da RNP apresentam o maior incremento no número de caminhos adicionais ao aumentar φ para 1. No entanto, os caminhos adicionais contabilizados em cada uma das redes afetam a classificação dos nós com intensidade diferente. Na rede da RNP, 59% dos nós são reclassificados quando se aumenta φ para 1, conforme ilustrado na Figura 6.3(a). Apesar de menor do que a frequência de reclassificação da RENATER e de todas as outras redes, 59% ainda é um valor significativamente alto quando se considera que apenas 1 caminho disjunto



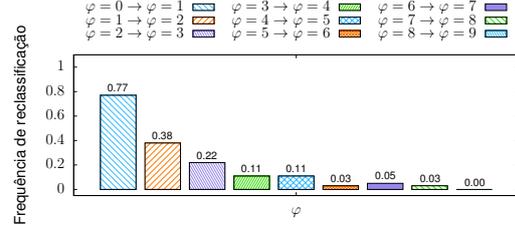
(a) RNP, grau médio 2,4 e densidade 0,094, $\varphi_{max} = 2$.



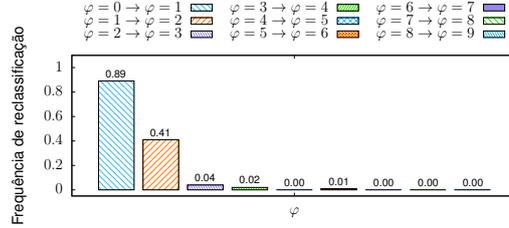
(b) GEANT, grau médio 3,2 e densidade 0,079, $\varphi_{max} = 3$.



(c) RENATER, grau médio 2,7 e densidade 0,062, $\varphi_{max} = 2$.



(d) Train Bombing, grau médio 7,6 e densidade 0,121, $\varphi_{max} = 8$.



(e) Netscience, grau médio 4,8 e densidade 0,013, $\varphi_{max} = 6$.

Figura 6.3: Frequência de reclassificação dos nós de acordo com a variação do fator de conectividade φ . A rede RENATER apresenta a maior frequência de reclassificação quando se começa a considerar os caminhos disjuntos ($\varphi = 0 \rightarrow \varphi = 1$), enquanto a rede RNP apresenta a menor frequência. A rede Train Bombing possui o decaimento da frequência de reclassificação mais suave do que o das demais redes, e estabiliza em $\varphi = 8$.

adicional entre cada par origem-destino foi contabilizado para computar a métrica. O mais notório dessas reclassificações é o desempate de nós classificados na mesma posição segundo a proximidade tradicional. Por exemplo, os nós 22 e 10 da rede da RNP são classificados no segundo lugar quando se utiliza a proximidade tradicional, mas o nó 10 perde uma posição no ranqueamento obtido para a métrica proposta. De forma geral, os nós que ocupam as primeiras posições tendem a coincidir no ranqueamento das métricas de centralidade [2]. Apesar disso, dentre as reclassificações na rede RENATER, a primeira posição, ocupada pelo nó 24 no ranqueamento da proximidade tradicional ($\varphi = 0$), passa a ser ocupada pelo nó 37 no ranqueamento da métrica proposta.

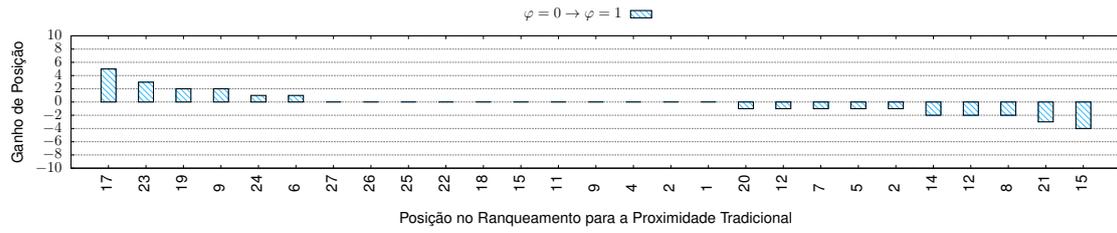
A Figura 6.3 mostra, ainda, o valor de φ para o qual cada ranqueamento torna-se estável, ou seja, nenhum nó muda de posição. Ambas as redes RENATER (Fi-

Figura 6.3(c)) e da RNP (Figura 6.3(a)) atingem essa estabilidade a partir de $\varphi = 2$. Logo, 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 nas Figuras 6.1(c) e 6.1(a), respectivamente. Os ranqueamentos nas redes GEANT (Figura 6.3(b)), Train Bombing (Figura 6.3(d)) e Netscience (Figura 6.3(e)) estabilizam para φ igual a 3, 8 e 6, respectivamente. Ainda, observa-se respectivamente nas Figuras 6.3(b) e 6.3(e), que as redes GEANT e Netscience possuem uma frequência de reclassificações muito pequena para $\varphi > 3$. A rede Train Bombing ilustrada na Figura 6.3(d) é a que possui a redução da frequência de reclassificação mais lenta dentre todas as topologias. Isso ocorre porque a proporção entre número de caminhos e número de nós, e também o grau médio da rede, é maior para essa rede. A Tabela 6.1 sumariza o valor máximo de φ para o qual ainda houve variação de posição na classificação dos nós de cada rede.

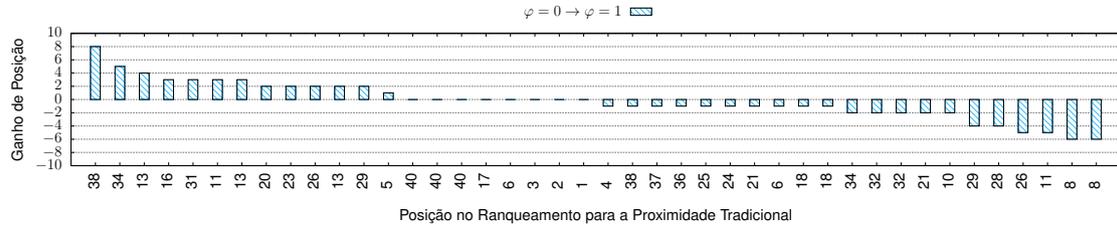
Tabela 6.1: Valores máximos de φ quando ocorre estabilização do ranqueamento, para cada conjunto de dados.

Topologia	RNP	GEANT	RENATER	Train Bombing	Netscience
φ_{max}	2	3	2	8	6

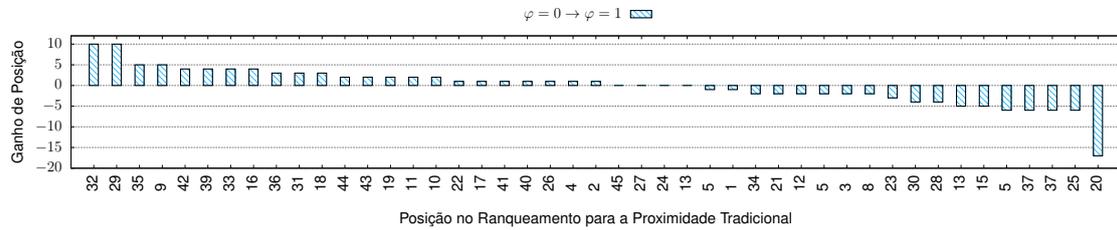
A maior frequência de reclassificação ocorre de $\varphi = 0$ para $\varphi = 1$ em todos os cenários, isto é, da proximidade tradicional (C) para $\varphi = 1$. A Figura 6.4 apresenta como esta reclassificação ocorre para cada posição dos nós, de forma que as variações positivas indicam ganho, e as negativas, perdas de posições para os nós que inicialmente ocupavam a posição descrita no eixo x . Para a rede Netscience, devido a quantidade de nós apresenta-se as 20 posições cujos nós ocupantes ganharam mais posições e as 20 posições cujos nós ocupantes perderam mais posições. É possível observar que para todas os conjuntos de dados os nós mais bem classificados sofrem pequena ou nenhuma variação, e que a maior variação tende a ser nos nós que pertencem às posições finais e centrais do ranqueamento. A Figura 6.4(a) mostra que, para a rede da RNP, o maior ganho de posição ocorre para o nó ocupante da 17^a, que passa a ocupar a 12^a posição, sendo promovido de 5 posições. A maior perda nessa rede é de 4 posições, o 15^o nó é rebaixado para a 19^a posição. A Figura 6.4(b) mostra que para a rede GEANT o maior ganho foi de 8 posições e a maior perda de 6 posições, essas variações ocorrem para os nós que ocupam posições 38 e 8 para $\varphi = 0$, respectivamente. A Figura 6.4(c) mostra a variação para a rede RENATER, sendo essa a rede que possui mais reclassificações dentre os nós mais bem classificados, destacando-se a variação da primeira posição, como já mencionado. Ainda assim, as maiores variações ocorrem para nós nas posições mais intermediárias do ranqueamento. Destaca-se o nó que ocupava a posição 20 e perde 17 posições. Ao analisar com mais detalhes trata-se de um nó que não



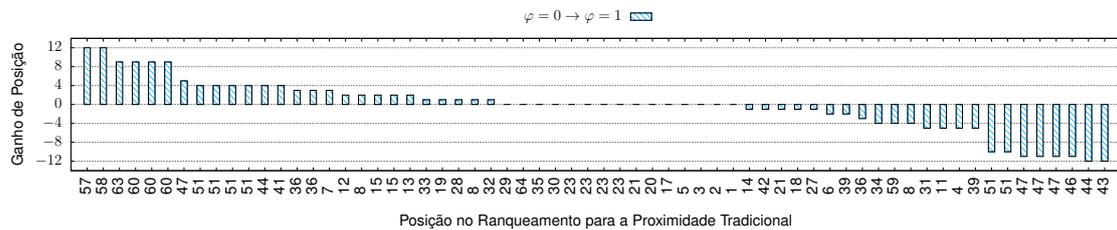
(a) RNP, total de nós 27.



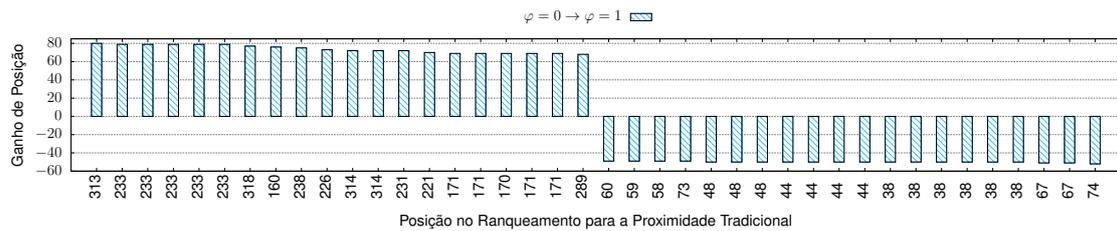
(b) GEANT, total de nós 42.



(c) RENATER, total de nós 45.



(d) Train Bombing, total de nós 64.



(e) Netscience, total de nós 379.

Figura 6.4: Variação da Reclassificação de $\varphi = 0$ para $\varphi = 1$ por nó de cada topologia. As maiores variações ocorrem para nós que ocupam as posições mais centrais da classificação. As variações para nós mais bem classificados são menores mas bastante significativas. A RENATER é a única topologia analisada a variar a 1ª posição da classificação.

possui caminhos disjuntos na rede e, por isso, é rebaixado de muitas posições. A Figura 6.4(d) mostra o comportamento para a topologia Train Bombing. Nessa rede,

a maior perda e o maior ganho são de 12 posições. Observando a Figura 6.4(e), o maior ganho, de 80 posições, ocorre na parte central do ranqueamento dos nós da rede. Nessa rede, nós a partir da 38ª posição já sofrem rebaixamentos intensos, perdendo mais de 40 posições. Além disso, devido à formação topológica da rede, o empate na classificação ainda é frequente.

6.3 Correlação entre as métricas

A fim de investigar a correlação entre a centralidade de proximidade por múltiplos caminhos disjuntos, e as outras métricas de proximidade, utiliza-se o coeficiente de correlação W de Kendall. O coeficiente de Kendall avalia a concordância entre diferentes classificações. Use o coeficiente de concordância de Kendall (Coef) para avaliar a associação entre avaliadores quando as classificações forem ordinais e você tiver três ou mais níveis de classificação.

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 [28]. Através desse coeficiente é possível identificar o grau de concordância entre os ranqueamentos produzidos pelas métricas. A Figura 6.5 mostra o valor do coeficiente W de Kendall para os máximos valores de φ de cada conjuntos de dados, conforme sumário da Tabela 6.1. Quanto mais próxima da borda do gráfico de radar estiver a área preenchida, maior é o grau de concordância entre as métricas comparadas. Todos os radares apresentam a correlação até $\varphi = 8$ para permitir melhor comparação visual entre as diferentes topologias.

Observa-se na Figura 6.5, que a centralidade de informação possui a menor correlação com as demais métricas para todas as topologias, exceto na rede Train Bombing (Figura 6.5(d)) que é a mais densa e que possui maior grau médio dentre os conjuntos de dados analisados. Além disso, grande parte de todos os caminhos existentes entre os nós são utilizados considerando maiores valores de φ . Para a rede Netscience, Figura 6.5(e), a correlação com a centralidade de informação é muito pequena, devido à baixa densidade da rede. É seguro afirmar, então, que a correlação entre a métrica proposta e a centralidade de informação tende a ser maior quanto maior for a densidade da rede. É interessante notar na Figura 6.5(a) que a rede da 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 da RNP possuir poucos caminhos disjuntos. Para as demais redes observa-se

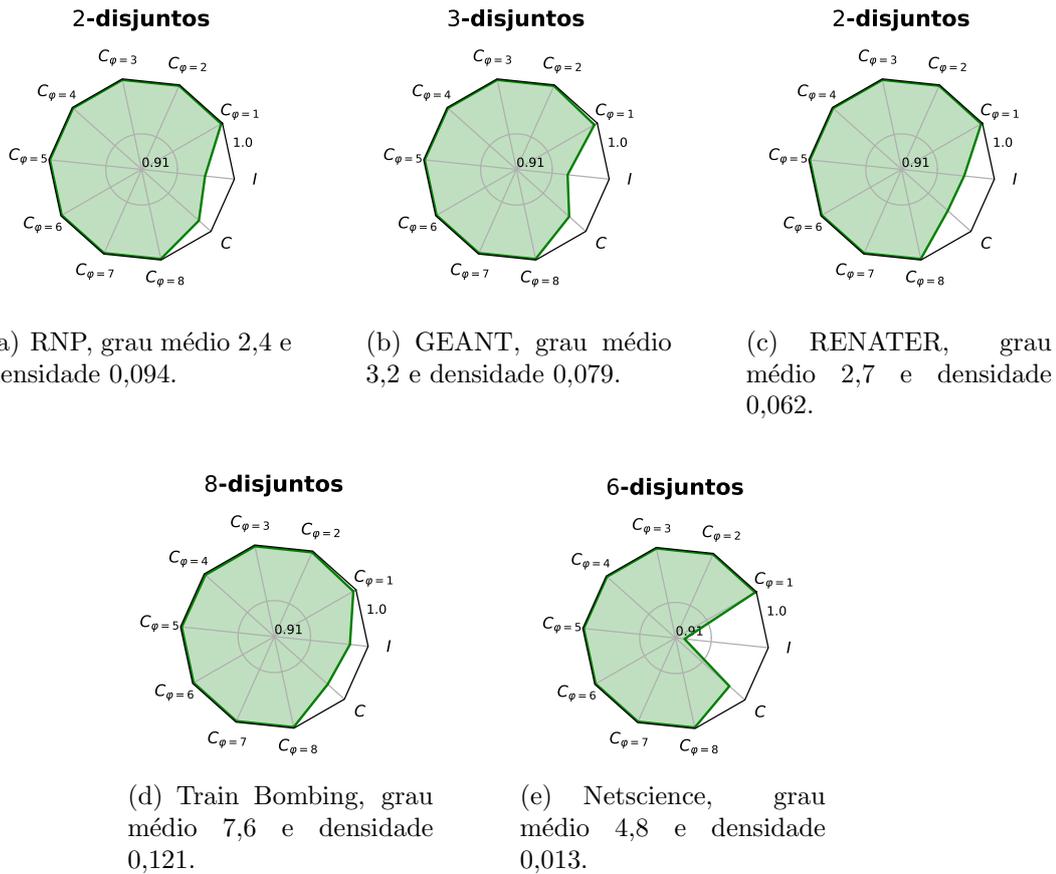


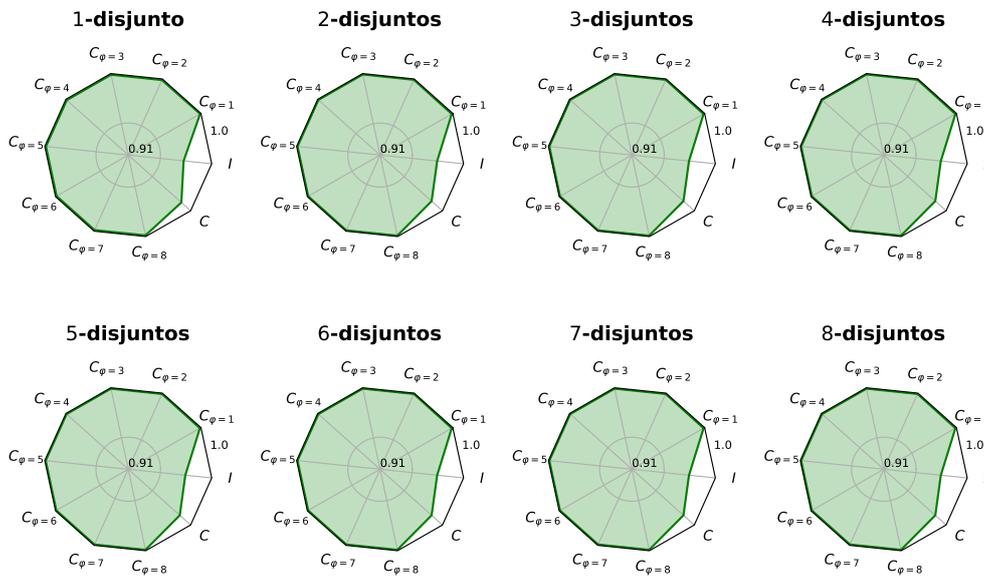
Figura 6.5: O coeficiente de correlação W de Kendall mostra o grau de concordância entre os ranqueamentos produzidos pelas métricas. A centralidade de informação possui a menor correlação com a métrica proposta para a maioria das topologias analisadas.

que a correlação entre a proximidade por múltiplos caminhos disjuntos e a tradicional é menor do que na rede RNP, porém a correlação ainda é elevada. É notório através dessa análise que, por um lado, em redes densas com valores elevados para φ , a métrica proposta tende à centralidade de informação, porque muitos dos caminhos existentes acabam sendo usados para computar a métrica proposta. Por outro lado, em redes menos densas, a métrica proposta tende à centralidade de proximidade tradicional, porque a quantidade de caminhos disjuntos adicionais contabilizados para computar a métrica é muito pequena.

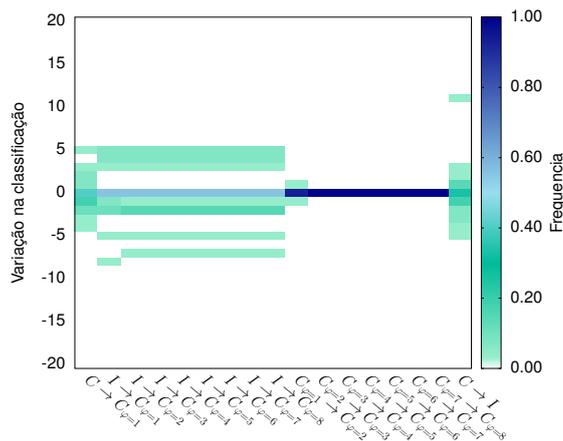
Apesar de o grau de concordância entre a centralidade de proximidade por múltiplos caminhos disjuntos e as demais métricas ser elevado, inúmeros nós são reclassificados ao se utilizar a métrica proposta. Nesse sentido, apresenta-se as Figuras 6.6, 6.7, 6.8, 6.9 e 6.10 que mostram o coeficiente W de Kendall entre as métricas e a frequência de variação na classificação dos nós ao trocar de uma métrica para outra, para reforçar o estudo da correlação entre as métricas. As Figuras 6.6(a), 6.7(a), 6.8(a), 6.9(a) e 6.10(a) mostram o coeficiente W de Kendall para todos os

valores de $\varphi \leq 9$ para melhor comparação visual entre os conjuntos de dados. As Figuras 6.6(b), 6.7(b), 6.8(b), 6.9(b) e 6.10(b) mostram a frequência de variação na classificação, sendo que o eixo- y representa o número de posições que os nós ganham ou perdem ao se utilizar cada métrica, a escala de cor indica o percentual de nós que sofreram uma mudança de y posições ao se trocar a métrica classificatória e o eixo- x define a 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, enquanto valores negativos indicam o oposto.

A correlação entre as métricas para a rede da RNP, na Figura 6.6, mostra que



(a) Coeficiente W de Kendall.

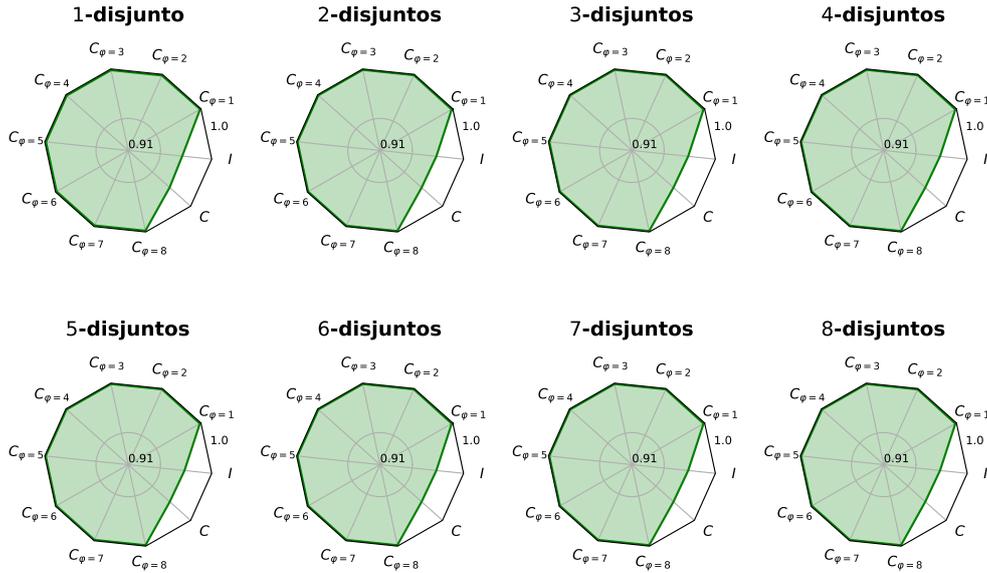


(b) Frequência de variação na classificação.

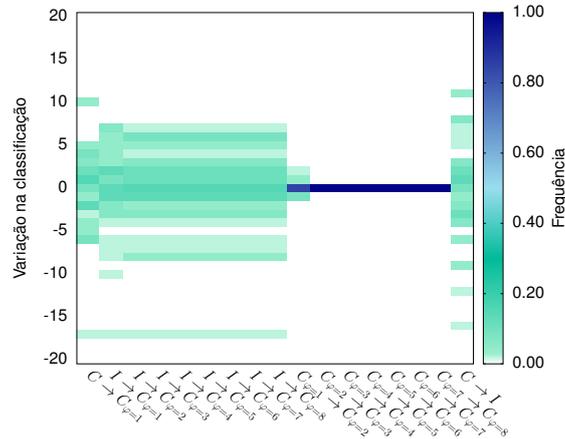
Figura 6.6: RNP. As métricas apresentam alta correlação, ainda assim, nota-se que ocorre reclassificação dos nós em maior ou menor amplitude. A rede da RNP é a que apresenta maior correlação entre a proximidade por múltiplos caminhos disjuntos e a tradicional, mesmo assim nós chegam a ganhar 5 posições.

apesar de apresentar alta correlação com a métrica tradicional, diversos nós perdem ou ganham várias posições quando são avaliados sob a visão da métrica proposta. O mesmo é válido quando ocorre a transição da centralidade de informação para a métrica proposta. Observa-se na Figura 6.6(b) que ao se utilizar a proximidade por múltiplos caminhos disjuntos com $\varphi = 1$ em vez da tradicional, alguns nós ganham até 5 posições. Ao se aumentar φ de 1 para 2, poucos nós 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 nas Figuras 6.1 e 6.2. Além disso, em relação aos ranqueamentos produzidos pela proximidade por múltiplos caminhos disjuntos e pela centralidade de informação, observa-se que 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 para $\varphi > 2$ ($I \rightarrow \varphi = 2$, $I \rightarrow \varphi = 3$, ..., $I \rightarrow \varphi = 8$).

Apesar da Figura 6.7(a) mostrar grande semelhança para todos os radares de φ para a rede RENATER, a Figura 6.7(b) mostra elevada frequência de variação na classificação dos nós da rede. Nota-se que o ranqueamento dos nós se estabiliza para $\varphi = 2$, pelo mesmo motivo da rede da RNP. A transição da proximidade tradicional para a por múltiplos caminhos disjuntos com $\varphi = 1$ mostra que menos nós permanecem na mesma posição quando comparado à rede da RNP. Alguns nós chegam a ganhar 10 posições e perder mais de 15. A correlação dos ranqueamentos produzidos pela proximidade por múltiplos caminhos disjuntos e pela centralidade de informação possui comportamento semelhante ao observado na rede da RNP. Já para rede GEANT é possível notar na Figura 6.8(a), que a correlação entre $\varphi = 1$ e os demais valores de φ é levemente menor. Além disso, a correlação entre a proximidade tradicional e a por múltiplos caminhos disjuntos com $\varphi = 1$ é maior do que para $\varphi > 3$. Para corroborar esse resultado a Figura 6.8(b) mostra que com $\varphi = 3$ a classificação dos nós torna-se estável, que para φ de 1 para 2 alguns nós ganham até 8 posições, e que alguns nós ganham até 5 posições quando φ muda de 2 para 3. Na rede GEANT, Figura 6.8(b), a amplitude da variação de posições é reduzida quando a centralidade de informação deixa de ser utilizada em detrimento da 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 perdem 11 posições para $\varphi = 2$ ($I \rightarrow \varphi = 2$) passa a perder apenas 10 posições. Para valores maiores de φ , não são encontrados novos caminhos disjuntos, de forma que a variação



(a) Coeficiente W de Kendall.

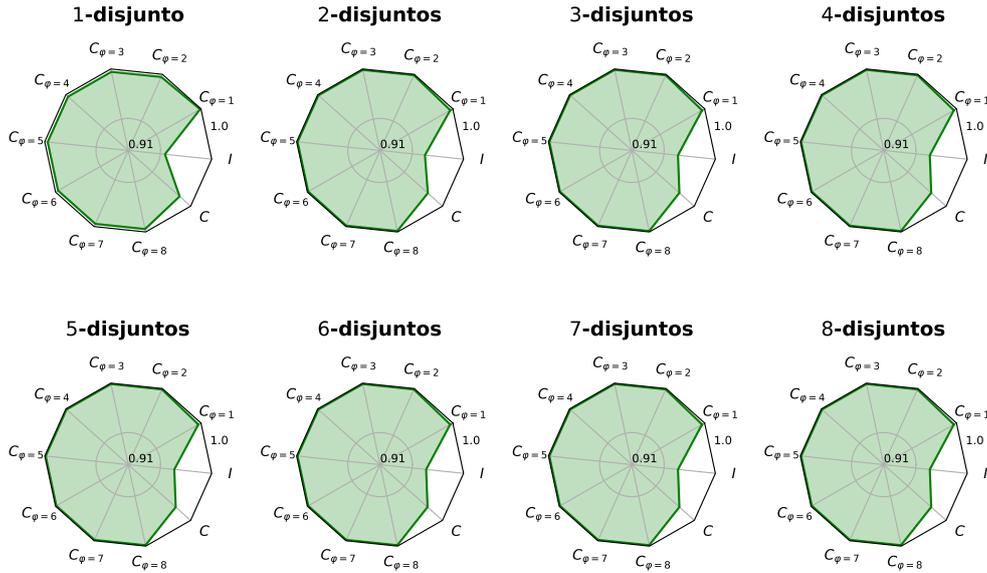


(b) Frequência de variação na classificação.

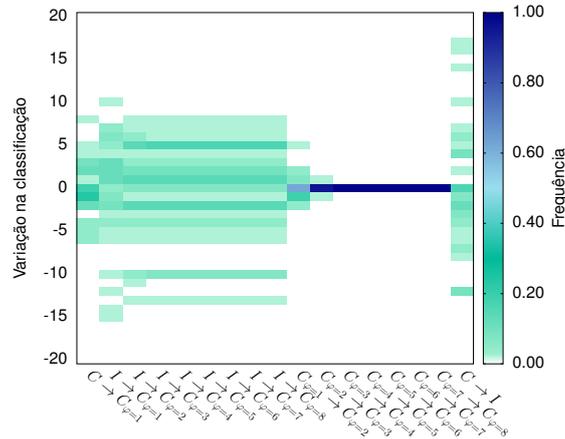
Figura 6.7: RENATER. As métricas apresentam alta correlação, ainda assim, nota-se que ocorre reclassificação nos nós em maior ou menor amplitude. A correlação entre a proximidade por múltiplos caminhos disjuntos e a tradicional ou de informação são próximas. Em ambos os casos nós perdemos até 17 posições.

no ranqueamento se estabiliza.

A Figura 6.9 apresenta a correlação para a rede social Train Bombing. Na Figura 6.9(a) é possível notar que há uma correlação menor entre $\varphi = 1$ e os demais valores, além disso, essa é a topologia que apresenta maior correlação entre a métrica proposta e a de informação. Apesar dessa alta correlação, cerca de 14% dos nós, apenas, permanecem na mesma posição e um nó chega a ganhar 17 posições. Em relação à tradicional, os nós chegam a ganhar 12 posições. Em relação ao aumento dos valores de φ , de 1 para 2 existem nós que chegam a perder 7 posições, e apesar da alta correlação, ao aumentar φ de 7 para 8 ainda ocorre uma pequena variação



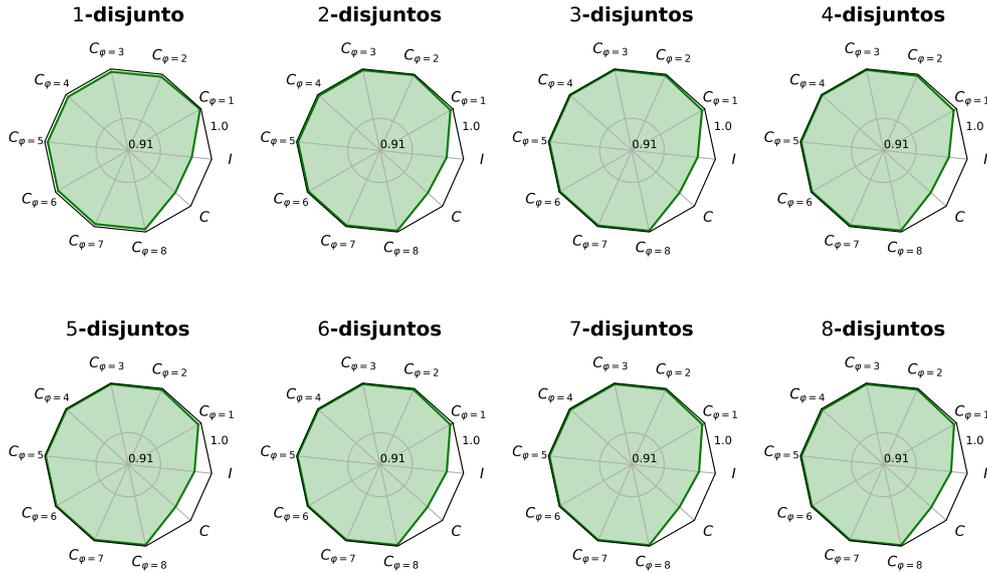
(a) Coeficiente W de Kendall.



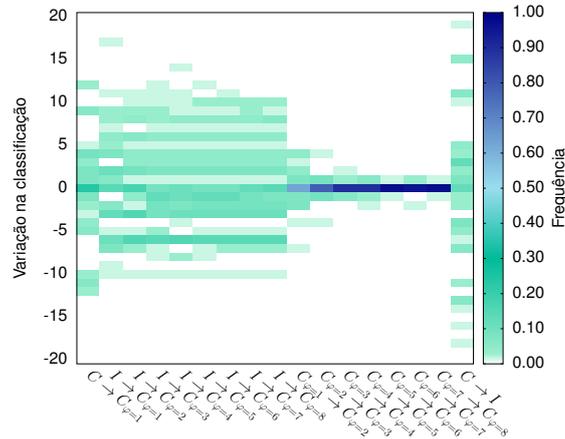
(b) Frequência de variação na classificação.

Figura 6.8: GEANT. As métricas apresentam alta correlação, ainda assim, nota-se que ocorre reclassificação dos nós em maior ou menor amplitude. É possível notar a redução da correlação entre a métrica tradicional e a proposta ao aumentar φ de 1 para 3.

na classificação. Para a rede social Netscience, a Figura 6.10(a) mostra que essa é a rede que possui menor correlação com a centralidade de informação. Isso se deve a sua topologia, e pode ser confirmada ao observar a dispersão na Figura 6.10(b) para todos os valores de φ em relação à centralidade de informação. Cerca de 2% dos nós, apenas, se mantêm na mesma posição, enquanto alguns nós chegam a ganhar 177 posições. As variações que ocorrem ao se aumentar os valores de φ são muito menores, comparando-se com as variações em relação à centralidade de informação. Ainda assim, existe uma pequena variação de apenas 1% quando se aumenta φ de 5 para 6. A partir de $\varphi = 6$ não ocorre mais variação na posição dos nós.



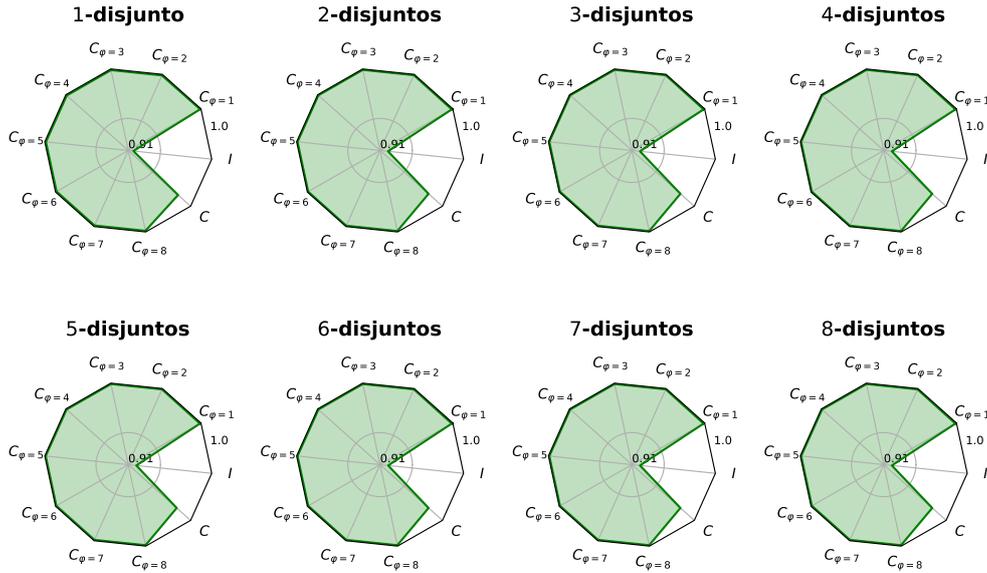
(a) Coeficiente W de Kendall.



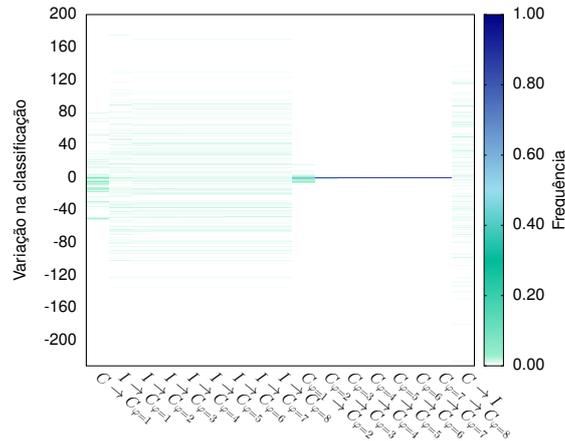
(b) Frequência de variação na classificação.

Figura 6.9: Train Bombing. As métricas apresentam alta correlação, ainda assim, nota-se que ocorre reclassificação so nós em maior ou menor amplitude. A rede GEANT maior correlação entre a métrica proposta e de informação, mesmos assim nós chegamos a ganhar 17 posições.

Os resultados apresentados nesta seção corroboram os resultados da Seção 6.2, observados nas Figuras 6.1, 6.3 e 6.4. Os resultados 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 Netscience, Figura 6.10(b), essa amplitude é maior, chegando a 419, com nós perdendo 194 posições e nós ganhando 225 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



(a) Coeficiente W de Kendall.



(b) Frequência de variação na classificação.

Figura 6.10: Netscience. As métricas apresentam alta correlação, ainda assim, nota-se que ocorre reclassificação os nós em maior ou menor amplitude. Esta rede possui menor correlação entre a proximidade por múltiplos caminhos disjuntos e de informação. a variação de classificação entre as métricas possui maior dispersão e uma quantidade menor de nós permanecem na mesma posição.

os caminhos pela centralidade de informação para computar a centralidade do nó.

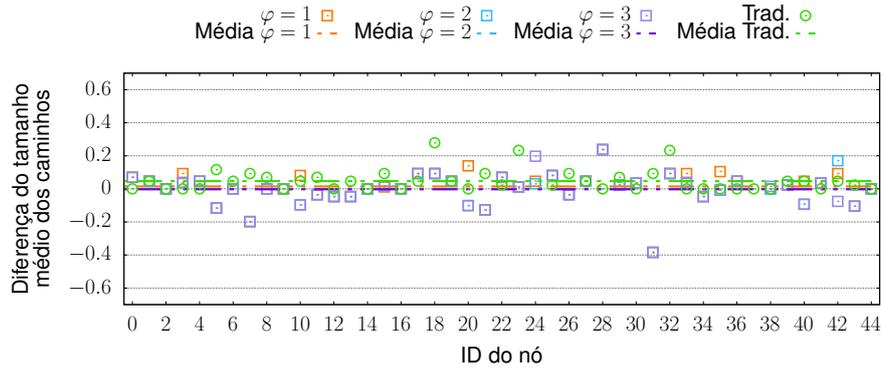
6.4 Alcançabilidade dos nós em presença de falha

Apesar de apresentar elevada correlação com a proximidade tradicional, a métrica proposta tem como objetivo destacar 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 re-

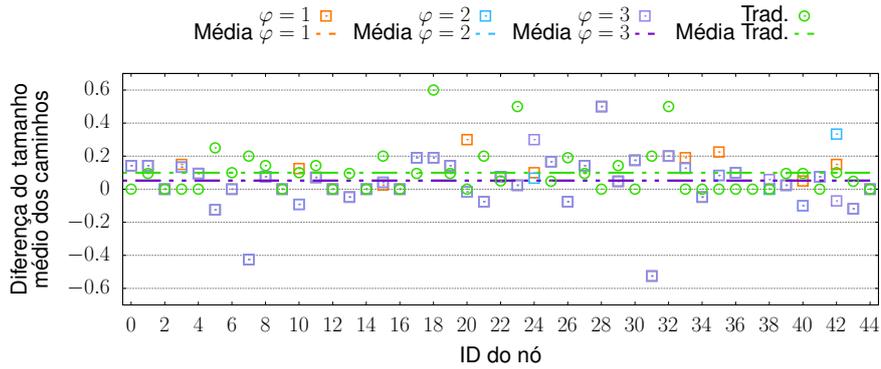
classificados, investiga-se o ranqueamento dos nós em redes propensas a falhas. O objetivo é estudar como a alcançabilidade dos nós varia quando se considera que eles podem ser acessados por caminhos mais curtos e múltiplos caminhos disjuntos. Para tanto, inicialmente investiga-se o custo para alcançar o nó mais bem posicionado segundo a métrica tradicional e a métrica proposta na presença de falha única na rede RENATER. Essa rede é utilizada para essa análise porque é a única em que o nó da primeira posição muda. Considera-se que a falha única pode ocorrer (i) em qualquer nó da rede ou (ii) apenas em nós que pertencem aos caminhos mais curtos para alcançar o nó mais bem classificado. A escolha na falha dos nós pertencentes aos caminhos mais curtos é consequência da propensão à falha nesses nós devido à concentração de informação neles quando apenas caminhos mais curtos são considerados. Por esse motivo, realiza-se também uma segunda análise, envolvendo múltiplas falhas dos nós em caminhos mais curtos. Para tanto, considera-se um grupo de nós mais bem classificados em cada conjunto de dados analisado nesta dissertação.

6.4.1 Cenário 1: acessibilidade do nó mais central da rede RENATER em presença de falha única

Investiga-se a alcançabilidade do nó mais central por todos os outros nós da rede quando ocorre uma falha única, aleatória, em quaisquer desses nós. Para tanto, falha-se cada um dos nós da rede. Em seguida, é avaliado se todos os nós restantes ainda alcançam o nó mais central e quão maior se tornou o custo para alcançar esse nó. Assim, primeiramente calcula-se para cada nó o tamanho médio dos caminhos para alcançar o nó mais central segundo cada métrica. Calcula-se, também, o tamanho médio dos caminhos quando a rede está em pleno funcionamento. Depois, verifica-se a diferença entre os tamanhos médios dos caminhos antes e depois das falhas. A Figura 6.11 mostra o resultado obtido para a rede RENATER. Nos demais conjuntos de dados, o nó identificado como o mais bem classificado é o mesmo, tanto pela proximidade tradicional quanto pela métrica proposta, de forma que os resultados sobre alcançabilidade obtidos para essas redes, entre os nós da rede na presença de falha, não acrescenta informação. Na Figura 6.11, o eixo- x identifica cada nó da rede, enquanto as linhas horizontais apresentam as médias das diferenças dos tamanhos médios dos caminhos considerando todos os nós, antes e depois da falha. Para a proximidade tradicional, os caminhos usados são apenas os mais curtos, enquanto para a proximidade por múltiplos caminhos disjuntos são usados todos os caminhos disjuntos que atendam ao fator 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



(a) Falha em nós aleatórios.



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

Figura 6.11: 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. Tanto para (a) falhas aleatórias, quanto para (b) falhas em nós dos caminhos mais curtos, valores próximos a $y = 0$ pelo lado positivo indicam menor aumento do custo médio para alcançar o nó mais central.

é o aumento no custo médio. Assim, quanto mais próximo de $y = 0$ pelo lado positivo, menor é o aumento no custo médio para acessar o nó mais central. Por outro lado, valores negativos são encontrados quando o tamanho médio dos caminhos após a falha é menor do que antes da falha. Isso pode ocorrer quando a falha aleatória provoca a perda de um caminho quase mais curto que anteriormente era utilizado.

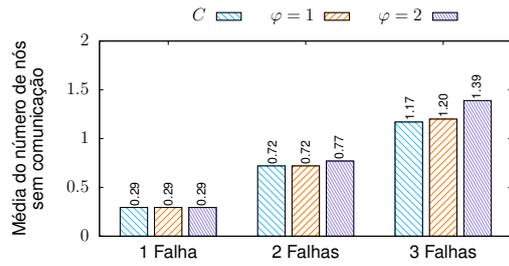
A Figura 6.11(a) mostra que, após a ocorrência de uma falha aleatória na rede, os demais nós da rede alcançam o nó classificado como mais central pela métrica proposta com aumento menor no custo médio dos caminhos usados para acessá-lo, quando comparado ao uso dos caminhos mais curtos apenas. Para a métrica proposta a média da diferença dos custos médios é de $-0,0023$ usando $\varphi = \{2, 3\}$, e para a proximidade tradicional a média é de $0,047$. Quando a falha ocorre apenas nos nós que pertencem a caminhos mais curtos, conforme Figura 6.11(b), a média das diferenças para a métrica proposta é de $0,051$, também usando $\varphi = \{2, 3\}$, e para a proximidade tradicional é de $0,099$. Apesar dos valores pequenos, decorrente da

pequena topologia avaliada, o custo médio para a métrica proposta fica mais próximo de 0 nos dois casos. Além disso, caso ocorram falhas em nós dos caminhos mais curtos, no pior caso 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.

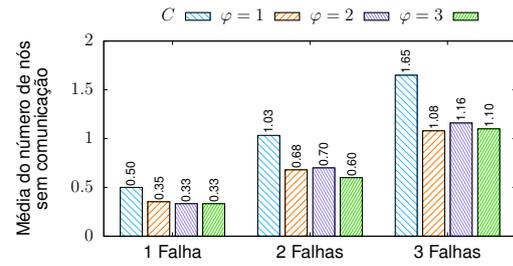
6.4.2 Cenário 2: acessibilidade dos nós mais centrais da rede RENATER em presença de múltiplas falhas

A fim de estender a análise de perda de comunicação com os nós mais bem posicionados da rede iniciada no final da seção anterior, investiga-se o impacto causado por múltiplas falhas na comunicação com um grupo de nós mais bem classificados por cada métrica. Para tanto, na ocorrência de falha de um, dois ou três nós, verifica-se a média do número de nós que perdem comunicação com 20% dos nós mais bem posicionados segundo cada métrica. Esse valor foi escolhido porque permite selecionar para analisar cada métrica um conjunto composto por nós diferentes. No caso de falha única, todas as possibilidades de falha são contabilizadas, isto é, todos os nós da rede são falhados, um de cada vez. No caso de duas ou três falhas sorteia-se um conjunto de 100 combinações de falhas dentre todas as possíveis combinações de falhas dos nós pertencentes aos caminhos mais curtos até os grupos de nós mais bem classificados. A Figura 6.12 apresenta os resultados encontrados para cada conjunto de dados analisado. O eixo- y representa o número médio de nós que perdem a comunicação com pelo menos um dos nós pertencentes ao grupo dos 20% mais bem classificados. O eixo- x por sua vez, indica o número de nós que falham. Os valores de φ considerados para cada topologia está relacionado à estabilização do ranqueamento e estão indicados na Tabela 6.1.

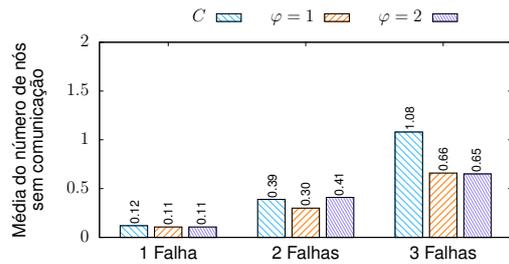
Em todos os cenários, os valores da média de nós sem comunicação aumentam com o aumento da quantidade de nós que falham na rede. A Figura 6.12(a) apresenta o resultado para a rede RNP. Essa é a única topologia para a qual o valor máximo de φ implica em maior média de número de nós sem comunicação com um dos nós do grupo composto pelos 20% nós mais bem classificados, quando ocorrem múltiplas falhas. Esse grupo é formado pelos 5 nós mais bem classificados. Para falha única, obtém-se os mesmos valores para $\varphi = 1$ e para a proximidade tradicional. Acima de três falhas os valores da média são maiores do que 1. Isso significa que para cada falha, pelo menos 1 dos nós da rede perde comunicação com um dos 5 mais centrais. Para a rede GEANT, o grupo de 20% dos nós é composto pelos 8 nós mais centrais de acordo com cada métrica. A Figura 6.12(b) mostra que a média do



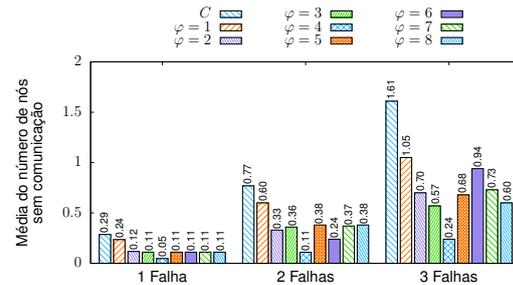
(a) RNP, grupo dos 20% nós mais bem classificados possui 5 nós.



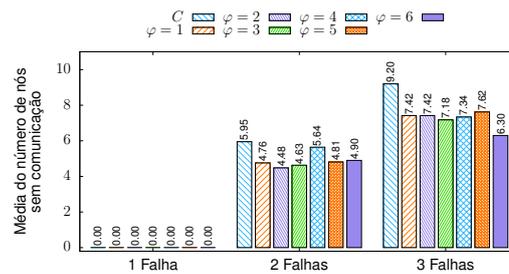
(b) GEANT, grupo dos 20% nós mais bem classificados possui 8 nós.



(c) RENATER, grupo dos 20% nós mais bem classificados possui 9 nós.



(d) Train Bombing, grupo dos 20% nós mais bem classificados possui 13 nós.



(e) Netscience, grupo dos 20% nós mais bem classificados possui 76 nós.

Figura 6.12: Média de nós que perdem a comunicação com o grupo formado pelos 20% nós mais bem classificados segundo a métrica tradicional e a proximidade de caminhos disjuntos. Considera-se φ de 1 até o valor máximo que mantém o ranqueamento estável para cada conjunto de dados. A métrica proposta apresenta melhores resultados para a maioria das topologia em quaisquer cenários de falha avaliados.

número de nós que perdem a comunicação com pelo menos um nó do grupo mais central é sempre menor para a métrica. Ou seja, ao se utilizar caminhos mais curtos e múltiplos caminhos disjuntos, menos nós deixam de se comunicar com os nós mais centrais. Nota-se que para três falhas ocorre um leve aumento da média quando se aumenta φ de 1 para 2 ou 3. Acredita-se que essa variação está relacionada aos conjuntos de amostras de falhas selecionados.

Para a rede RENATER o grupo de 20% corresponde aos 9 nós mais centrais da rede. A Figura 6.12(c) mostra que para uma falha, a diferença entre as métricas é insignificante. Para duas falhas as médias da quantidade de nós sem comunicação

ainda é muito semelhante para todas as métricas. No entanto, para três falhas, a comunicação com o grupo de nós mais bem classificado pela proximidade por múltiplos caminhos disjuntos é mantida para uma maior quantidade de nós da rede. Na rede Train Bombing, o acesso ao grupo mais bem classificado pela métrica proposta é o menos impactado pelas falhas, sejam elas únicas ou múltiplas. O grupo é composto por 13 nós. Esse resultado está ilustrado na Figura 6.12(d). Existem oscilações ao se aumentar os valores de φ , mas todas as médias da quantidade de nós sem comunicação são menores do que para a proximidade tradicional. Considerando todos os fatores de conectividade φ no caso de 3 falhas, o valor médio das médias do número de nós sem comunicação é igual a 0,69, enquanto para a métrica tradicional a média do número de nós sem comunicação é de 1,61. Para a rede Netscience o grupo de 20% dos nós mais bem conectados é composto por 76 nós. É interessante notar na Figura 6.12(e) que na ocorrência de uma falha, a média de nós sem comunicação é 0 independente da métrica usada para classificar os nós do grupo mais importante. No entanto, ao se considerar múltiplas falhas, a média da quantidade de nós sem comunicação apresenta um número muito mais expressivo para ambas as métricas, comparada às médias nos demais conjuntos de dados analisados. Isso se deve à topologia dessa rede. No geral, apesar de existirem muitos caminhos considerando todos os nós da rede, entre um par de nós qualquer existem poucos caminhos. Dessa forma, é maior a probabilidade de inviabilizar a comunicação entre os pares dos nós quando ocorrem falhas. Ainda assim, a proximidade por múltiplos caminhos disjuntos apresenta melhor resultado, chegando a apresentar aproximadamente 2 nós a menos sem comunicação quando comparada à proximidade tradicional.

Os resultados mostram que para os conjuntos de dados apresentados, a proximidade por múltiplos caminhos disjuntos elege nós que permanecem mais conectados aos demais nós da rede quando a informação flui pelos caminhos contabilizados pela métrica, principalmente em caso de múltiplas falhas em nós que pertençam aos caminhos mais curtos.

Capítulo 7

Conclusões

Esta dissertação estuda como a importância dos nós é afetada quando outros caminhos além dos mais curtos são levados em consideração. A ideia é utilizar múltiplos caminhos para quantificar a importância dos nós, de forma a eleger como nós mais centrais aqueles que são mais acessíveis por todos os outros nós da rede e com menor custo. Para tanto, propõe-se uma nova métrica de centralidade, a *centralidade de proximidade por múltiplos caminhos disjuntos*, que considera caminhos mais curtos e múltiplos caminhos vértice-disjuntos, privilegiando sempre os mais curtos, para computar a centralidade de um nó. A proposta de métrica tem como inspiração o dilema de escolher o nó da rede mais adequado para executar um papel central, como a instalação de um serviço que deve ser acessado por diversos nós, de forma a oferecer a maior rapidez de acesso para todos os nós e o menor risco de indisponibilidade. Aplicada a outras redes, como as redes sociais, a ideia adaptada é permitir a identificação de nós eficientes em transmitir mensagem por caminhos que não se interceptam. Isso é importante para garantir a informação chegue ao destino, mesmo que caminhos sejam eliminados.

O número de caminhos contabilizados para calcular a centralidade dos nós utilizando a métrica proposta é limitado pelo fator de conectividade φ . Esse fator é definido como o número *desejável* de caminhos totalmente disjuntos adicionais entre um par de nós v_s, v_t , sendo $\varphi \in \mathbb{N}$. Para um fator de conectividade $\varphi = 0$, a métrica proposta torna-se equivalente à proximidade tradicional. A avaliação da métrica proposta é feita de forma comparativa com duas outras métricas de centralidade de proximidade bem conhecidas, a proximidade tradicional, que considera apenas os caminhos mais curtos; e a centralidade de informação, que considera todos os caminhos existentes, priorizando os mais curtos. A avaliação é feita através da (i) investigação da influência do fator de conectividade no ranqueamento dos nós, da (ii) correlação entre as métricas e da (iii) alcançabilidade de nós centrais em presença de falha. As três métricas são aplicadas a 5 conjuntos de dados com características distintas, o que permite avaliar a métrica proposta em cenários diferentes. Os con-

juntos de dados usados são todos estáticos, sendo três deles representantes de redes de comunicação reais (RNP, RENATER e GEANT) e dois deles, de redes sociais (Train Bombing e Netscience).

Os resultados mostraram que, ao considerar múltiplos caminhos disjuntos, a métrica proposta é capaz de reclassificar de 59% a 91% dos nós usando um fator de conectividade $\varphi = 1$. Além disso, identificou-se que o valor máximo de φ para estabilizar a reclassificação dos nós está diretamente relacionado à topologia analisada. A tendência é que o fator de conectividade seja maior quando o grau médio da rede é maior. Ao avaliar a correlação da métrica proposta com a proximidade tradicional e a centralidade de informação, conclui-se que apesar da alta correlação, a métrica proposta foi capaz de reclassificar uma grande quantidade de nós melhorando ou piorando a posição deles de forma significativa. A rede da RNP, por exemplo, é o cenário em que ocorre a menor intensidade de reclassificação, variando de até 5 posições a classificação dos nós. A maior amplitude de variação na classificação ocorre ao se comparar a proximidade tradicional e a centralidade de informação. Dessa forma, a métrica proposta se apresenta como um meio termo entre essas duas.

A alcançabilidade de nós centrais em presença de falha é feita através da investigação de falhas únicas e falhas múltiplas. Para falha única, verifica-se quão acessível o nó mais central da rede está quando uma falha única aleatória acontece. Assim, apenas a rede RENATER foi avaliada, uma vez que essa é a única rede para a qual o nó mais central muda de acordo com a métrica utilizada. Para múltiplas falhas, todas as redes foram analisadas. Ao analisar a presença de falha única na rede RENATER, os resultados mostraram que a métrica proposta é capaz de identificar o nó cujos caminhos até ele são menos afetados. Além disso, observou-se que a comunicação com esse nó é perdida para uma menor quantidade de nós quando comparado à proximidade tradicional. Para investigar melhor a alcançabilidade dos nós, o cenário é estendido para múltiplas falhas e um conjunto de nós centrais. Assim, verificou-se como as múltiplas falhas afetam a comunicação com o grupo de 20% dos nós mais centrais de cada conjunto de dados. Os resultados mostraram que, em geral, um menor número de nós perde acesso a pelo menos um dos nós do grupo mais central classificado pela proximidade por múltiplos caminhos disjuntos. Isso é verdade quando se considera que a informação flui pelos caminhos importantes para cada métrica. A maior perda de comunicação ocorre para a rede Netscience, para ambas as métricas avaliadas. No entanto, em média 2 nós a mais perdem comunicação com algum nó do grupo mais central da proximidade tradicional, quando comparado ao grupo mais central da métrica proposta.

Como trabalhos futuros, pretende-se propor um mecanismo para encontrar um valor ótimo de φ para cada conjunto de dados a ser aplicado baseado nas caracte-

terísticas da rede. Além disso, pretende-se aprimorar o algoritmo que computa os caminhos disjuntos e, assim, tornar possível a análise de cenários de maior escala. Pretende-se também, ampliar as análises de falhas na rede e estender o estudo da métrica com aplicação em cenário cujos enlaces sejam ponderados. Essa extensão também será feita para cenários dinâmicos, a fim de identificar nós que possam ser mais acessíveis aos demais por um maior período de tempo. Por fim, pretende-se desenvolver uma aplicação baseada na proximidade por múltiplos caminhos disjuntos para tomada de decisão de alocação de funcionalidades, com o objetivo de tornar a rede mais resiliente a falhas.

Referências Bibliográficas

- [1] OLIVEIRA, A. K. M. D. *Estudo de casos de complexidade de colorações gulosa de vértices e de arestas*. Tese de Doutorado, Universidade Federal do Ceará, 2011.
- [2] FREEMAN, L. C. “Centrality in social networks conceptual clarification”, *Social Networks*, v. 1, n. 3, pp. 215–239, 1978.
- [3] BAVELAS, A. “A Mathematical Model for Group Structures”, *Human Organization*, v. 7, n. 3, pp. 16–30, 1948.
- [4] BOUET, M., LEGUAY, J., COMBE, T., et al. “Cost-based placement of vDPI functions in NFV infrastructures”, *International Journal of Network Management*, v. 25, n. 6, pp. 490–506, 2015.
- [5] MACCARI, L., CIGNO, R. L. “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*, pp. 1–9. IEEE, 2016.
- [6] STEPHENSON, K., ZELEN, M. “Rethinking centrality: Methods and examples”, *Social Networks*, v. 11, n. 1, pp. 1–37, 1989.
- [7] NEWMAN, M. J. “A measure of betweenness centrality based on random walks”, *Social Networks*, v. 27, n. 1, pp. 39–54, 2005.
- [8] MEDEIROS, D. S. V., CAMPISTA, M. E. M., MITTON, N., et al. “The Power of Quasi-Shortest Paths: ρ -Geodesic Betweenness Centrality”, *IEEE Transactions on Network Science and Engineering*, v. 4, n. 3, pp. 187–200, July 2017. doi: 10.1109/TNSE.2017.2708705.
- [9] MEDEIROS, D. S. V., CAMPISTA, M. E. M., MARCELO DIAS DE AMORIM, N. M., et al. “Eficiência dos Caminhos Quase Mais Curtos em Redes Dinâmicas”. In: *SBRC*, pp. 544–557, 2017.
- [10] BORGATTI, S. P., EVERETT, M. G. “A Graph-theoretic perspective on centrality”, *Social Networks*, v. 28, n. 4, pp. 466 – 484, 2006.

- [11] BRANDES, U., FLEISCHER, D. “Centrality Measures Based on Current Flow”. In: *STACS*, pp. 533–544, 2005.
- [12] BAVELAS, A. “Communication patterns in task-oriented groups”, *The Journal of the Acoustical Society of America*, v. 22, n. 6, pp. 725–730, 1950.
- [13] BEAUCHAMP, M. A. “An improved index of centrality”, *Behavioral science*, v. 10, n. 2, pp. 161–163, 1965.
- [14] BARBOSA, M. S. M., MEDEIROS, D. S. V., CAMPISTA, M. E. M. “Centralidade de Proximidade por Múltiplos Caminhos Disjuntos: Aplicação em Redes de Longa Distância”. In: *SBRC*, pp. 88–101, 2019.
- [15] AMARAL, P., PINTO, P. F., BERNARDO, L., et al. “SDN based traffic engineering without optimization: A centrality based approach”. In: *Communications (ICC), 2017 IEEE International Conference on*, pp. 1–7. IEEE, 2017.
- [16] HUI, L., TAO, L., XIANGLIN, H., et al. “Detection algorithm based on closeness rank and signal transimission”. In: *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 443–447. IEEE, 2017.
- [17] MARINA, M. K., DAS, S. R. “On-demand multipath distance vector routing in ad hoc networks”. In: *Proceedings Ninth International Conference on Network Protocols. ICNP 2001*, pp. 14–23. IEEE, 2001.
- [18] SIDHU, D., NAIR, R., ABDALLAH, S. “Finding disjoint paths in networks”, *ACM SIGCOMM Computer Communication Review*, v. 21, n. 4, pp. 43–51, 1991.
- [19] SUURBALLE, J. “Disjoint paths in a network”, *Networks*, v. 4, n. 2, pp. 125–145, 1974.
- [20] TAFT-PLOTKIN, N., BELLUR, B., OGIER, R. “Quality-of-service routing using maximally disjoint paths”. In: *1999 Seventh International Workshop on Quality of Service. IWQoS’99.(Cat. No. 98EX354)*, pp. 119–128. IEEE, 1999.
- [21] COUTO, R. S., SECCI, S., CAMPISTA, M. E. M., et al. “Latência Versus Sobrevivência no Projeto de Centros de Dados Geograficamente Distribuídos”, *XXXII SBRC*, pp. 809–822, 2014.
- [22] COMITÊ GESTOR RNP. *Rede Ipê: Política de Uso*. In: Report, Rede Nacional de Pesquisa, 2007.

- [23] SCHAFER, V. “Part of a whole: RENATER, a twenty-year-old network within the Internet”, *Information & Culture*, v. 50, n. 2, pp. 217–235, 2015.
- [24] *Géant: Transforming the way researchers collaborate*. In: Report, DANTE, 2007.
- [25] HAYES, B. “Connecting the Dots”, *American Scientist*, v. 94, n. 5, pp. 400–404, 2006.
- [26] NEWMAN, M. E. J. “Finding community structure in networks using the eigenvectors of matrices”, *Phys. Rev. E*, v. 74, pp. 036104, Sep 2006. doi: 10.1103/PhysRevE.74.036104. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevE.74.036104>>.
- [27] ROSSI, R. A., AHMED, N. K. “The Network Data Repository with Interactive Graph Analytics and Visualization”. In: *AAAI*, 2015. Disponível em: <<http://networkrepository.com>>.
- [28] CROUX, C., DEHON, C. “Influence functions of the Spearman and Kendall correlation measures”, *Statistical methods & applications*, v. 19, n. 4, pp. 497–515, 2010.