



Jonas Freire de Alcântara Marques de Barros

Uso de Técnicas de Detecção de Comunidades para Análise de Redes Ópticas

Recife

2021

Jonas Freire de Alcântara Marques de Barros

Uso de Técnicas de Detecção de Comunidades para Análise de Redes Ópticas

Monografia apresentada ao Curso de Bacharelado em Ciências da Computação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Ciências da Computação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Computação

Curso de Bacharelado em Ciências da Computação

Orientador: Danilo Ricardo Barbosa de Araújo

Recife

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

B277u Barros, Jonas Freire de Alcântara Marques de
Uso de Técnicas de Detecção de Comunidades para Análise de Redes Ópticas / Jonas Freire de
Alcântara Marques de Barros. - 2021.
59 f. : il.

Orientador: Danilo Ricardo Barbosa de Araujo.
Inclui referências.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco,
Bacharelado em Ciência da Computação, Recife, 2021.

1. Ciência de redes. 2. Redes complexas. 3. Topologias de rede. 4. Detecção de comunidades. I.
Araujo, Danilo Ricardo Barbosa de, orient. II. Título

CDD 004



**MINISTÉRIO DA EDUCAÇÃO E DO DESPORTO
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO (UFRPE)
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

<http://www.bcc.ufrpe.br>

FICHA DE APROVAÇÃO DO TRABALHO DE CONCLUSÃO DE CURSO

Trabalho defendido por Jonas Freire de Alcântara Marques de Barros às 16 horas do dia 09 de dezembro de 2021, no link <https://meet.google.com/ypd-ugyy-bss>, como requisito para conclusão do curso de Bacharelado em Ciência da Computação da Universidade Federal Rural de Pernambuco, intitulado “Uso de Técnicas de Detecção de Comunidades para Análise de Redes Ópticas”, orientado por Danilo Ricardo Barbosa de Araújo e aprovado pela seguinte banca examinadora:

Danilo Ricardo Barbosa de Araújo
DC/UFRPE

Pablo Azevedo Sampaio
DC/UFRPE

Assim, Ao Rei eterno, imortal, invisível, Deus único, honra e glória pelos séculos dos séculos. Amém!

(1 Timóteo 1:17)

Agradecimentos

Agradeço em primeiro lugar ao meu Senhor, que me salvou. E através de Sua providência tem me feito crescer em conhecimento e maturidade. A Ele seja dada toda a honra, glória e louvor.

Agradeço em segundo lugar aos meus pais, Edson Marques e Maria Ida, que sempre cuidaram de mim, e me instruíram no caminho que deveria andar, que oraram por mim. Que sempre estiveram presentes nas horas de angústia e de felicidade. Agradeço a eles por todo o investimento que fizeram e ainda fazem em mim. Se eu conseguir fazer pelos meus filhos metade do que eles fizeram por mim, certamente, serei um grande pai. Ao meu querido irmão Jasom, muito obrigado também pelos seus conselhos e repreensões que foram e tem sido tão importantes para meu amadurecimento.

Agradeço também a todos os professores que durante o decorrer do curso me instruíram e me deram muitos conhecimentos que diariamente me auxiliam no exercício da minha profissão. Em especial ao professor Danilo Araújo, meu orientador. Que tem me ajudado no processo de formação acadêmica e profissional, sua paciência é notável professor.

Por fim, agradeço também a todos que de alguma forma têm contribuído para a minha formação. Aos meus amigos e colegas que têm compartilhado seus conhecimentos, a todos os funcionários da Universidade Federal Rural de Pernambuco que têm feito desta, uma segunda casa. E também a todos que direta ou indiretamente me auxiliaram, obrigado.

*“Porque os meus pensamentos não são os vossos pensamentos, nem os vossos caminhos, os meus caminhos, diz o SENHOR.”
(Isaías 55:8)*

Resumo

O crescimento da utilização dos serviços na Internet tem promovido uma crescente demanda por elevadas taxas de transmissão. Essa demanda tem sido suprida principalmente por redes ópticas. Na etapa de projeto dessas redes, o engenheiro deve possuir a capacidade de avaliar o desempenho de uma determinada rede antes de sua efetiva implantação física. Nesse processo de projeto várias topologias são consideradas. A comparação entre topologias é feita através de métricas que indicam um determinado aspecto da rede. Normalmente as métricas consideradas são os indicadores de desempenho, como *Throughput*, Probabilidade de Bloqueio, Resiliência e também outros indicadores, como o Custo da rede. Os indicadores de desempenho são importantes, pois eles informam a qualidade de uma determinada topologia. Portanto, esses indicadores de desempenho são imprescindíveis para os projetos de *design* dessas redes. A forma mais confiável de calcular os valores desses indicadores de desempenho é através de simulações. Porém, as simulações apresentam um alto custo computacional, elevando o tempo necessário para obter informações acerca das topologias; visto que, nesses projetos um número muito grande de diferentes topologias precisa ser considerado. Por outro lado, um grande número de pesquisas dos mais diversos domínios do conhecimento vem sendo feitas na área de detecção de comunidades em grafos. Entretanto, não existem aplicações destas técnicas em redes de fibra óptica de alta capacidade. Desta forma, o presente trabalho tem por objetivo investigar a existência de correlação entre a capacidade que uma rede de fibra óptica possui de formar comunidades e os seus indicadores de desempenho. Mais especificamente, a sua Probabilidade de Bloqueio e indicadores de Resiliência. A análise foi realizada comparando a Probabilidade de Bloqueio e Resiliência dessas redes com as métricas de clusterização através de gráficos de dispersão. De acordo com os resultados, existe uma correlação positiva entre as métricas de comunidades e os indicadores de desempenho das redes e comparativamente foi obtido um *speedup* de aproximadamente 4500 vezes entre métricas de clusterização e as simulações.

Palavras-chave: Ciência de redes, Redes complexas, Topologias de rede, Clusterização, Detecção de comunidades, Redes de fibra óptica, Design de topologias físicas, Resiliência de redes, Probabilidade de bloqueio.

Abstract

The growth in the use of services on the Internet has promoted an increasing demand for high transmission rates. This demand has been met by optical networks. At the design stage of these networks, the engineer must be able to assess the performance of a given network before its actual physical implementation. In this design process, several topologies are considered. The comparison between topologies is made through metrics that indicate a certain aspect of the network. Typically the metrics considered are performance indicators, such as Throughput, Blocking Probability, Resilience and also other indicators, such as the network Cost. Performance indicators are important because they inform about the quality of a particular topology. Therefore, performance metrics are essential for the design projects of such networks. The most reliable way to calculate the values of these performance indicators is through simulations. However, simulations have a high computational cost, increasing the time needed to obtain information about topologies; since, in these projects, a very large number of different topologies must be considered. On the other hand, a large number of researches in the most diverse domains of knowledge have been carried out on the theme of community detection in graphs. However, there are no applications of these techniques in high capacity fiber-optic networks. Thus, the present work aims to investigate the existence of a correlation between the ability of a fiber-optic network to form communities and its performance indicators. More specifically, it's Blocking Probability and indicators of Resilience. The analysis was performed comparing the Blocking Probability and Resilience of these networks and the clustering metrics using scatter plots. According to the results, there is a positive correlation between the community metrics and the network performance indicators, and comparatively a speedup of approximately 4,500 times was obtained between the community metrics and the simulations.

Keywords: Network science, Complex networks, Network topologies, Clustering, Community detection, Optical fiber networks, Physical topology design, Network robustness, Blocking probability.

Lista de ilustrações

Figura 1 – CARNET (Croatian Academic and Research Network). Fonte: (TOPOLOGY-ZOO, b)	14
Figura 2 – Kentucky Datalink. Fonte: (TOPOLOGY-ZOO, c)	15
Figura 3 – Exemplo de Grafo. Fonte: Autor	18
Figura 4 – Exemplo de Comunidade. Fonte: Autor	19
Figura 5 – Exemplo de atualização de rótulo. Fonte: Autor	20
Figura 6 – Exemplo de expansão das comunidades. Fonte: (PARÉS et al., 2017)	23
Figura 7 – Panorama da pesquisa. Fonte: Autor	34
Figura 8 – ARNES (The Academic and Research Network of Slovenia). Fonte: (TOPOLOGY-ZOO, a)	36
Figura 9 – Variações de topologias relacionadas com as posições de nós originais da rede. Fonte: (ARAÚJO; BARROS, 2020)	36
Figura 10 – Processamento do <i>dataset</i> . Fonte: Autor	37
Figura 11 – Panorama dos experimentos. Fonte: Autor	39
Figura 12 – Comparação entre PB e DC utilizando o algoritmo LPA. Fonte: Autor	41
Figura 13 – Comparação entre PB e DC utilizando o algoritmo GMO. Fonte: Autor	42
Figura 14 – Comparação entre PB e a Modularidade após DC utilizando algoritmo GMO. Fonte: Autor	43
Figura 15 – Tempo necessário para calcular a PB usando simulação. Fonte: Autor	44
Figura 16 – Tempo necessário para calcular a PB usando o algoritmo <i>Greedy Modularity Optimization</i> . Fonte: Autor	44
Figura 17 – Comparação entre PB e DC utilizando o algoritmo FC2. Fonte: Autor	45
Figura 18 – Comparação entre PB e DC utilizando o algoritmo FC3. Fonte: Autor	45
Figura 19 – Comparação entre PB e DC utilizando o algoritmo FC4. Fonte: Autor	46
Figura 20 – Comparação entre PB e DC utilizando o algoritmo FC5. Fonte: Autor	46
Figura 21 – Comparação entre PB e DC utilizando o algoritmo FC6. Fonte: Autor	46
Figura 22 – Variação da performance ao se alterar o número de <i>clusters</i> utilizando o algoritmo FC<k>. Fonte: Autor	47
Figura 23 – Comparação entre Resiliência de falhas simples e dupla e a Modularidade utilizando o algoritmo LPA. Fonte: Autor	48
Figura 24 – Comparação entre Resiliência de falhas simples e dupla e a Cobertura utilizando o algoritmo GMO. Fonte: Autor	48
Figura 25 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo LPA. Fonte: Autor	49
Figura 26 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo GMO. Fonte: Autor	49

Figura 27 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo FC2. Fonte: Autor	50
Figura 28 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo FC3. Fonte: Autor	50
Figura 29 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo FC4. Fonte: Autor	50
Figura 30 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo FC5. Fonte: Autor	51
Figura 31 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo FC6. Fonte: Autor	51

Lista de tabelas

Tabela 1 – Comparativo dos Trabalhos Relacionados. Fonte: Autor	32
---	----

Lista de abreviaturas e siglas

DC	Detecção de Comunidades
DTF	Design de Topologias Físicas
RTFO	Redes de Telecomunicações de Fibra Óptica
PB	Probabilidade de Bloqueio
RNA	Redes Neurais Artificiais
EMQ	Erro Médio Quadrático
MLP	Multi Layer Perceptron
QT	Qualidade de Transmissão
DWDM	Dense Wavelength-Division Multiplexing
EON	Elastic Optical Network
GML	Geography Markup Language
CSV	Comma-Separated Values
LPA	Label Propagation Algorithm
GMO	Greedy Modularity Optimization
FC	Fluid Communities

Sumário

	Lista de ilustrações	7
1	INTRODUÇÃO	13
1.1	Motivação	13
1.2	Justificativa	15
1.3	Problema de Pesquisa	16
1.4	Objetivos	16
1.4.1	Objetivos Específicos	17
1.5	Estrutura do Trabalho	17
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1	Detecção de Comunidades	18
2.2	Algoritmos para Clusterização	19
2.2.1	<i>Label Propagation</i>	20
2.2.2	<i>Greedy Modularity Communities</i>	21
2.2.3	<i>Fluid Communities</i>	22
2.3	Métricas para Avaliação de Clusterização	23
2.3.1	Modularidade	24
2.3.2	Performance de Clusterização	25
2.3.3	Cobertura	25
2.4	Avaliação da Qualidade de RTFOs	25
2.4.1	Resiliência em Redes de Fibra Óptica	26
2.4.2	Probabilidade de Bloqueio	27
3	TRABALHOS RELACIONADOS	29
3.1	Trabalhos Relacionados com Análise e Projeto de Redes Ópticas	29
3.2	Trabalhos que Aplicaram Técnicas de Detecção de Comunidades	30
3.3	Comparativo dos Trabalhos	31
4	METODOLOGIA	33
4.1	Etapas da pesquisa	33
4.2	<i>Dataset</i> de Resiliência e Probabilidade de Bloqueio	35
4.3	Arranjo Experimental	37
4.4	Especificações do Ambiente de Experimentos	38
5	RESULTADOS	41
5.1	Resultados usando Probabilidade de Bloqueio	41

5.1.1	<i>Label Propagation Algorithm</i>	41
5.1.2	<i>Greedy Modularity Optimization</i>	42
5.1.3	<i>Fluid Communities</i>	45
5.2	Resultados usando Resiliência	47
5.2.1	<i>Label Propagation Algorithm</i>	47
5.2.2	<i>Greedy Modularity Optimization</i>	49
5.2.3	<i>Fluid Communities</i>	49
6	CONCLUSÃO E TRABALHOS FUTUROS	52
6.1	Detecção de Comunidades para Obtenção da Resiliência e Probabilidade de Bloqueio	52
6.2	Conclusão dos Resultados	52
6.3	Trabalhos Futuros	54
	REFERÊNCIAS	56

1 Introdução

Desde o seu surgimento, a teoria dos grafos tem provido para a humanidade uma grande ferramenta que permite analisar as entidades e seus relacionamentos. Usando a teoria dos grafos, é possível mapear para um modelo matemático um sistema do mundo real, permitindo extrair, através desta modelagem, informações importantes sobre o sistema (BOCCALETTI et al., 2006).

De forma geral, qualquer sistema do mundo real no qual existam relações binárias entre as suas entidades pode ser modelado como um grafo. Alguns exemplos de sistemas que podem ser modelados são: relações comportamentais em grupos de pessoas, redes de amizades entre pessoas, redes de infraestrutura de comunicação, relações entre entidades do mercado financeiro, etc (LEWIS, 2011).

As conexões entre as entidades nesses sistemas podem indicar relações de dependência, que comunicam algum tipo de propriedade em comum ou de interdependência. Logo, não apenas os nós importam, mas também as relações entre eles (ZACHARY, 1977), (JONSSON et al., 2006). Por exemplo, em redes de comunicação, um conjunto de nós, com um alto número de conexões entre si poderá possuir uma resiliência maior; visto que, quanto maior o número de rotas redundantes entre os nós, maior a capacidade da rede suportar uma falha em algum *link*, pois ainda assim haverá uma rota redundante conectando os nós (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015).

Nesse sentido, o desenvolvimento das técnicas que consigam extrair informações das redes utilizando a sua estrutura topológica se torna de grande importância. Dentre essas técnicas se destaca a Detecção de Comunidades (DC) ou clusterização. Estas técnicas permitem que, a partir da estrutura topológica de um grafo, sejam identificados grupos de nós que possuam um grande número de conexões entre si e um baixo número de conexões para nós que pertençam a outros grupos (MOSCATO; SPERLÌ, 2021).

1.1 Motivação

A demanda por conexões de alta qualidade, com alta taxa de transmissão e baixa latência tem crescido nos últimos anos (LIRA et al., 2019). E para suportar essa demanda, Redes de Telecomunicações de Fibra Óptica (RTFO) têm sido utilizadas (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015a).

O *design* de Projeto de Topologia Física (DTF) dessas redes consiste na escolha

do número de *links*, escolha de quais nós deverão ser conectados, onde e quais amplificadores e repetidores serão usados, etc (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015a). Para essas escolhas deverão ser observadas diversas variáveis, como o custo dos dispositivos e da fibra, a qualidade dos dispositivos, e alguns indicadores que informam a qualidade da rede resultante, como a *Resiliência* da rede (LIRA et al., 2019) e a *Probabilidade de Bloqueio* (PB) (ARAÚJO; BARROS, 2020). Esses indicadores são fundamentais pois dão ao engenheiro uma visão quantificável da qualidade de uma rede sob determinada distribuição de carga. Logo, o estudo para aprimoramento do cálculo desses indicadores é de grande importância para os DTFs.

Por outro lado, é possível perceber um crescimento nas pesquisas e aplicações da DC nas mais diversas áreas do conhecimento (MOSCATO; SPERLÌ, 2021), (LIU et al., 2020). Porém, ainda que seja possível perceber um crescimento das aplicações da DC, na área de RTFO as pesquisas são praticamente inexistentes.

Neste contexto, surge a seguinte questão: Por que não utilizar as técnicas de DC para obtenção de informações das RTFOs? Dado o fato de que as RTFOs possuem uma natureza de grafo, as conexões entre os nós poderiam comunicar algum tipo de informação a respeito daquela rede. Mais especificamente, seria possível encontrar uma correlação entre as comunidades e os indicadores de qualidade dessas redes como a PB e Resiliência, por exemplo.

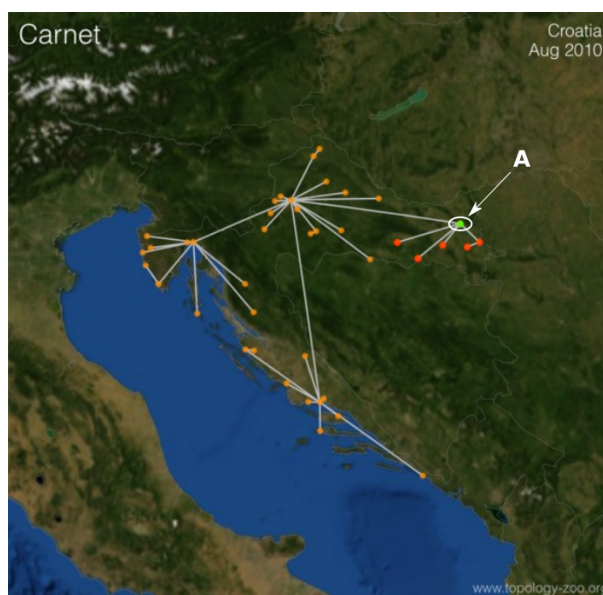


Figura 1 – CARNET (Croatian Academic and Research Network). Fonte: (TOPOLOGY-ZOO, b)

A relevância da DC para as RTFOs fica mais evidente quando se observa uma rede real. Por exemplo, na Figura 1, é possível perceber que os nós se conectam em um padrão que forma comunidades, alguns nós são mais centrais, e portanto, mais relevantes para a resiliência (hubs ou concentradores). Caso algum desses nós por

algum motivo seja perdido, a capacidade de comunicação entre uma grande parte da rede será perdida. Como no exemplo apresentado na Figura 1, se o nó A for perdido, os nós vermelhos serão desconectados do resto da rede. Dessa forma, apenas intuitivamente, é possível perceber que entre esses dois conceitos já existe algum tipo de correlação.

Cada rede possui uma topologia própria. Por exemplo, na Figura 2 a estrutura é bem distinta da rede na Figura 1, o que resulta em características diferentes. Porém, ainda assim serão características que estarão atreladas a disposição dos nós e de suas conexões. Logo, os indicadores que informam sobre a qualidade dessas redes poderiam possuir correlação com a capacidade dessas redes formarem comunidades.

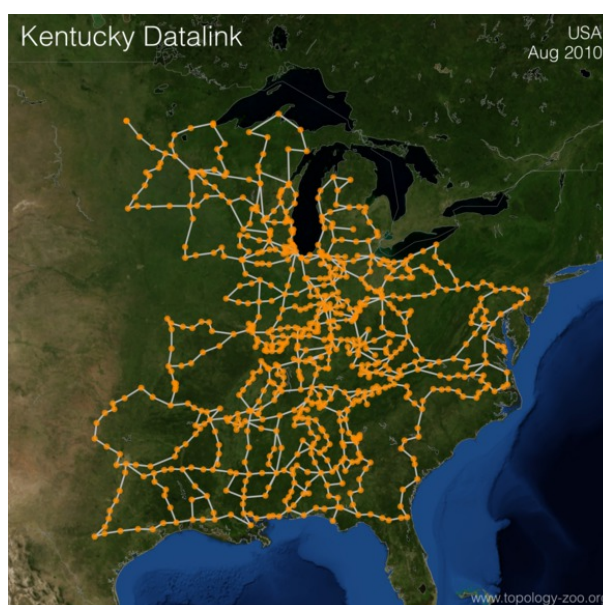


Figura 2 – Kentucky Datalink. Fonte: ([TOPOLOGY-ZOO](http://www.topology-zoo.org), c)

Outro ponto é que a DC nessas redes permitiria que várias análises fossem possíveis. A análise da capacidade de tráfego naquela comunidade, a possibilidade de aprimoramento de rotas entre os nós de uma dada comunidade, detecção de sub-redes que possuem nós com propriedades semelhantes como localização, e também análise de resiliência dessas comunidades, são alguns exemplos.

1.2 Justificativa

Como já mencionado, os indicadores de desempenho como PB e Resiliência são fundamentais para os projetos de DTF e isso é possível de se constatar pelas pesquisas realizadas ([LIRA et al., 2019](#)), ([ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015](#)), ([ROCCO et al., 2018](#)), ([ARAÚJO; BARROS, 2020](#)). Nessas pesquisas, a obtenção dos indicadores é realizada através de simuladores ([CHAVES et al., 2010](#)), ou de redes neurais de regressão ([LIRA et al., 2019](#)), ([ARAÚJO; BASTOS-FILHO; MARTINS-FILHO,](#)

2015). As simulações apresentam um resultado mais fiel ao real, pois conseguem modelar as características que interessam do mundo real na simulação, porém apresentam um alto custo computacional. Já os métodos baseados em aprendizagem de máquina como os regressores, embora sejam bastante rápidos para a regressão, apresentam um custo computacional elevado para treinamento e também são mais complexos de lidar.

Por outro lado, a utilização de DC é um problema já relativamente bem conhecido em diversos domínios de aplicação, com algoritmos clássicos apresentando boa eficiência computacional para resultados sub-ótimos, e baixa complexidade de operação. Contudo, há uma carência de pesquisas que usem as técnicas de DC para os projetos de DTFs. Neste sentido, a possibilidade de utilizar a técnica de DC para obtenção de informações da rede seria uma grande contribuição para a área.

Outro fator importante a se observar é que algumas iniciativas de projeto de DTF se utilizam de algoritmos genéticos (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015a). Porém a avaliação do resultado das populações, realizada pela função de *fitness* é feita com simulações ou com regressores na primeira etapa para identificação de redes muito ruins. A possibilidade de aprimorar a eficiência dessas iniciativas usando DC como uma primeira etapa permitiria uma maior simplificação e ganho computacional, visto que seria desnecessário um treinamento prévio, ou mesmo criação de um *dataset* para treinamento. Neste sentido, uma integração futura das técnicas estudadas neste trabalho com meta-heurísticas de otimização poderia diminuir o tempo total de projeto de redes de diversos dias para poucos minutos.

1.3 Problema de Pesquisa

Dado os pontos apresentados, um questionamento foi feito. Existe alguma correlação entre os indicadores de desempenho de uma RTFO e a sua capacidade de formar comunidades? Ou seja, seria possível obter informação acerca da rede utilizando-se das técnicas de DC?

Para tanto, o presente estudo objetiva investigar a possibilidade da obtenção de informações de uma RTFO utilizando DC. Mais especificamente, investigar se existe alguma correlação entre os indicadores de desempenho de uma RTFO e a capacidade que essa rede apresenta de formar comunidades.

1.4 Objetivos

O objetivo geral desta pesquisa é investigar a possibilidade da utilização de técnicas de DC para obter informações de uma RTFO, avaliando se há alguma correlação

entre os indicadores de PB e Resiliência e o nível de clusterização que essa redes apresentam.

1.4.1 Objetivos Específicos

Os objetivos específicos da pesquisa foram:

1. Desenvolvimento de uma plataforma para execução dos experimentos adequada para suportar as diversas técnicas de detecção de comunidades em grafos;
2. Investigação da correlação entre o nível de clusterização de uma rede e a sua PB com foco em redes reais, implantadas atualmente;
3. Investigação da correlação entre o nível de clusterização de uma rede e a sua Resiliência ou robustez com foco em redes reais, implantadas atualmente.

1.5 Estrutura do Trabalho

O presente texto foi organizado da seguinte forma. No capítulo 2 será apresentado toda a fundamentação teórica necessária para a compreensão do resto do trabalho. O capítulo 3 apresentará os trabalhos relacionados à pesquisa. A metodologia da pesquisa será apresentada no capítulo 4. No capítulo 5 serão apresentados os resultados obtidos dos experimentos. E por fim, no capítulo 6 a conclusão e discussão de possíveis trabalhos futuros.

2 Fundamentação Teórica

Neste capítulo serão apresentados os conceitos teóricos fundamentais para a compreensão do tema e dos resultados obtidos pelos experimentos.

2.1 Detecção de Comunidades

Para a compreensão do conceito de DC é necessário primeiro entender o que é um grafo. A teoria de grafos define um grafo como uma tupla $G = (V, E)$ no qual V é o conjunto de vértices e E o conjunto de arestas que conectam esses vértices. Esses vértices podem ter pesos de forma que a definição fica $G = (V, E, W)$ no qual W é o conjunto dos pesos para cada vértice de E (SHUN et al., 2016).

Um exemplo pode ser visto na Figura 3. A representação matemática para este grafo é $G = (V, E)$ no qual $V = \{A, B, C, D, E\}$ são os nós e o conjunto de arestas $E = \{(A, B), (B, C), (B, D), (C, D), (C, E)\}$, observe que $(A, B) = (B, A)$ pois este é um grafo não direcionado ou bidirecional.

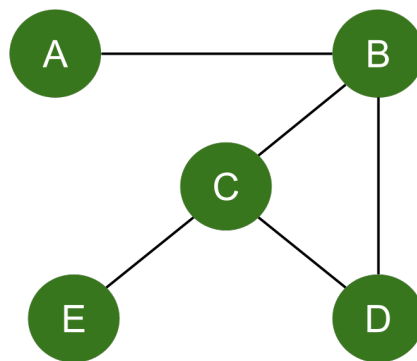


Figura 3 – Exemplo de Grafo. Fonte: Autor

Existem várias definições de *comunidade* ou *cluster* em um grafo. De forma geral, comunidade pode ser definida como um conjunto de nós no qual existam muitas conexões entre os nós do conjunto e poucas conexões com o resto do grafo (MOSCATO; SPERLÌ, 2021).

Um exemplo de comunidade é apresentado na Figura 4. É possível perceber como existem conjuntos de nós que possuem uma alta densidade de conexões internas e uma baixa densidade de conexões externas.

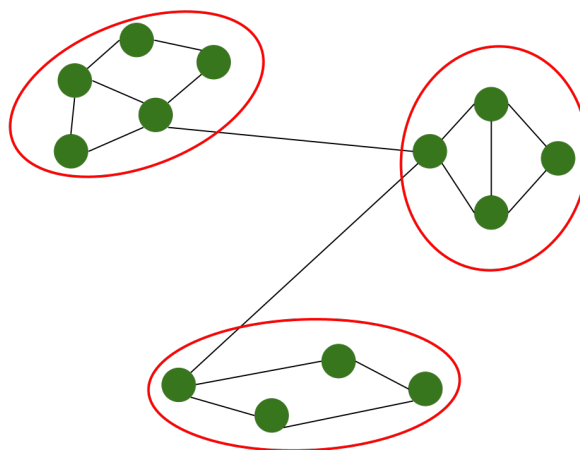


Figura 4 – Exemplo de Comunidade. Fonte: Autor

O problema de clusterização ou DC consiste em identificar as comunidades e seus respectivos nós (BOCCALETTI et al., 2006). As comunidades podem ser sobrepostas ou não. Em caso de serem sobrepostas os nós podem pertencer a uma ou mais comunidades e em caso de não serem sobrepostas os nós não poderão fazer parte de múltiplas comunidades. No caso de não sobreposição é comum também a operação ser chamada de particionamento (FORTUNATO, 2010).

No presente trabalho foi optado por utilizar apenas detecção de comunidades não sobrepostas, ou seja, algoritmos que fazem o particionamento do grafo. Isso foi feito para restringir o escopo de pesquisa e permitir a sua conclusão em tempo hábil.

2.2 Algoritmos para Clusterização

Encontrar a solução ótima de divisão das comunidades exige um alto poder computacional visto que a clusterização é um problema NP-Difícil (RAGHAVAN; ALBERT; KUMARA, 2007), (FORTUNATO, 2010). A obtenção da solução ótima é infactível para redes que não sejam pequenas.

Para os problemas do mundo real, onde se tem um grande número de vértices e de relações (MOSCATO; SPERLÌ, 2021), (STROGATZ, 2001), os algoritmos de aproximação de solução têm sido amplamente pesquisados. São algoritmos que fornecem resultados satisfatórios e que sejam factíveis, no que diz respeito ao tempo de processamento e de espaço de armazenamento (FORTUNATO, 2010).

Diversas abordagens já foram propostas (RAGHAVAN; ALBERT; KUMARA, 2007), (JONNALAGADDA; KUPPUSAMY, 2018), (ZENG et al., 2019), (PARÉS et al., 2017), (CLAUSET; NEWMAN; MOORE, 2004). Dentre elas, as famílias dos algoritmos gulosos, de lógica *fuzzy* e baseados em teoria dos jogos (MOSCATO; SPERLÌ, 2021).

Alguns dos algoritmos se tornaram bastante populares na área de DC. O presente

estudo fará uso de três desses algoritmos para condução dos experimentos. A motivação para sua escolha foi o fato dos mesmos serem algoritmos de particionamento e terem sua implementação disponibilizada pela biblioteca *NetworkX* (HAGBERG; SWART; CHULT, 2008). A seguir, será feita uma pequena explanação a respeito do funcionamento desses algoritmos.

2.2.1 Label Propagation

Em (RAGHAVAN; ALBERT; KUMARA, 2007), um algoritmo de DC chamado *Label Propagation Algorithm* (LPA) efetua o particionamento da rede, sem necessidade de conhecer previamente o número de comunidades ou o seu tamanho. Ele se baseia unicamente na estrutura topológica da grafo e possui um tempo de execução próximo ao linear, proporcional ao número de nós.

Como o próprio nome sugere, o funcionamento do algoritmo é baseado no conceito de propagação de um rótulo através dos nós. Este rótulo define a qual comunidade um nó pertence, e cada nó poderá apenas possuir um rótulo, o que o torna um algoritmo de particionamento.

A cada iteração um nó é visitado e o seu rótulo é atualizado conforme a maioria dos rótulos dos nós adjacentes. Na Figura 5 é possível ver um exemplo de atualização do rótulo. Se o nó A possui três nós adjacentes B, C e D com os respectivos rótulos azul, azul e verde, o nó A se tornará azul pois este é o rótulo mais frequente entre seus vizinhos. Em situações nas quais ocorre empate, a atualização do rótulo é feita aleatoriamente (RAGHAVAN; ALBERT; KUMARA, 2007).

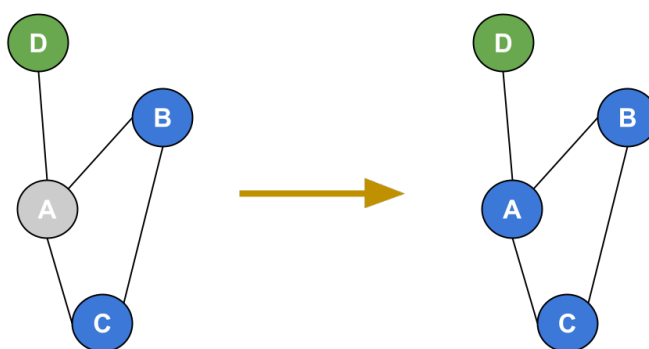


Figura 5 – Exemplo de atualização de rótulo. Fonte: Autor

Os nós são visitados em ordem aleatória, e duas maneiras de atualização são possíveis, a **síncrona**, que se baseia no estado da rede da iteração anterior, e a **assíncrona**, que se baseia no estado da rede da iteração atual. Formalmente, na síncrona, a atualização de um nó x na t -ésima iteração se baseia nos rótulos dos seus vizinhos da iteração $t - 1$, ou seja, $C_x(t) = f(C_{x_1}(t - 1), \dots, C_{x_k}(t - 1))$, no qual $C_x(t)$ é o rótulo do nó x no tempo t , e $\{x_1, x_2, \dots, x_k\}$ são os vizinhos de x . Na

assíncrona, $C_x(t) = f(C_{x_1}(t), \dots, C_{x_m}(t), C_{(m+1)}(t-1), \dots, C_k(t-1))$, no qual $\{x_1, \dots, x_m\}$ são os vizinhos de x que já foram atualizados e $\{x_{(m+1)}, \dots, x_k\}$ os que ainda não foram atualizados na iteração atual (RAGHAVAN; ALBERT; KUMARA, 2007).

Inicialmente cada nó possui um rótulo próprio, e conforme o algoritmo é executado estes rótulos são modificados. O comportamento do algoritmo fará com que regiões de mais densa conexão cheguem rapidamente em um consenso a respeito do rótulo. E este, por sua vez, irá expandir até encontrar outras regiões que são densamente conectadas e que possuam rótulos diferentes. Desta forma, em um dado momento a rede irá convergir para um estado no qual os seus nós não mais irão mudar seus rótulos.

O algoritmo irá parar quando a rede convergir ou quando atingir um valor máximo de iterações permitido. No fim, os nós que possuem rótulos em comum pertencerão à mesma comunidade.

2.2.2 Greedy Modularity Communities

O *Greedy Modularity Communities* (GMO) é um algoritmo guloso que busca identificar as comunidades através da maximização da modularidade considerando as junções entre as comunidades (CLAUSET; NEWMAN; MOORE, 2004).

Para compreender este algoritmo, é necessário primeiro entender o que é a modularidade. A modularidade é uma métrica que ganhou destaque na área de DC nos últimos anos (BRANDES et al., 2007) e é um indicador da qualidade da divisão de um grafo em suas comunidades.

Ela mede quão bom um particionamento é, no sentido de que o número de conexões internas das comunidades é maior que o número de conexões externas (BRANDES et al., 2007). A hipótese básica é que uma rede com conexões aleatórias entre seus vértices não apresenta comunidades. Desta forma, uma comparação entre o número de conexões internas de uma dada comunidade e o número de conexões em uma rede aleatória desta mesma comunidade é feita. Quanto maior o número de conexões em comparação com a rede aleatória maior será a modularidade, pois significa que aquela comunidade foi bem definida.

Uma modularidade igual a 0 significa que o número de conexões internas das comunidades é o mesmo do esperado em uma rede com conexões aleatórias. Valores diferentes de zero representam um desvio da aleatoriedade. De acordo com (CLAUSET; NEWMAN; MOORE, 2004), valores acima de 0,3 já são indicadores de uma boa divisão de comunidades.

Portanto, se valores maiores de modularidade significam melhores divisões da rede, então é possível encontrar boas divisões, escolhendo as aglomerações que

maximizem a modularidade. Este é o princípio de funcionamento do algoritmo GMO (CLAUSET; NEWMAN; MOORE, 2004). Inicialmente cada vértices pertencem a uma comunidade própria daquele vértice. A cada iteração, o par de comunidades adjacentes cuja junção trará um maior ganho de modularidade para a rede será escolhido para fazer a aglomeração.

O algoritmo efetua uma aglomeração das comunidades a cada iteração, e termina quando restar apenas uma comunidade. Todo o processo pode ser representado como uma árvore no qual as folhas são os vértices (comunidades iniciais), e cada nó interno corresponde às junções entre as comunidades.

Ao final, de posse do dendrograma que representa todas as junções entre as comunidades em todos os níveis, o algoritmo seleciona aquela que apresenta maior modularidade como a melhor divisão da rede em suas comunidades.

2.2.3 *Fluid Communities*

O algoritmo *Fluid Communities* (FC), apresentado em (PARÉS et al., 2017), também utiliza a metodologia de propagação como no *Label Propagation*. Porém adiciona um comportamento de propagação de fluidos que se expandem e "lutam" por território até que um equilíbrio seja alcançado.

Para se ter o comportamento de fluidos é utilizado o conceito de densidade. A densidade é distribuída igualmente através de todos os nós que pertencem a uma mesma comunidade. Formalmente, seja d a densidade total de uma comunidade C , então $d_i = \frac{d}{n}$ no qual d_i é a densidade de cada nó em C e n é a quantidade de nós da comunidade.

O número k de comunidades deve ser previamente definido. Ao iniciar, k nós são escolhidos aleatoriamente como ponto inicial da comunidade. Estes nós possuirão densidade máxima $d = 1$.

A cada iteração todos os nós serão visitados em ordem aleatória. Para cada nó visitado sua atualização será conforme a soma total das densidades de cada comunidade dos nós adjacentes; levando em conta também a densidade do próprio nó, caso ela já pertença a uma comunidade. Por exemplo, suponha um nó A que pertença a comunidade Φ e sua densidade seja de $0,125$, ele está conectado a mais outros três nós B, C e D , no qual B e C pertencem a comunidade Ω com densidades $0,25$ cada um e D pertence a comunidade Φ com densidade de $0,125$. Portanto, o somatório da densidade resultante na comunidade Ω será $0,25 + 0,25 = 0,5$ e o somatório da densidade resultante da comunidade Φ será $0,125 + 0,125 = 0,25$. Desta forma, o nó A irá mudar para a comunidade Ω , pois a soma das densidades para Ω foi maior. Em caso de empate, o nó permanecerá na comunidade inicial caso a sua comunidade faça parte

do empate. caso contrário, a escolha será aleatória nas comunidades que obtiveram maior densidade.

Desta forma, quanto mais nós uma comunidade adquire, menor será a densidade em seus nós, e portanto, menor a sua capacidade de conquistar mais nós. Chegará um momento em que as comunidades alcançarão a fronteira umas das outras e competirão para ganhar mais nós. Esta competição levará em conta a quantidade de relações que um dado nó possui e também na quantidade de nós que a comunidade em questão já possui.

Na Figura 6 é possível ver um exemplo da expansão de duas comunidades. Inicialmente as comunidades são posicionadas aleatoriamente com suas respectivas densidades $d = 1$. O nó da iteração, destacado com uma borda azul, é escolhido aleatoriamente e seu próximo valor obedece à regra de atualização de rótulo citada anteriormente.

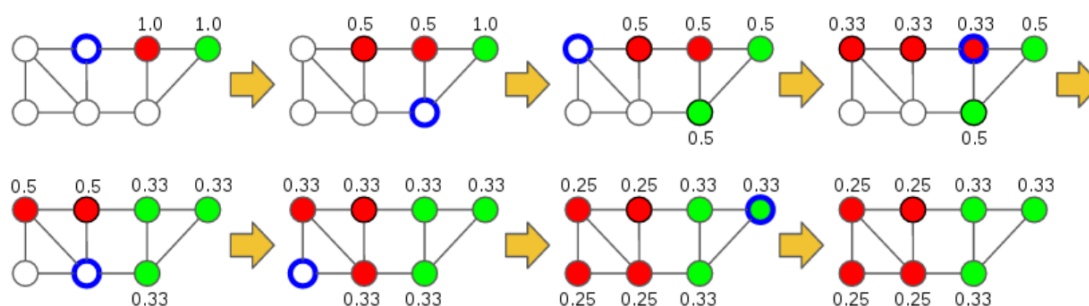


Figura 6 – Exemplo de expansão das comunidades. Fonte: (PARÉS et al., 2017)

FC é um algoritmo assíncrono, no qual cada vértice é atualizado usando o último estado parcial do grafo, podendo ele estar olhando para nós que já foram atualizados na iteração atual ou não. O algoritmo para quando a rede converge para um estado de estabilidade (PARÉS et al., 2017).

2.3 Métricas para Avaliação de Clusterização

Tão importante quanto os algoritmos que efetuam a DC são as métricas que avaliam a qualidade dessa detecção. As métricas de avaliação quantificam o quão bom é um dado particionamento (RAGHAVAN; ALBERT; KUMARA, 2007), (BRANDES et al., 2007).

Como dito anteriormente, embora haja uma ideia geral de que uma comunidade consiste em um *cluster* de muitas conexões internas e poucas conexões para o resto da rede, não existe uma definição universalmente aceita de comunidade, e esse entendimento pode levar à subjetividade do que seria uma boa divisão das comunidades.

Por outro lado, as métricas de avaliação permitem efetuar uma comparação objetiva dentre as diferentes técnicas de DC e também quantificar a capacidade que uma rede possui de formar comunidades, o que por si só já revela a sua importância. Além disso, as métricas são muito importantes para avaliação da qualidade de uma clusterização em redes que são muito grandes (MOSCATO; SPERLÌ, 2021), visto que uma análise manual seria impossível de ser feita.

O estudo atual fez uso de três métricas de avaliação de clusterização. O objetivo foi investigar se há alguma correlação entre os valores dessas métricas e os indicadores de qualidade de uma RTFO. Se existe alguma correlação entre a capacidade que uma dada rede possui de formar comunidades e a sua Resiliência e Probabilidade de Bloqueio. A seguir serão feitas breves descrições das métricas escolhidas e suas respectivas definições formais.

2.3.1 Modularidade

A modularidade compara a densidade de conexões entre os nós dentro de uma comunidade e a densidade esperada de conexões dessa mesma comunidade em uma rede de conexões aleatórias (RAGHAVAN; ALBERT; KUMARA, 2007). O pressuposto básico é de que grafos nos quais todos os pares de nós possuem a mesma probabilidade de estarem conectados não apresentam comunidades (NEWMAN; GIRVAN, 2004).

A definição matemática formal de modularidade é como segue. Seja k o número de comunidades da clusterização, pode-se definir uma matriz simétrica $\mathbf{e} = k \times k$ cujo elemento e_{ij} é a fração do número de arestas que conectam os *clusters* i e j , dividido pelo número de todas as arestas na rede. Desta forma, o traço da matriz $Tr \mathbf{e} = \sum_i e_{ii}$ representa a fração das arestas da rede que conectam os nós que pertence a uma mesma comunidade. Para boas divisões de comunidade o traço apresentará um valor alto. Porém o traço por si só não seria suficiente para identificar uma boa divisão de comunidades, por exemplo, uma divisão de comunidades com todos os nós pertencendo a uma comunidade teria $Tr \mathbf{e} = 1$, mas isso não é necessariamente uma boa divisão de comunidades. Logo se faz necessária a seguinte definição, seja $a_i = \sum_j e_{ij}$ o somatório das linhas ou colunas da matriz \mathbf{e} , que representa a fração das arestas que conectam os nós da comunidade i . Em uma rede no qual os nós se conectam aleatoriamente, pode-se afirmar que $e_{ij} = a_i a_j$ (NEWMAN; GIRVAN, 2004). Logo a modularidade pode ser definida formalmente como:

$$M = \sum_i (e_{ii} - a_i^2) \quad (2.1)$$

De acordo com (NEWMAN; GIRVAN, 2004) as redes normalmente apresentam modularidades no intervalo entre 0,3 e 0,7. Segundo (FORTUNATO, 2010), a relevância

dessa métrica pode ser observada pela frequência na qual ela é utilizada nas pesquisas.

2.3.2 Performance de Clusterização

A performance de clusterização é calculada inicialmente pela soma entre os pares de nós que estão conectados e que pertencem uma mesma comunidade e o número de pares de nós que pertencem às comunidades diferentes, mas não estejam conectados. Por fim, esta soma é dividida pela quantidade máxima de pares de nós possíveis (BRANDES; GAERTLER; WAGNER, 2003).

$$P = \frac{|\{(i, j) \in E, C_i = C_j\}| + |\{(i, j) \notin E, C_i \neq C_j\}|}{n(n-1)/2} \quad (2.2)$$

Formalmente a performance de clusterização pode ser definida como na Equação 2.2, no qual C_i é a comunidade ao qual o nó i pertence, E é o conjunto de arestas do grafo e n é o número de nós do grafo. O valor da performance de clusterização sempre será $0 \leq P \leq 1$ de forma tal que valores maiores significam uma maior qualidade da clusterização.

2.3.3 Cobertura

A métrica de avaliação da detecção de comunidades chamada cobertura consiste na razão entre o número de conexões internas das comunidades e o número total de conexões existentes no grafo (FORTUNATO, 2010). A definição formal da cobertura pode ser vista na Equação 2.3, no qual E é o conjunto de arestas e C_i é a comunidade ao qual o nó i pertence.

$$C = \frac{|\{(i, j) \in E, C_i = C_j\}|}{|E|} \quad (2.3)$$

O valor da cobertura sempre será $0 \leq C \leq 1$. Como bem observado por (FORTUNATO, 2010) por definição uma clusterização ideal seria aquela no qual as comunidades estivessem desconectadas umas das outras, o valor da cobertura nesse caso seria o máximo 1.

2.4 Avaliação da Qualidade de RTFOs

Como já mencionado no capítulo 1 as RTFOs têm sido utilizadas para suportar a demanda crescente por altas taxas de transmissão, baixa latência e alta disponibilidade. De acordo com (PANDEY; PAL et al., 2020), durante a pandemia de COVID-19 a utilização por serviços na Internet cresceram de 40% até 100% se comparado com os níveis anteriores aos *lockdowns*. O próprio distanciamento social levou a uma

aceleração no processo de transformação digital nas empresas, fazendo com que muitos colaboradores trabalhassem remotamente (PANDEY; PAL et al., 2020). Além disso, com o advento da tecnologia 5G os requisitos para as redes apenas tendem a crescer.

As redes de *backbone* são muito importantes, pois elas permitem a conexão entre as diversas redes que compõem a Internet. Portanto, os projetos dessas redes precisam apresentar um nível mínimo de qualidade para que justifique o seu custo e manutenção. O processo de avaliação da qualidade de uma RTFO leva em conta diversos aspectos, normalmente são utilizados indicadores que permitem quantificar o desempenho da rede (ARAÚJO, 2015). Dentre esses indicadores estão *Throughput* (ZHU; MUKHERJEE, 2002), Atraso Médio (BANERJEE; KUMAR, 2007), Probabilidade de Bloqueio (PB) (ARAÚJO; BARROS, 2020) e Resiliência (LIRA et al., 2019).

Para a presente pesquisa os indicadores usados foram a PB e Resiliência. Desta forma, uma explanação a respeito de suas definições e características será necessária, tendo em vista a boa compreensão do tema abordado no trabalho. O primeiro indicador apresentado será a Resiliência, e em seguida a PB.

2.4.1 Resiliência em Redes de Fibra Óptica

Resiliência é a capacidade que uma entidade possui de suportar uma adversidade. No contexto de redes de comunicação a resiliência à falha consiste na capacidade de sobrevivência da rede após sofrer uma falha em um link ou nó (LIRA et al., 2019).

A resiliência contra falhas é um importante requisito para a maioria das redes de alta capacidade. Visto que essas redes transportam muitos dados a falha em um link ou nó causaria muitos problemas (LIRA et al., 2019). De acordo com (RAMASWAMI; SIVARAJAN; SASAKI, 2009) a quebra de um único cabo de fibra óptica poderia desconectar milhões de usuários e resultar em milhões de dólares em perdas financeiras.

Como já foi dito, a Resiliência de uma rede óptica é a capacidade que uma rede possui de continuar funcionando mesmo após uma falha. Uma conexão na rede normalmente é composta por uma rota que possui diversos nós e links. Para a conexão ser estabelecida, é necessário que toda a rota esteja funcionando corretamente. Se algum link ou nó falhar, a rota é perdida.

Dessa forma, uma rede que sofre uma falha precisará encontrar uma rota redundante que conecte o mesmo par de nó para que a conexão não seja perdida. Logo, a capacidade de sobrevivência de uma rede está relacionada com a presença de rotas redundantes (LIRA et al., 2019).

De acordo com (LIRA et al., 2019) existem três tipos de serviço de fornecimento de conexão em uma RTFO:

- (1 + 1) o cliente possui uma rota de trabalho e uma rota redundante que permanece ativa durante todo o tempo; neste caso o serviço é chamado de *protegido*;
- (1 + R) o cliente possui uma rota de trabalho e uma rota redundante, porém esta não permanece ativa, ela é ativada apenas quando a rota de trabalho é perdida; neste caso o serviço é chamado *restaurável*;
- (1 + 1 + R) o cliente possui um serviço que é *protegido* e *restaurável* ao mesmo tempo;

A principal diferença entre os serviços (1 + 1) e (1 + R) é que o serviço (1 + R) precisa de um tempo para descobrimento da rota redundante e comutação que é da ordem de 60 ms (RAMASWAMI; SIVARAJAN; SASAKI, 2009).

A medição da Resiliência de uma rede é feita através de simuladores ou uso de indicadores próprios que quantificam rotas redundantes (LIRA et al., 2019). O presente trabalho fará uma comparação usando os resultados obtidos por simulações do SIMTON (CHAVES et al., 2010).

O SIMTON é um simulador de redes ópticas que modela um tráfego dinâmico de rede. Durante a sua execução são feitas falhas de um único enlace (falhas simples) ou falhas de dois enlaces ao mesmo tempo (falhas duplas). As conexões que foram comprometidas pela falha serão roteadas por outros caminhos caso exista alguma rota disponível. A Resiliência será o percentual de demandas de clientes que conseguiram ser recuperadas.

$$R = \frac{D_r}{D_t} \quad (2.4)$$

Na Equação 2.4 é possível ver a definição formal de um indicador de Resiliência, que foi adotada no trabalho, no qual D_r é a quantidade de demandas afetadas pelas falhas que foram recuperadas e D_t é a quantidade total de demandas afetadas.

2.4.2 Probabilidade de Bloqueio

A PB permite avaliar a capacidade de uma rede sujeita a tráfego dinâmico (ARAÚJO; BARROS, 2020). Ela provê ao arquiteto de rede uma quantificação da performance da rede sob determinada distribuição de carga, o que é imprescindível para comparação com as outras topologias cujo arquiteto precisará escolher.

Formalmente a PB é definida como na Equação 2.5, no qual R_b corresponde ao número total de requisições bloqueadas e R_t o número total de requisições efetuadas

na rede (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015b).

$$PB = \frac{R_b}{R_t} \quad (2.5)$$

A PB consiste em um percentual $0 \leq PB \leq 1$ no qual uma PB próxima a 0 corresponde a uma rede que sob determinada distribuição de carga apresentou um bom desempenho. Uma PB próxima a 1 corresponde a uma rede que sob determinada distribuição de carga apresentou um péssimo desempenho, visto que praticamente todas as requisições efetuadas foram bloqueadas.

Dado o fato de serem redes de *backbone*, existe uma grande exigência por alta disponibilidade, ou seja, os projetos de DTF tem como um dos requisitos mínimos uma baixa PB. De acordo com (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015b) a maioria das redes reais apresentam $PB < 1\%$. O que mostra a sua relevância e importância.

3 Trabalhos Relacionados

Neste capítulo serão apresentados os trabalhos relacionados à presente pesquisa. Uma categorização dos trabalhos foi feita, a saber: trabalhos que se relacionam com a área do problema de pesquisa e trabalhos que se relacionam com a técnica utilizada que foi a Detecção de Comunidades.

3.1 Trabalhos Relacionados com Análise e Projeto de Redes Ópticas

O estudo tem por objetivo a investigação de possíveis relações entre as métricas de clusterização, e a Resiliência e PB em RTFOs. Desta forma, os temas levados em conta para apresentação dos trabalhos relacionados são sobre métodos para avaliação de Resiliência e PB em RTFOs.

Na temática de Resiliência em redes ([ARAÚJO et al., 2017](#)) apresentou um modelo de Redes Neurais Artificiais (RNA) para predição da Resiliência em RTFOs. No trabalho um *benchmarking* envolvendo 9 topologias de RTFOs reais foi usado, o Erro Médio Quadrático (EMQ) encontrado para falhas aleatórias, ficou entre $1,76 \times 10^{-3}$ e $4,41 \times 10^{-6}$, a depender da topologia. O tempo de cálculo foi 300 vezes mais rápido do que as simulações tradicionais.

De forma semelhante, ([LIRA et al., 2019](#)) conduziu uma investigação sobre as configurações dos hiperparâmetros de uma *Multi Layer Perceptron* (MLP) para o cálculo da Resiliência de uma RTFO que foi submetida às falhas simples e dupla. O trabalho alcançou um ganho no tempo de cálculo 51.000 vezes mais rápido do que as simulações tradicionais e um EMQ em torno de 3×10^{-3} .

Em ([ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015b](#)) foi apresentado um modelo de RNA para predição de PB em RTFOs densas. Semelhantemente, em ([ARAÚJO; BARROS, 2020](#)) uma MLP de predição de PB em RTFOs elásticas foi usada. Ambos os trabalhos apresentaram um ganho no tempo de cálculo de 7.500 e 60, respectivamente, se comparado com as simulações. Os EMQs dos modelos nos trabalhos foram $2,57 \times 10^{-4}$ e $1,85 \times 10^{-5}$, respectivamente.

Outra abordagem que foi feita se utilizando da PB em RTFOs foi a apresentada por ([ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015a](#)). Nela uma RNA foi usada para diminuir o tempo necessário da função de *fitness* do algoritmo evolucionário. O objetivo foi utilizar a RNA para calcular a PB e identificar redes ruins. Caso a rede fosse

considerada de boa qualidade pelo modelo, uma segunda etapa de avaliação usando o simulador (CHAVES et al., 2010) era usada para a obtenção de uma PB mais precisa. Houve uma redução no tempo de execução do algoritmo entre 84% e 88%.

De forma geral, todos esses trabalhos que se relacionam com a pesquisa atual apresentam uma lacuna em comum. Embora essas abordagens que se utilizam de RNAs para predição da PB ou da Resiliência apresentem resultados promissores com EMQs baixos e baixo custo computacional para predição, existem desvantagens nessas abordagens. A geração de *datasets* e também o custo computacional para o treinamento e avaliação da qualidade de um determinado modelo elevam a complexidade de sua efetiva aplicação na indústria. Logo, seria de grande relevância uma abordagem que unisse a capacidade de se obter informações de uma rede com baixo custo computacional e que tivesse uma baixa complexidade de implementação.

3.2 Trabalhos que Aplicaram Técnicas de Detecção de Comunidades

A técnica utilizada no presente trabalho é a DC, mais especificamente, o particionamento da rede para avaliação da Resiliência e da PB. Uma aplicação muito interessante de DC é a apresentada em (JONSSON et al., 2006) no qual uma rede de relações entre proteínas foi inicialmente criada se utilizando de homologia. Isso permitiu que, ao serem detectadas as comunidades grupos funcionais de proteínas fossem encontrados. Ou seja, grupos de proteínas com funções iguais ou semelhantes, no qual espera-se que estejam envolvidas no mesmo processo (FORTUNATO, 2010). No trabalho foi feita uma ilustração da técnica usando comunidades de proteínas relacionadas na metástase de câncer em ratos.

Em (LI; DONG; MOSTAFAVI, 2020) foi realizado um estudo para entender as relações que existem entre a formação de comunidades em organizações de infraestrutura que formam sistemas interdependentes e a capacidade de colaboração entre esses agentes para reagirem às situações de contingência. Foi usado como exemplo o furacão Harvey na cidade de Harris County, no Texas. O objetivo era encontrar algum padrão entre a estrutura das comunidades e as redes de colaboração entre os sistemas. A análise foi feita em níveis de interações diária, semanal, mensal e anual (LI; DONG; MOSTAFAVI, 2020). Esse trabalho é interessante pois se utiliza da quantificação da métrica de modularidade para avaliar como a frequência de interação entre os sistemas poderia ser impactada.

O trabalho realizado por (ROCCO et al., 2018) é o que mais se assemelha à proposta desta pesquisa. Nela uma abordagem de avaliação dos efeitos de falhas em uma rede é apresentada. O objetivo foi identificar a capacidade que uma rede possui de

se recuperar de falhas sob determinadas sequências de restauração. Foram observado como esses eventos disruptivos afetavam as comunidades e a rede como um todo. O trabalho fez um estudo de caso em redes de transporte de energia elétrica (ROCCO et al., 2018).

De forma geral, esses trabalhos apresentam aplicações importantes da DC em problemas do mundo real e com grande relevância nas suas respectivas áreas. E isto, inclusive no que diz respeito à resiliência em determinados domínios, como por exemplo resiliência entre sistemas interdependentes (LI; DONG; MOSTAFAVI, 2020) e resiliência em redes de transporte elétrico (ROCCO et al., 2018). Porém ainda assim, existem lacunas de pesquisa no que tange utilização de técnicas de DC para avaliar o desempenho de RTFOs.

3.3 Comparativo dos Trabalhos

No trabalho atual, serão feitas algumas investigações acerca das relações existentes entre a capacidade de clusterização e os indicadores de Resiliência e PB em RTFOs. Esta abordagem é promissora no sentido de permitir a avaliação de Resiliência e PB com baixo custo computacional e de forma simples. Este trabalho não tem por objetivo esgotar o assunto, de modo que trabalhos futuros poderão avaliar a possibilidade de substituição das simulações custosas pelas técnicas estudadas aqui. Ou seja, essa pesquisa pode ser entendida como um primeiro passo na direção da utilização dessa abordagem como mecanismo de avaliação de qualidade de redes em ferramentas de planejamento complexas (como avaliação de aptidão em algoritmos evolucionários multiobjetivos) (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015a).

Nesta seção será feita uma comparação entre os trabalhos relacionados apresentados e a proposta da pesquisa. Na lista abaixo são apresentados os conceitos levados em conta para a comparação:

- (C_1) Utilização de DC;
- (C_2) Possibilidade de se obter Resiliência de uma RTFO;
- (C_3) Possibilidade de se obter PB de uma RTFO;
- (C_4) Baixo nível de complexidade para implantação prática - complexidade aqui refere-se à necessidade de gerar *dataset*, treinar modelos, usar os modelos, tudo isso exige um conhecimento técnico avançado e alto custo computacional; o que aumenta a complexidade da implantação prática;

- (C_5) Baixo custo computacional - refere-se ao custo computacional envolvido em todo o processo, no contexto de RNAs diz respeito ao custo de treinamento e avaliação e no contexto dos simuladores, o processo das simulações em si.

Na Tabela 1 é apresentado de forma gráfica como os trabalhos relacionados se complementam e como a abordagem da pesquisa objetiva sanar as lacunas que estão abertas no que se refere à aplicação de DC em RTFOs para cálculo de PB e Resiliência.

Trabalho / Contribuições	C_1	C_2	C_3	C_4	C_5
(LIRA et al., 2019)		X		X	
(ARAÚJO et al., 2017)		X		X	
(ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015b)		X		X	
(ARAÚJO; BARROS, 2020)			X	X	
(ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015a)		X	X		
(JONSSON et al., 2006)	X			X	X
(LI; DONG; MOSTAFAVI, 2020)	X			X	X
(ROCCO et al., 2018)	X				X
Abordagem da Pesquisa	X	X	X	X	X

Tabela 1 – Comparativo dos Trabalhos Relacionados. Fonte: Autor

4 Metodologia

Neste capítulo serão mostrados os passos necessários para a execução da pesquisa e também alguns detalhes específicos a respeito do *dataset* utilizado e da condução dos experimentos. O capítulo foi dividido da seguinte forma. As etapas da pesquisa serão descritas na Seção 4.1. Na Seção 4.2 é explicado como o *dataset* com as topologias e suas Resiliências e PBs foram gerados. A Seção 4.3 apresenta o arranjo experimental. E por fim, na Seção 4.4 é apresentado o ambiente no qual os experimentos foram conduzidos.

4.1 Etapas da pesquisa

Na Figura 7 é apresentado um diagrama com as etapas da pesquisa. A primeira etapa foi a pesquisa bibliográfica que foi feita para obter uma melhor compreensão dos trabalhos que têm sido realizados na área de DC. A estratégia adotada inicialmente para o levantamento bibliográfico consistiu em entender o panorama da área por meio de *surveys*, para em seguida afunilar o escopo da pesquisa para um nicho com contribuição clara. As buscas foram feitas inicialmente nas bases IEEE, Elsevier, ACM e também pelo Google Scholar. As buscas foram feitas utilizando as seguintes *queries*: “*Community Detection Survey*”, “*Fiber Optic Networks Performance Evaluation*”, “*Applications of Community Detection Survey*”, “*Artificial Intelligence Community Detection*”.

Durante o processo de levantamento bibliográfico foi detectado uma lacuna na aplicação de DC em RTFOs. Logo a primeira definição do escopo da pesquisa foi se aprofundar no tópico de DC em RTFOs. A estratégia inicial consistia no desenvolvimento de uma plataforma de experimentos para comparação dos resultados obtidos usando algoritmos clássicos de DC e as técnicas de *Deep Learning* para DC.

Durante a implementação da plataforma dois pontos ficaram muito claros. O primeiro consistia no tempo para condução da pesquisa que seria insuficiente dada a complexidade de se terem implementações de algoritmos clássicos e também de algoritmos que usassem *Deep Learning*. O segundo ponto consistia em um detalhe observado nas RTFOs, que é o número de nós. As iniciativas que usam *Deep Learning* para DC são justificadas pelo grande número de nós e de arestas que um determinado grafo possui. Porém, as RTFOs não possuem um número elevado de nós, da ordem de milhares. Normalmente as redes ópticas apresentam um número de nós menor que 100.

Neste estágio da pesquisa foi realizada uma análise para refinar o escopo da pes-

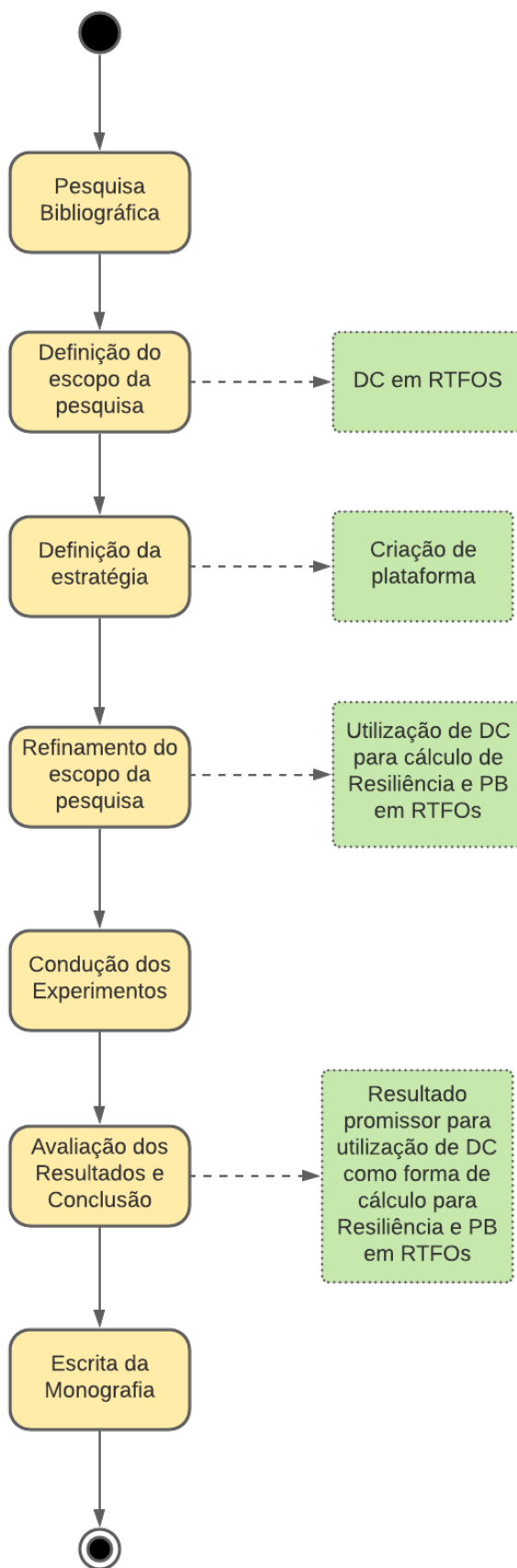


Figura 7 – Panorama da pesquisa. Fonte: Autor

quisa para um nicho promissor do ponto de vista científico e tecnológico considerando o eixo de redes de longa distância. Durante esta análise, o seguinte questionamento foi levantado: *“Existe alguma relação entre as métricas de clusterização de uma RTFO e a sua qualidade?”*. Foi percebido que havia uma lacuna na literatura a respeito dessa questão e também que uma resposta para esse questionamento traria contribuição para a área, mesmo se a resposta fosse não. Sendo assim, o escopo da pesquisa foi refinado para responder o questionamento da possibilidade de correlação entre a capacidade que uma RTFO possui de formar comunidades e a sua qualidade.

Definido o escopo da pesquisa, na próxima etapa os experimentos foram executados em diferentes cenários e muitos dados foram coletados para análise. O instrumento gráfico adotado para análise da correlação entre as variáveis foram gráficos de dispersão, que são úteis considerando uma pesquisa inicial exploratória para estabelecer a correlação entre as grandezas. Por fim, a monografia foi escrita para divulgação do trabalho.

4.2 *Dataset* de Resiliência e Probabilidade de Bloqueio

O *dataset* necessário para a execução do trabalho consiste em um conjunto de topologias de RTFOs com suas respectivas PB e Resiliência.

Para a Resiliência, foi utilizado o *dataset* resultante do trabalho de (LIRA et al., 2019), no qual o simulador utilizado foi o (CHAVES et al., 2010) em redes conhecidas como densas ou DWDM (Dense Wavelength-Division Multiplexing). As simulações geravam eventos de falhas simples e dupla, e então as demandas afetadas tentavam encontrar rotas alternativas. A Resiliência de uma rede foi considerada como a proporção entre o número de demandas que conseguiram se recuperar e o número total de demandas afetadas pelas falhas.

Para a PB, foi utilizado o *dataset* resultante do trabalho de (ARAÚJO; BARROS, 2020), o simulador utilizado também foi o (CHAVES et al., 2010), porém em redes ópticas elásticas ou EON (Elastic Optical Network). O bloqueio foi considerado como uma nova demanda que não pudesse ser atendida pela rede por falta de rotas disponíveis. A PB calculada foi a proporção entre o número de demandas bloqueadas e o número total de demandas feitas à rede.

Em ambos os *datasets*, as topologias das RTFOs foram geradas a partir da rede ARNES (KNIGHT et al., 2011) que pode ser vista na Figura 8. Baseado na posição geográfica diferentes topologias foram geradas através de variações no conjunto de arestas. Ou seja, para cada topologia embora os nós estivessem com a mesma posição geográfica as conexões entre esses nós se diferenciavam. Por exemplo, na Figura 9 é possível perceber que embora as posições dos nós sejam as mesmas, existem



Figura 8 – ARNES (The Academic and Research Network of Slovenia). Fonte: (TOPOLOGY-ZOO, a)

alguns pares de nós que possuem conexões na rede 9b, mas não estão conectados na rede 9a. Essas pequenas variações na topologia geram resultados de Resiliência e PB bastante diferentes, pois embora a posição geográfica dos nós seja igual a disposição das rotas é modificada, melhorando ou degradando a qualidade da rede.

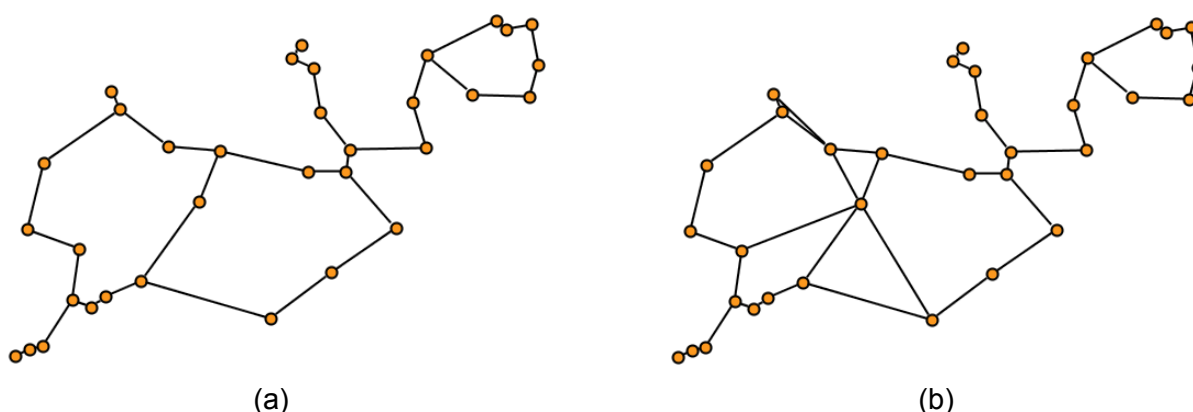


Figura 9 – Variações de topologias relacionadas com as posições de nós originais da rede. Fonte: (ARAÚJO; BARROS, 2020)

O número total de topologias gerados no *dataset* de Resiliência foi de 15.000. No *dataset* de PB foram geradas 15.000 topologias diferentes nos quais para cada topologia foram realizadas 10 simulações com diferentes valores de carga na rede totalizando 150.000 entradas. Para o presente trabalho, do *dataset* que possui diferentes valores de carga foram consideradas apenas as entradas com topologias diferentes e que foram submetidas à mesma carga para o cálculo da PB.

Um ponto importante a ser destacado é o fato de que as topologias não possuíam toda a sua matriz de adjacência nos arquivos do *dataset*. Como as RTFOs consideradas possuíam links bidirecionais, a matriz de adjacência é simétrica em relação à diagonal principal, logo apenas as entradas que ficam acima da diagonal principal foram salvas

no *dataset* inicial. Sendo necessário um pré-processamento para transformar esses vetores de entradas em um arquivo GML (*Geography Markup Language*) válido que pudesse ser utilizado posteriormente pela plataforma desenvolvida. Na Figura 10 é possível ver um panorama do processo para pré-processamento do *dataset* que foi usado nos experimentos.

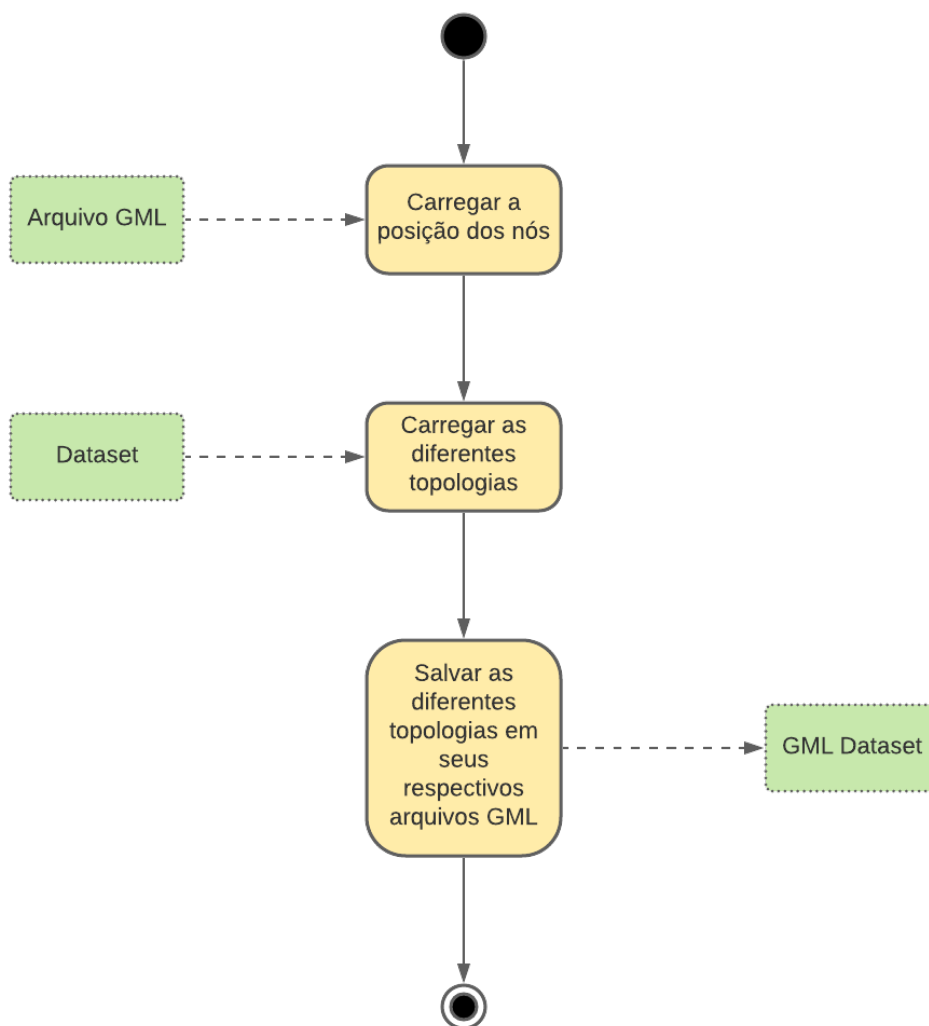


Figura 10 – Processamento do *dataset*. Fonte: Autor

4.3 Arranjo Experimental

Os experimentos conduzidos foram realizados através das seguintes etapas:

1. Transformação do vetor de entradas do *dataset* para o formato GML, para cada topologia no qual seria feita a DC um arquivo GML seria criado;
2. Carregamento do arquivo GML para a memória, em uma estrutura de dado que permitisse a execução dos algoritmos;

3. Detecção de Comunidades propriamente dita feita pelos algoritmos de clusterização LPA, GMO e FC;
4. Calcular os valores das métricas *cobertura*, *modularidade* e *performance* sobre o particionamento resultante;
5. Para cada topologia adicionar uma entrada no arquivo CSV (Comma-Separated Values) de saída, com os valores da Resiliência ou PB e os valores das métricas de clusterização sobre o particionamento realizado pelos algoritmos de DC.

Para cada algoritmo de DC, foi gerado um arquivo CSV com os resultados das métricas, e também uma entrada com os valores da Resiliência, e a PB. Desta forma seria possível fazer comparações isoladas entre cada indicador de qualidade da rede e as métricas de avaliação de qualidade do particionamento.

O método escolhido para avaliar se existem relações entre a Resiliência e PB e as métricas de avaliação da qualidade de DC foi a utilização de gráficos de dispersão sob um conjunto grande de topologias diferentes. Se houvesse algum tipo de correlação entre esses valores, um padrão no gráfico de dispersão seria formado. Outros instrumentos de análise podem ser usados de forma complementar no futuro, como o uso de indicadores de correlação como o de Pearson. Na Figura 11 é apresentado um diagrama com as etapas específicas dos experimentos.

Como já dito anteriormente, o número total de topologias disponíveis para os experimentos é de 15.000. Porém após a geração dos gráficos de dispersão foi percebido que com um número tão elevado de topologias o gráfico ficara com muitos pontos, o que acabava poluindo o gráfico e atrapalhando a análise. Foram feitos testes com valores menores de topologia, e um número que conseguia representar bem os resultados sem poluir o gráfico foi de 1.500 topologias. Contudo, o número reduzido foi escolhido de forma aleatória, de modo que a base resultante não possua nenhum viés de escolha

Portanto, foram executados particionamentos sobre as diferentes topologias, 1.500 topologias do *dataset* de Resiliência e 1.500 topologias do *dataset* de PB. No total foram realizados 3.000 execuções para cada algoritmo de DC e sobre as comunidades resultantes foram calculadas as métricas de clusterização.

4.4 Especificações do Ambiente de Experimentos

A máquina no qual foram conduzidos os experimentos possui uma memória RAM de 8GB, processador Intel I5-7^oGen com frequência de 2.5GHz, e sistema operacional GNU/Linux distribuição Ubuntu 20.04.

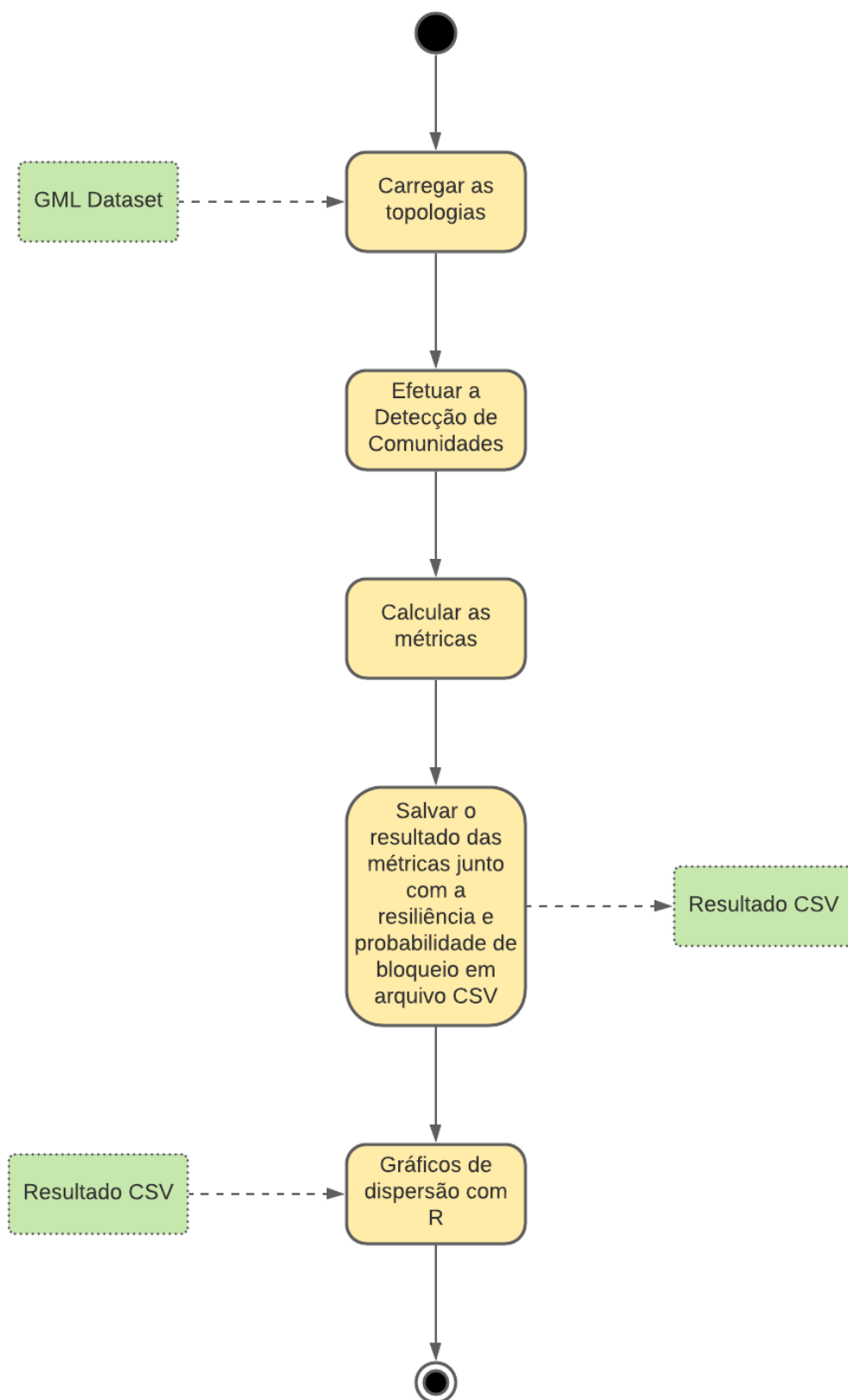


Figura 11 – Panorama dos experimentos. Fonte: Autor

Para a execução dos experimentos foi criada uma plataforma utilizando a linguagem Python, nela foram implementados os códigos úteis para todo o processo de experimentação. O objetivo é que novas pesquisas poderão se utilizar dela para alavancar seus trabalhos e também fazerem novas contribuições para a plataforma. A plataforma se chama *Community Detection Research Initiative* e seu código-fonte pode ser encontrado em ([BARROS, 2021](#)).

Uma das bibliotecas utilizadas foram a NetworkX ([HAGBERG; SWART; CHULT, 2008](#)) uma biblioteca que forneceu os algoritmos de DC e as métricas de clusterização usadas no presente trabalho. Além disso a biblioteca provê uma estrutura de dados própria para lidar com grafos, que pode ser facilmente carregada e escrita em arquivos no formato GML. Outra biblioteca usada foi a Folium ([VISUALIZATION, 2020](#)), que permitiu a plotagem dos nós e das arestas no mapa usando as suas coordenadas.

Por fim, os resultados foram consolidados pela plotagem de gráficos de dispersão e gráficos de linha para avaliação do tempo de computação usando a linguagem R ([R Core Team, 2021](#)).

5 Resultados

Neste capítulo serão apresentados em duas seções os resultados encontrados. A primeira seção mostrará a comparação da proposta em relação à PB, e a segunda seção a comparação em relação à Resiliência. A sequência escolhida para apresentação dos algoritmos foi a seguinte:

- *Label Propagation Algorithm* (LPA).
- *Greedy Modularity Optimization* (GMO).
- *Fluid Communities* (FC) com k clusters, $k \in \{2, 3, \dots, 6\}$.

5.1 Resultados usando Probabilidade de Bloqueio

Nesta seção serão apresentados os resultados da comparação das métricas de clusterização e da PB. Como já dito anteriormente a PB consiste na razão entre o número de requisições bloqueadas e o número total de requisições efetuadas na rede (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015b).

5.1.1 *Label Propagation Algorithm*

O algoritmo LPA foi o que apresentou os resultados menos relevantes da pesquisa. Na Figura 12 são apresentados os resultados da comparação entre as métricas de modularidade, cobertura e performance. Os resultados não apresentaram uma correlação bem definida, porém foi possível perceber que a alteração das métricas gera resultados diferentes. Por exemplo a diferença entre as Figuras 12b e 12c sugere que é possível obter informações de diferentes perspectivas ao se alterar as métricas de avaliação de um determinado particionamento.

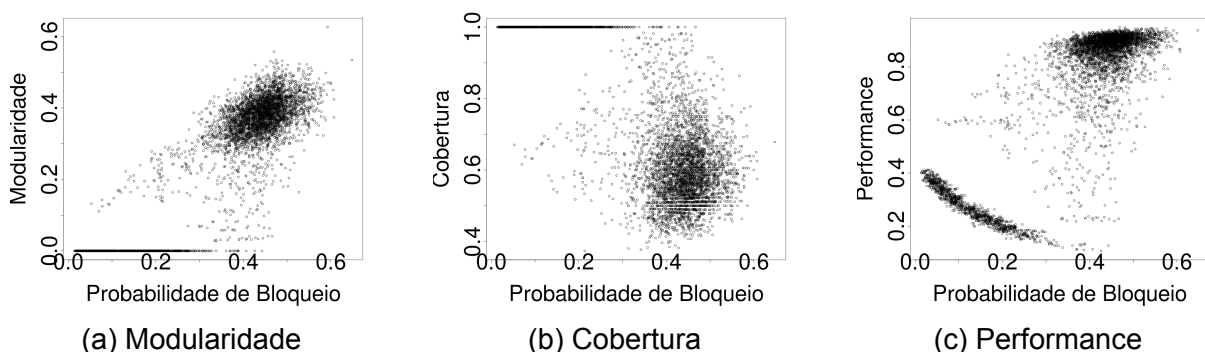


Figura 12 – Comparação entre PB e DC utilizando o algoritmo LPA. Fonte: Autor

5.1.2 Greedy Modularity Optimization

O algoritmo GMO apresentou resultados bastante promissores. Na Figura 13 os resultados para todas as métricas de clusterização apresentaram uma clara correlação ao passo que os valores dessas métricas aumentam.

As três métricas utilizadas são de maximização, ou seja, quanto maior o seu valor, maior a qualidade do particionamento, portanto, maior a capacidade que a rede possui de formar comunidades. Logo, um padrão interessante que pode ser observado daqui, é justamente o fato de que isso é bem o que caracteriza uma correlação positiva. Ou seja, quanto maior é a capacidade da rede apresentar um padrão de formação de comunidades, maior será a chance desta rede apresentar uma maior PB. Observe também que embora todos os gráficos mostrem uma padrão de crescimento, as diferentes métricas mostram perspectivas ligeiramente diferentes a respeito dos resultados.

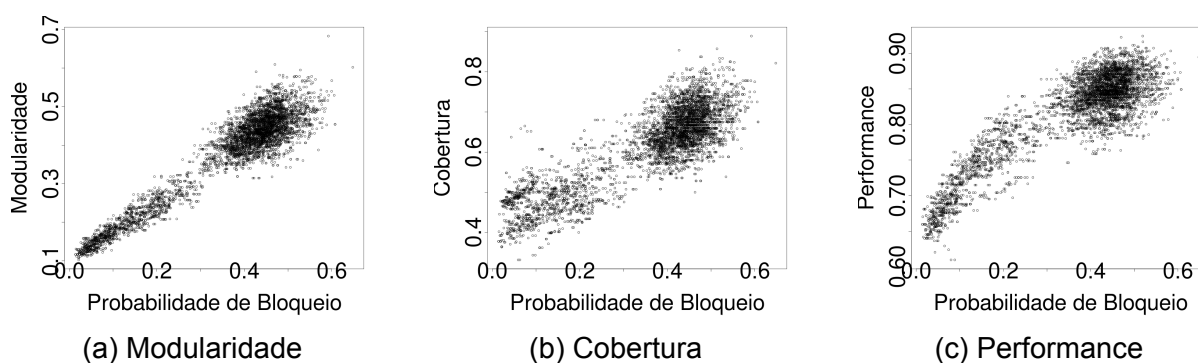


Figura 13 – Comparação entre PB e DC utilizando o algoritmo GMO. Fonte: Autor

Na Figura 14 é apresentado de forma mais nítida o gráfico 13a. Nele é possível ver duas topologias que foram adicionadas, uma com baixa PB e outra com uma alta PB. O gráfico mostra o resultado da comparação entre a métrica de modularidade e a PB. Observe como os pontos apresentam um padrão de linearidade. Outro aspecto muito interessante é a amplitude dos valores da modularidade que se assemelha à amplitude da PB.

Outra questão importante que deve ser considerada é o que pode ser visto na Figura 14. Existe uma maior concentração dos pontos para valores mais baixos de PB e um maior dispersão dos pontos para valores mais altos de PB. Quanto mais afastados os pontos estão, mais complexas serão as aplicações práticas, pois seria difícil inferir qual a PB de uma rede dada uma modularidade que apresenta um intervalo possível de PB muito grande. Entretanto, de acordo com (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015b), os valores de PB para aplicações reais em RTFOs precisam ficar no intervalo $PB < 1\%$, como esse fenômeno ocorre para valores de PB muito grandes, para valores de PB pequenos, o intervalo possível mostrado pelos resultados apresenta uma pequena dispersão justamente na faixa de valores que mais interessa.

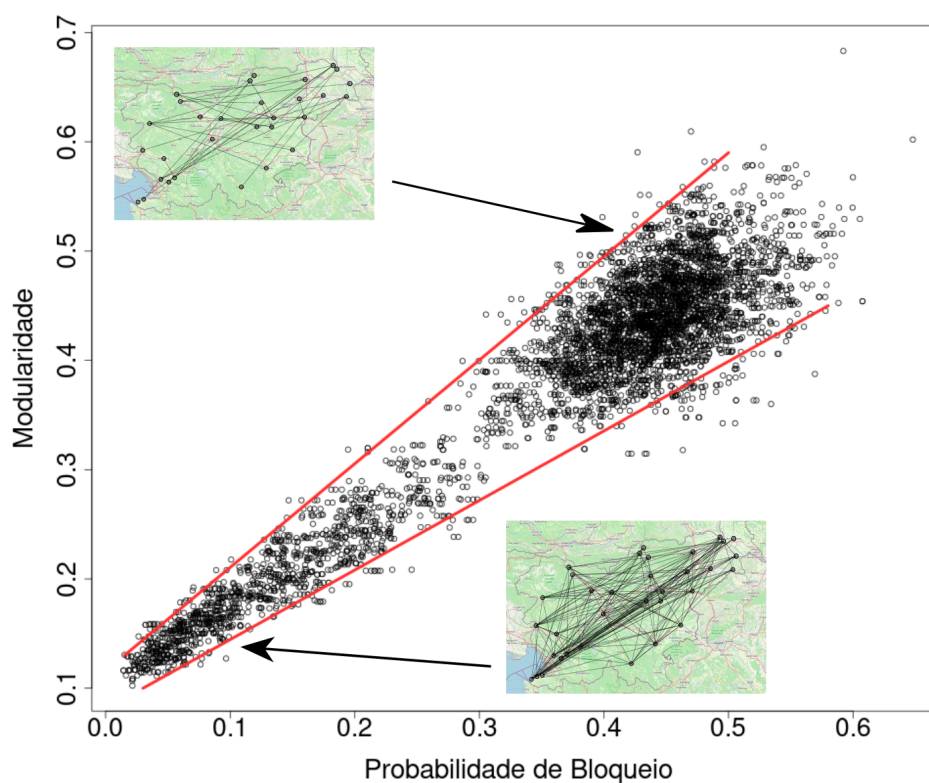


Figura 14 – Comparação entre PB e a Modularidade após DC utilizando algoritmo GMO.
Fonte: Autor

Uma possível aplicação prática seria a utilização desta técnica de DC como primeira etapa da função de *fitness* sobre uma população de topologias em algoritmos evolucionários. A ideia seria utilizá-lo como forma de detectar rapidamente redes muito ruins (ARAÚJO; BASTOS-FILHO; MARTINS-FILHO, 2015a). Caso ele indicasse um valor de PB abaixo de um limiar, a topologia seria avaliada por uma simulação.

Outro resultado muito interessante é o tempo necessário para fazer o cálculo da PB. Na Figura 15 é possível ver o tempo necessário para calcular a PB através de uma simulação. As topologias foram ordenadas pelo tempo de simulação em ordem crescente para facilitar a visualização. O simulador utilizado foi o (CHAVES et al., 2010). A Figura 16 apresenta a soma do tempo necessário para efetuar a DC e o cálculo das métricas de clusterização nas mesmas topologias. Observe que na simulação a escala é em minutos, porém no algoritmo GMO é milissegundos. A média de tempo para o cálculo no simulador foi $\mu = 506974ms$, enquanto no GMO foi apenas $\mu = 7ms$. Isto significa um ganho de 72000. Mesmo levando em conta a simulação que levou menos tempo, $\Delta_t = 396472ms$, e o tempo para particionamento e cálculo da modularidade que demorou mais $\Delta_t = 88ms$, ainda assim o ganho seria de 4500. Ou seja, a utilização de técnicas de DC para estimativa da PB em RTFOs permitiria um ganho computacional significativo que é de 3 a 4 ordens de grandezas.

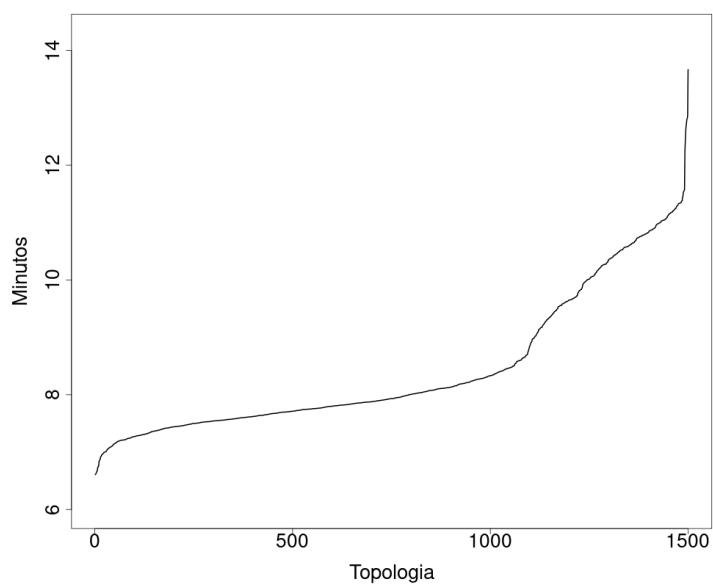


Figura 15 – Tempo necessário para calcular a PB usando simulação. Fonte: Autor

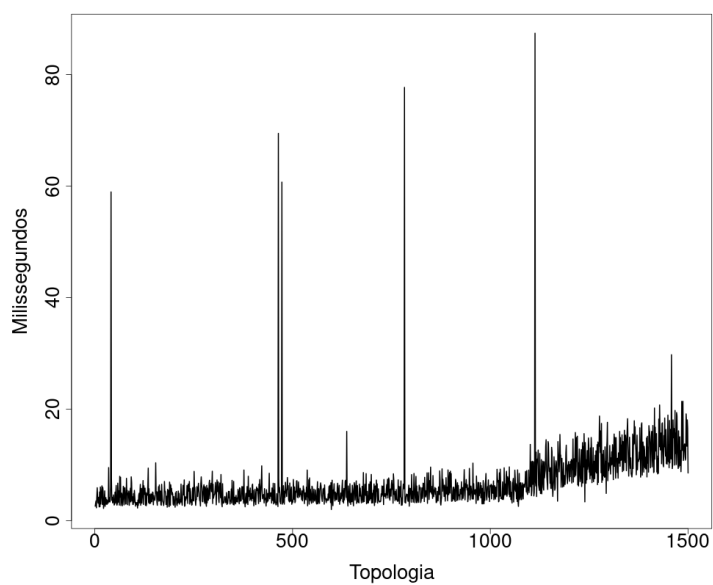


Figura 16 – Tempo necessário para calcular a PB usando o algoritmo *Greedy Modularity Optimization*. Fonte: Autor

5.1.3 Fluid Communities

O algoritmo FC precisa receber como parâmetro o número de clusters. Dessa forma, para apresentar os resultados, foi adotada a nomenclatura FC<k> no qual <k> representa o número de clusters. Assim, FC2 significa o algoritmo *Fluid Communities* com 2 clusters.

Assim como no algoritmo GMO, os resultados com o FC<k> foram promissores, para comparação da PB. Dentre os pontos em comum encontrados nestes e nos resultados utilizando o algoritmo GMO se destacam a baixa dispersão dos pontos para PBs mais baixas e uma maior dispersão quando existe uma maior PB. A presença de uma maior PB em redes que apresentam uma maior capacidade de clusterização, este padrão foi observado tanto nos algoritmos LPA e GMO. Um fenômeno interessante é a variação dos resultados da métrica performance que fica mais achatada para valores menores de k . Ou seja, não apresenta correlação alta para valores $k < 4$, enquanto nas outras métricas independentemente do número de clusters uma correlação alta sempre é mantida.

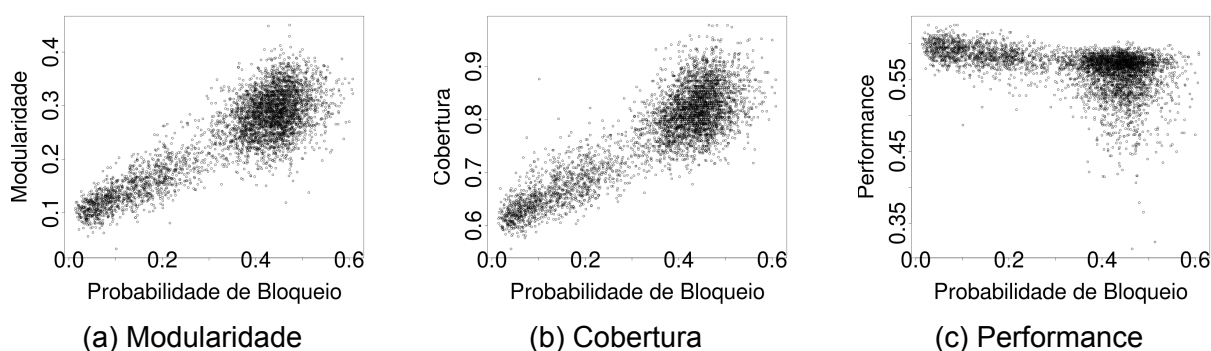


Figura 17 – Comparação entre PB e DC utilizando o algoritmo FC2. Fonte: Autor

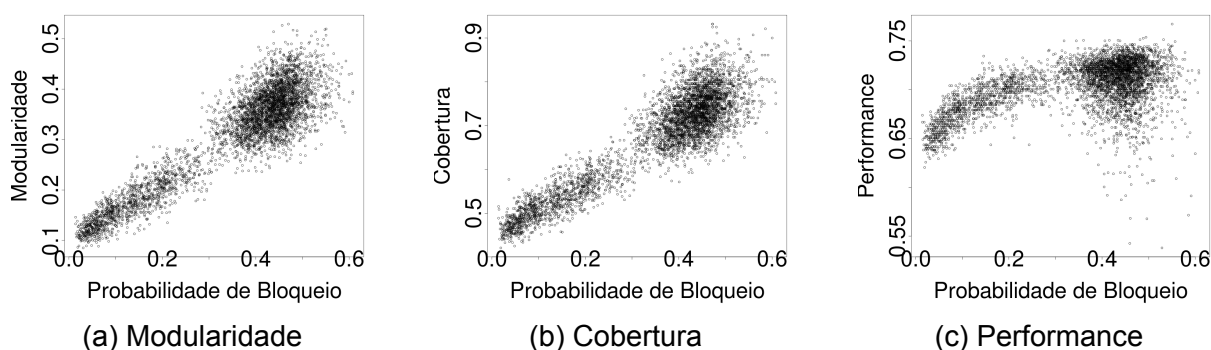


Figura 18 – Comparação entre PB e DC utilizando o algoritmo FC3. Fonte: Autor

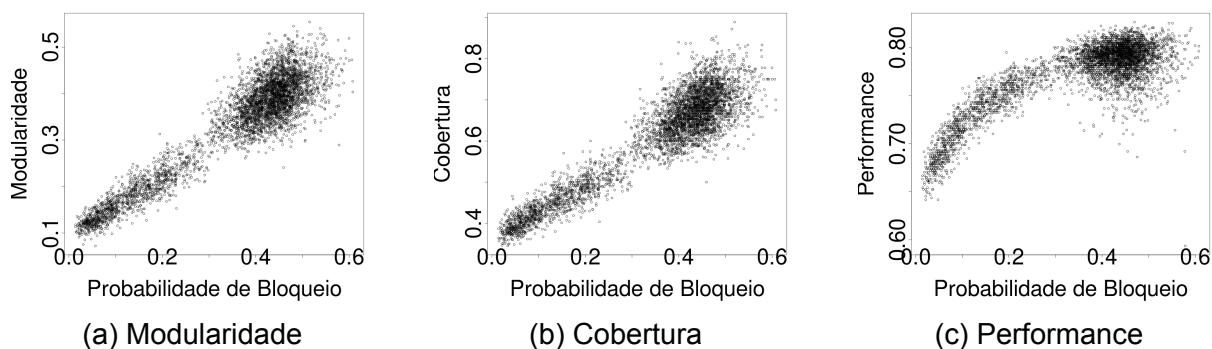


Figura 19 – Comparação entre PB e DC utilizando o algoritmo FC4. Fonte: Autor

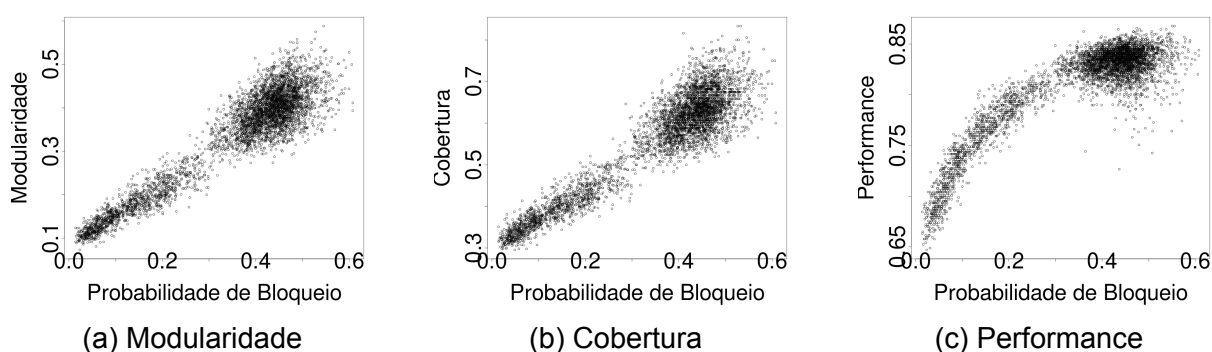


Figura 20 – Comparação entre PB e DC utilizando o algoritmo FC5. Fonte: Autor

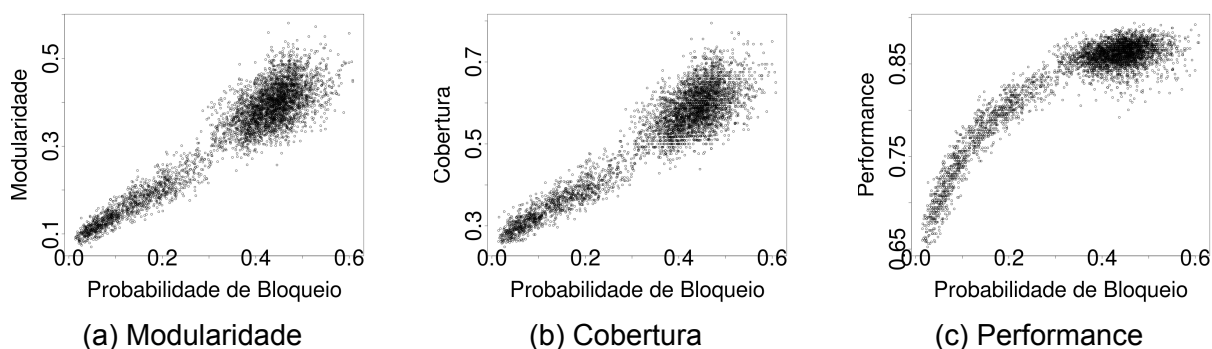


Figura 21 – Comparação entre PB e DC utilizando o algoritmo FC6. Fonte: Autor

Na Figura 22 é possível ver o padrão já mencionado. A métrica performance apresenta uma falta de variação quando há uma variação na PB, isto ocorre para valores pequenos de *clusters*. Porém, ao passo que o número de clusters vai crescendo a curva vai se moldando ao padrão de crescimento das outras métricas. Isso acontece devido ao fato de que um número menor de clusters disponíveis levará a um particionamento que não reflete bem a quantidade de comunidades presentes na topologia. Isto é muito interessante pois demonstra que, de fato, existem relações entre as comunidades e a PB em RTFOs.

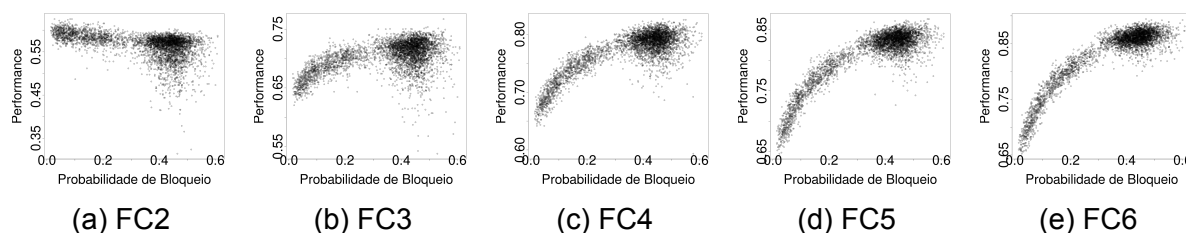


Figura 22 – Variação da performance ao se alterar o número de *clusters* utilizando o algoritmo FC<k>. Fonte: Autor

5.2 Resultados usando Resiliência

Como já mencionado anteriormente, a Resiliência no contexto do presente trabalho é definida pela razão entre o número de demandas afetadas que conseguem encontrar uma rota alternativa e o número total de demandas afetadas e as falhas simuladas foram simples e dupla.

Um detalhe percebido durante a análise dos resultados foi a similaridade entre os gráficos de comparação de Resiliência para falhas simples e dupla. Por exemplo, na Figura 23 é possível ver os resultados para as falhas simples 23a e dupla 23b em uma execução que utiliza o algoritmo LPA. Na Figura 24 é apresentado os gráficos de dispersão para as mesmas topologias, sob execução do algoritmo GMO. Aparentemente, as imagens estão duplicadas, mas na verdade são execuções independentes. O que gera estes resultados parecidos é a similaridade da saída das simulações. As falhas simples e duplas possuem resultados próximos, embora o valor de Resiliência para as falhas duplas seja ligeiramente menor do que para falhas simples. Este aspecto já fornece um *insight* inicial de que, embora as técnicas de DC sejam úteis para aproximar PB, a correlação com a Resiliência talvez seja menor.

Desta forma, como para o caso de Resiliência, os resultados gráficos foram todos semelhantes, a presente seção irá apenas mostrar os resultados para falhas simples.

5.2.1 Label Propagation Algorithm

Assim como no cenário de análise de PB, o algoritmo LPA foi o que apresentou os resultados menos relevantes para o tema proposto. Na Figura 25 é possível ver os resultados da comparação da PB em relação às métricas de avaliação da qualidade da DC nas RTFOs.

Um comportamento interessante observado desde a seção anterior é a diferença nos resultados apresentados pelas diferentes métricas de clusterização. Embora seja o mesmo particionamento, os gráficos de dispersão mostram resultados evidentemente diferentes. É possível de se perceber isto nas Figuras 25b e 25c, as mesmas comuni-

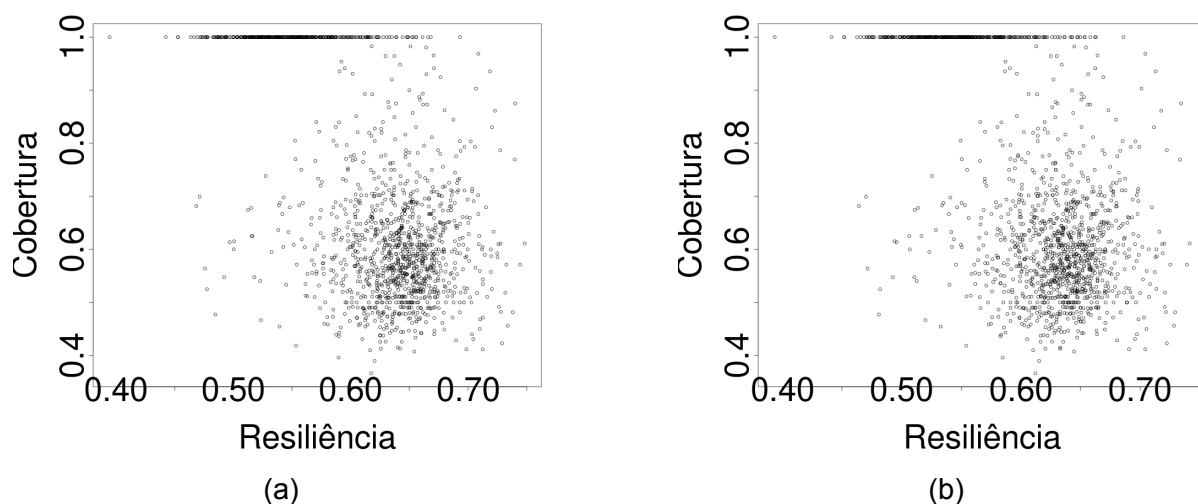


Figura 23 – Comparação entre Resiliência de falhas simples e dupla e a Modularidade utilizando o algoritmo LPA. Fonte: Autor

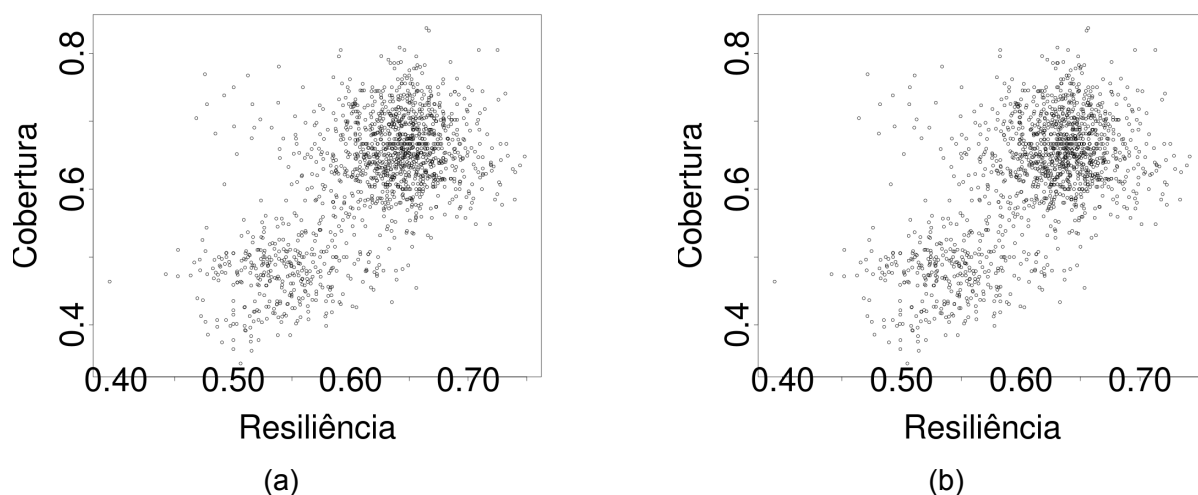


Figura 24 – Comparação entre Resiliência de falhas simples e dupla e a Cobertura utilizando o algoritmo GMO. Fonte: Autor

dades com diferentes resultados. Isso decorre das próprias definições das métricas, e mostra a relevância de um estudo específico sobre a escolha das métricas para uma futura aplicação de DC.

O gráfico que transmite um nível maior de informação é o [25c](#), nele existem dois *clusters* de pontos. As topologias que apresentaram uma maior Resiliência foram as que apresentaram um maior valor da métrica performance. As três métricas utilizadas são de maximização, ou seja, quanto maior o seu valor, maior a qualidade do particionamento, portanto, maior a capacidade que a rede possui de formar comunidades. Logo, redes com maiores capacidades de formarem comunidades, segundo a métrica performance, tendem a apresentarem uma maior Resiliência.

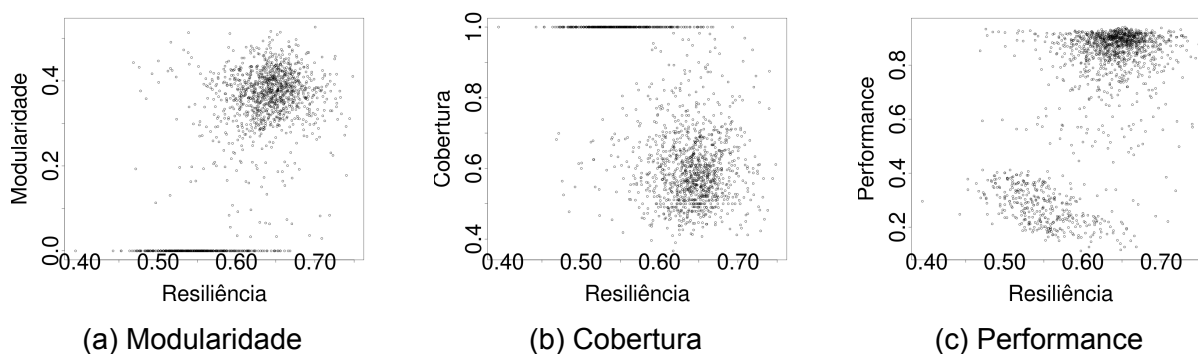


Figura 25 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo LPA. Fonte: Autor

5.2.2 Greedy Modularity Optimization

O algoritmo GMO apresentou resultados interessantes, na Figura 26 é possível ver uma correlação mais clara. Embora eles não estejam concentrados em uma única linha, os gráficos mostram de uma forma mais clara o que foi inferido do algoritmo LPA. Ao passo que a rede apresenta uma maior estrutura de comunidade a sua Resiliência também aumenta.

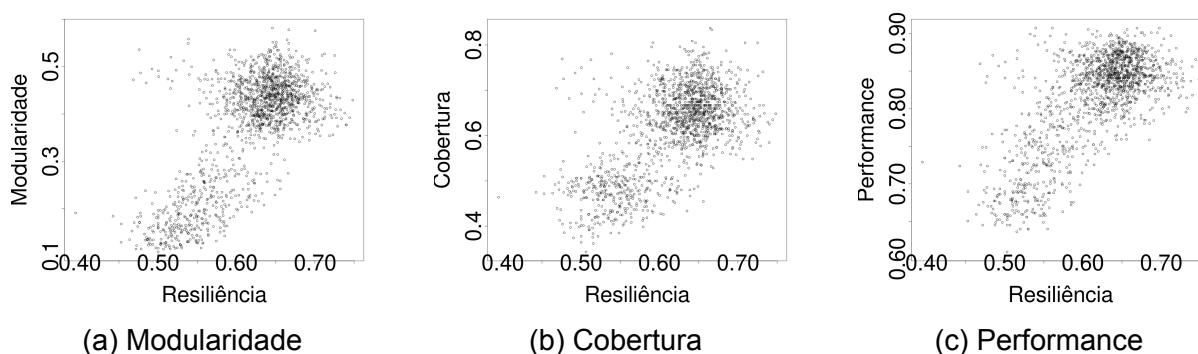


Figura 26 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo GMO. Fonte: Autor

5.2.3 Fluid Communities

Nas Figuras 27 e 28 é possível ver que para a métrica performance independentemente do valor da Resiliência, a métrica de performance não possuía variação. Isso aconteceu devido ao baixo número de comunidades, o que dificultava o algoritmo fazer um particionamento de maior qualidade.

Na Figura 29, é possível ver que a partir de valores $k > 4$, o gráfico que compara a métrica performance com a Resiliência mostrou uma correlação. Assim como nos outros algoritmos já apresentados, existe entre a DC e a Resiliência uma correlação positiva. Ou seja, quanto maior a capacidade de clusterização que uma dada topologia apresenta, maior será a tendência da rede apresentar uma melhor resiliência à falha.

Isso possivelmente decorre da propriedade de resiliência de uma RTFO, que depende da existência de rotas alternativas. Redes que apresentam um maior nível de clusterização, muito provavelmente possuirão um número maior de rotas alternativas, visto que, uma comunidade consiste em uma região de alta conexão entre os nós.

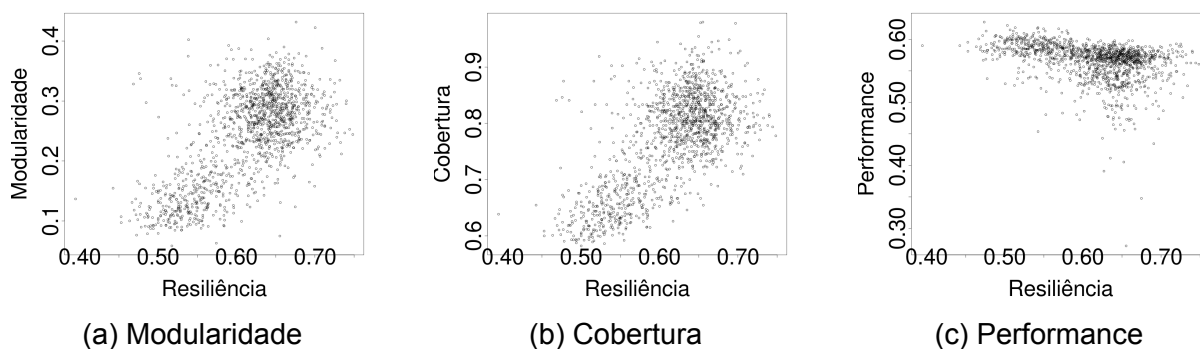


Figura 27 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo FC2. Fonte: Autor

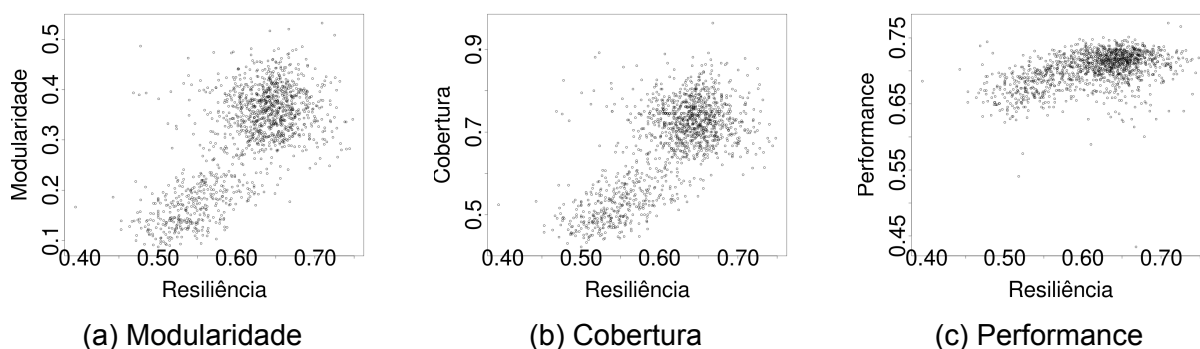


Figura 28 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo FC3. Fonte: Autor

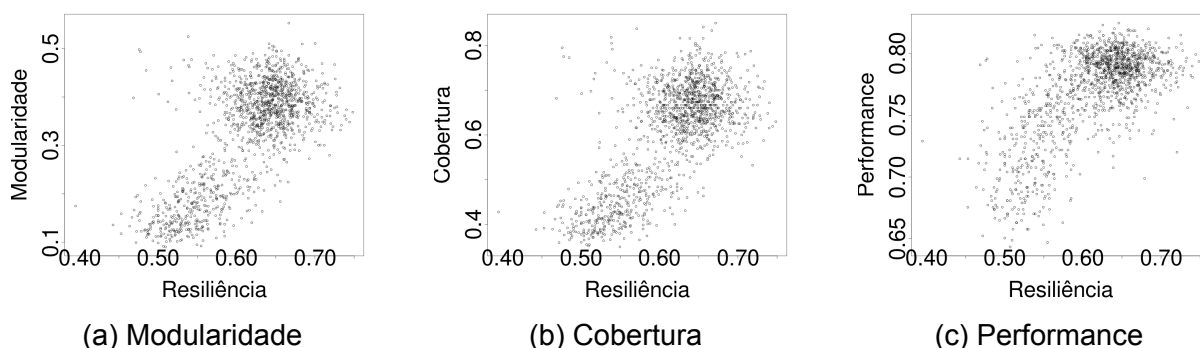


Figura 29 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo FC4. Fonte: Autor

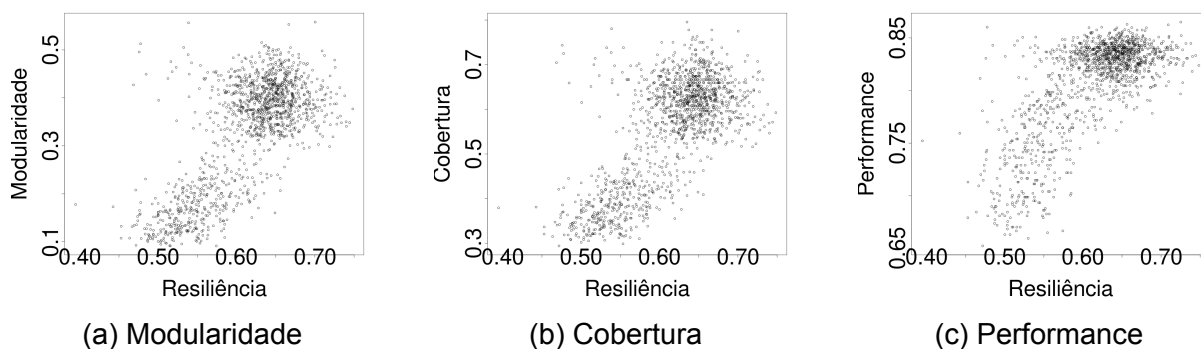


Figura 30 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo FC5. Fonte: Autor

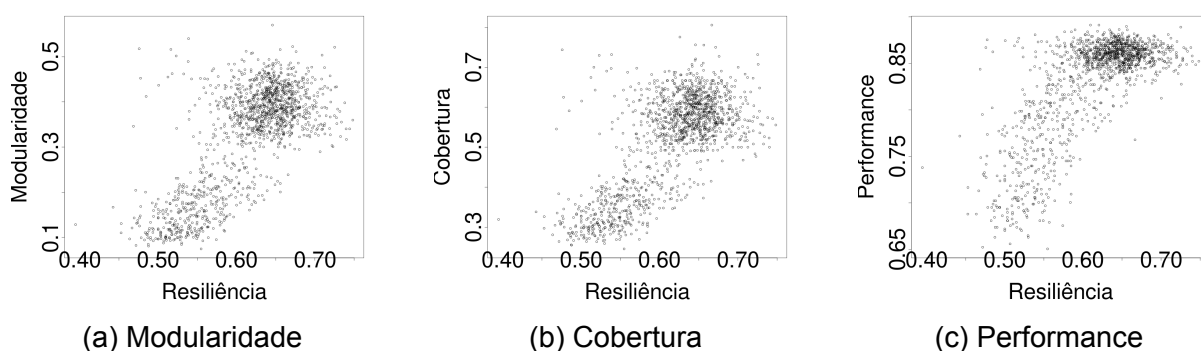


Figura 31 – Comparação entre Resiliência de falhas simples e DC utilizando o algoritmo FC6. Fonte: Autor

Outro ponto interessante é o fato de que para a Resiliência as métricas de modularidade apresentaram resultados similares, independentemente da quantidade de *clusters*. De forma geral, desses resultados apresentados, é nítida a existência de uma correlação entre as métricas de avaliação da qualidade de clusterização de uma DC e da Resiliência em uma RTFO.

6 Conclusão e Trabalhos Futuros

Neste capítulo serão apresentadas as conclusões a respeito dos resultados e possíveis trabalhos futuros. Na Seção 6.1 será feita uma recapitulação da proposta inicial e dos objetivos gerais e específicos da pesquisa. Na Seção 6.2 será feita a conclusão dos resultados sob a ótica dos objetivos da pesquisa. Por fim, a Seção 6.3 apresentará os trabalhos futuros que darão continuidade à pesquisa do tópico apresentado.

6.1 Detecção de Comunidades para Obtenção da Resiliência e Probabilidade de Bloqueio

A DC tem como objetivo o particionamento de um grafo em subgrafos nos quais o número de conexões para nós que pertencem ao subgrafo é maior do que o número de conexões para regiões do resto do grafo (MOSCATO; SPERLI, 2021). Este trabalho teve como objetivo a investigação da existência de correlação entre a Resiliência e PB em RTFOs e a capacidade que essas redes possuem de formar comunidades.

Nesse sentido foram utilizadas técnicas de DC para avaliar se havia padrões entre esses conceitos. A estratégia usada foi a comparação direta entre as métricas de clusterização e a Resiliência e PB que foram calculadas por simulações. A comparação se deu através de gráficos de dispersão que permitiam efetuar uma análise visual sobre as possíveis relações existentes entre as comunidades e os indicadores de desempenho das RTFOs.

O objetivo geral da pesquisa foi investigar a possibilidade da utilização de técnicas de DC para obter informações de uma RTFO, avaliando se há alguma correlação entre a Resiliência e PB e a capacidade que a rede possui de formar comunidades. Os objetivos específicos foram o desenvolvimento de uma plataforma para execução dos experimentos e a investigação de possíveis relações entre o nível de clusterização e a Resiliência e PB em RTFOs.

6.2 Conclusão dos Resultados

Os resultados de comparação para com a PB mostraram um padrão muito claro de correlação entre as métricas de clusterização e a PB. Enquanto a comparação para com a Resiliência, embora, também tenha mostrado padrões de correlação, em alguns

algoritmos de clusterização os resultados não apresentaram aparentes relações entre os dados.

Um ponto muito interessante é o fato de que diferentes métricas de clusterização apresentaram diferentes perspectivas de análise. Ou seja, para o mesmo particionamento, os gráficos de dispersão apresentaram padrões ligeiramente diferentes. Dessa forma, conclui-se que a aplicação prática dessas técnicas leve em consideração as particularidades de cada métrica, e idealmente, se utilize de múltiplas métricas para obtenção de resultados mais precisos.

Além disso, dos resultados apresentados, observou-se também a diminuição de precisão da estimativa para valores maiores de PB. Isto, entretanto, não ocasionaria a inutilidade da técnica, visto que, para projetos reais de DTF os valores aceitáveis de PB são pequenos, $PB < 1\%$.

O custo computacional também foi considerado na pesquisa. Observou-se que o tempo necessário para o particionamento da rede pelo algoritmo GMO foi menor do que o tempo necessário usado pela simulação. O ganho obtido levando em conta o melhor tempo da simulação e o pior tempo utilizando a DC foi de 4500. Isto ocorre devido a necessidade que a simulação precisa de efetuar vários eventos para calcular a PB. Embora os resultados no gráfico de dispersão tenham apresentado um nível de imprecisão, a velocidade é tão superior a simulação que justificaria, por si só, uma análise mais aprofundada da utilidade das técnicas de DC para obtenção de informações relevantes a respeito das RTFOs.

De forma geral, o presente trabalho teve como fundamento motivador o questionamento da existência de correlação entre a capacidade que uma RTFO possui de formar comunidades e a sua qualidade; que é comumente avaliada através de indicadores de desempenho como a PB e a Resiliência. Portanto, de tudo o que foi apresentado, observando os objetivos da pesquisa e os resultados obtidos, conclui-se que existe uma correlação entre o nível de clusterização de uma RTFO e os seus indicadores de desempenho. É possível, de fato, extrair informações úteis para os projetos de DTFs se utilizando de técnicas de Detecção de Comunidades.

Dentre os benefícios desta abordagem estão a simplicidade para aplicação industrial, visto que não há necessidade de gerar um *dataset* para treinamento de modelos de Inteligência Artificial. E também há o benefício da eficiência computacional, que justifica a sua pesquisa e utilização prática.

6.3 Trabalhos Futuros

Como já mencionado, a presente pesquisa teve como escopo uma análise exploratória e investigativa da possibilidade de existência de correlação entre as comunidades de uma RTFO e a sua qualidade em regime de operação. A conclusão do trabalho é que sim, existe relação entre esses dois conceitos, e é possível se utilizar de técnicas de DC para inferir a qualidade de uma RTFO.

Porém, ainda assim outras análises e pesquisas precisam ser conduzidas para um melhor entendimento do domínio em questão. Portanto, a seguir é apresentada uma lista com sugestões de trabalhos futuros:

- **Extensão dos experimentos realizados para outras redes:** O presente trabalho, se utilizou de *datasets* sintetizados a partir da topologia ARNES (KNIGHT et al., 2011). Sendo assim, uma extensão para outras topologias permitiria avaliar os resultados sob condições diferentes, como por exemplo, um número diferente de nós.
- **Estudo de outras métricas de redes complexas:** Utilização de outras medidas de redes complexas de forma individual para predição ou combinação dessas medidas com as métricas de DC para treinamento de um modelo de RNA e assim melhorar a precisão.
- **Análise das amplitudes dos resultados:** Para aplicações práticas um fator muito importante é conhecer o intervalo no qual uma determinada métrica irá variar. O fato de a PB poder variar em um intervalo $0 \leq PB \leq 1$ não significa que uma métrica como a modularidade também irá necessariamente variar nesse intervalo. Portanto uma análise das amplitudes de cada métrica é imprescindível para possibilitar aplicações práticas.
- **Extensão da investigação para outros indicadores de qualidade de RTFOs:** A pesquisa atual avaliou a correlação entre o nível de clusterização e a PB e Resiliência. Porém ainda existem outros indicadores de desempenho de uma rede que podem ser utilizadas.
- **Extensão da investigação para redes sob diferentes níveis de carga:** Uma investigação que é muito relevante e importante no contexto de RTFOs é o desempenho de redes sob diferentes níveis de carga. Desta forma, uma investigação levando em conta diferentes níveis de carga poderia ser conduzida, observando também diferentes padrões de distribuição de carga dentre os nós da rede.

- **Análise de correlação usando outros instrumentos:** Poderia ser conduzida uma análise utilizando outros instrumentos de avaliação de correlação, como o uso de indicadores de correlação estatísticos como o de Pearson.
- **Estudo da combinação de mais de um indicador de clusterização:** Uma pesquisa interessante seria a combinação das métricas de clusterização utilizando técnicas de regressão simples, como a regressão linear.
- **Integração da abordagem da análise proposta em motores de otimização baseados em meta-heurísticas:** A utilização de técnicas de DC permite um ganho computacional considerável se comparado com as simulações. Nesse sentido, a integração desta abordagem em motores de otimização que são baseados em meta-heurísticas poderia trazer ganhos no tempo de execução.

Referências

- ARAÚJO, D.; BARROS, J. Estimativa de probabilidade de bloqueio de redes Ópticas elásticas usando redes neurais artificiais. In: *19º SBMO ? Simpósio Brasileiro de Micro-ondas e Optoeletrônica*. [s.n.], 2020. p. 1–6. ISBN 978-65-89532-00-2. Disponível em: <<https://drive.google.com/file/d/1awjFbTcPKIOPFysm7B4sbeM2LxB-AcKG/view?ts=602fbb6b>>. Citado 9 vezes nas páginas 7, 14, 15, 26, 27, 29, 32, 35 e 36.
- ARAÚJO, D. R. et al. Surrogate models assisted by neural networks to assess the resilience of networks. In: IEEE. *2017 IEEE Latin American Conference on Computational Intelligence (La-Cci)*. [S.l.], 2017. p. 1–6. Citado 2 vezes nas páginas 29 e 32.
- ARAÚJO, D. R.; BASTOS-FILHO, C. J.; MARTINS-FILHO, J. F. Artificial neural networks to estimate blocking probability of transparent optical networks: A robustness study for different networks. In: IEEE. *2015 17th International Conference on Transparent Optical Networks (ICTON)*. [S.l.], 2015. p. 1–4. Citado 3 vezes nas páginas 13, 15 e 16.
- ARAÚJO, D. R. B. d. Planejamento de redes ópticas usando inteligência computacional e ciência das redes. Universidade Federal de Pernambuco, 2015. Citado na página 26.
- ARAÚJO, D. R. D.; BASTOS-FILHO, C. J.; MARTINS-FILHO, J. F. An evolutionary approach with surrogate models and network science concepts to design optical networks. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 43, p. 67–80, 2015. Citado 7 vezes nas páginas 13, 14, 16, 29, 31, 32 e 43.
- ARAÚJO, D. R. de; BASTOS-FILHO, C. J.; MARTINS-FILHO, J. F. Methodology to obtain a fast and accurate estimator for blocking probability of optical networks. *Journal of Optical Communications and Networking*, Optical Society of America, v. 7, n. 5, p. 380–391, 2015. Citado 5 vezes nas páginas 28, 29, 32, 41 e 42.
- BANERJEE, N.; KUMAR, R. Multiobjective network design for realistic traffic models. In: *Proceedings of the 9th annual conference on Genetic and evolutionary computation*. [S.l.: s.n.], 2007. p. 1904–1911. Citado na página 26.
- BARROS, J. *Community Detection Research Initiative*. 2021. <<https://github.com/JonasFreireAlcantara/community-detection-research>>. Citado na página 40.
- BOCCALETTI, S. et al. Complex networks: Structure and dynamics. *Physics reports*, Elsevier, v. 424, n. 4-5, p. 175–308, 2006. Citado 2 vezes nas páginas 13 e 19.
- BRANDES, U. et al. On finding graph clusterings with maximum modularity. In: SPRINGER. *International Workshop on Graph-Theoretic Concepts in Computer Science*. [S.l.], 2007. p. 121–132. Citado 2 vezes nas páginas 21 e 23.
- BRANDES, U.; GAERTLER, M.; WAGNER, D. Experiments on graph clustering algorithms. In: SPRINGER. *European Symposium on Algorithms*. [S.l.], 2003. p. 568–579. Citado na página 25.
- CHAVES, D. A. et al. Simton: A simulator for transparent optical networks. *Journal of Communication and Information Systems*, v. 25, n. 1, 2010. Citado 5 vezes nas páginas 15, 27, 30, 35 e 43.

CLAUSET, A.; NEWMAN, M. E.; MOORE, C. Finding community structure in very large networks. *Physical review E*, APS, v. 70, n. 6, p. 066111, 2004. Citado 3 vezes nas páginas 19, 21 e 22.

FORTUNATO, S. Community detection in graphs. *Physics reports*, Elsevier, v. 486, n. 3-5, p. 75–174, 2010. Citado 4 vezes nas páginas 19, 24, 25 e 30.

HAGBERG, A.; SWART, P.; CHULT, D. S. *Exploring network structure, dynamics, and function using NetworkX*. [S.l.], 2008. Citado 2 vezes nas páginas 20 e 40.

JONNALAGADDA, A.; KUPPUSAMY, L. Overlapping community detection in social networks using coalitional games. *Knowledge and Information Systems*, Springer, v. 56, n. 3, p. 637–661, 2018. Citado na página 19.

JONSSON, P. F. et al. Cluster analysis of networks generated through homology: automatic identification of important protein communities involved in cancer metastasis. *BMC bioinformatics*, Springer, v. 7, n. 1, p. 1–13, 2006. Citado 3 vezes nas páginas 13, 30 e 32.

KNIGHT, S. et al. The internet topology zoo. *IEEE Journal on Selected Areas in Communications*, IEEE, v. 29, n. 9, p. 1765–1775, 2011. Citado 2 vezes nas páginas 35 e 54.

LEWIS, T. G. *Network science: Theory and applications*. [S.l.]: John Wiley & Sons, 2011. Citado na página 13.

LI, Q.; DONG, S.; MOSTAFAVI, A. Community detection in actor collaboration networks of resilience planning and management in interdependent infrastructure systems. In: AMERICAN SOCIETY OF CIVIL ENGINEERS RESTON, VA. *Construction Research Congress 2020: Infrastructure Systems and Sustainability*. [S.l.], 2020. p. 675–683. Citado 3 vezes nas páginas 30, 31 e 32.

LIRA, C. et al. Redes neurais artificiais para predição de tolerância a falhas em redes Ópticas. In: *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*. Porto Alegre, RS, Brasil: SBC, 2019. p. 377–388. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/eniac/article/view/9299>>. Citado 8 vezes nas páginas 13, 14, 15, 26, 27, 29, 32 e 35.

LIU, F. et al. Deep learning for community detection: progress, challenges and opportunities. *arXiv preprint arXiv:2005.08225*, 2020. Citado na página 14.

MOSCATO, V.; SPERLÌ, G. A survey about community detection over on-line social and heterogeneous information networks. *Knowledge-Based Systems*, Elsevier, p. 107112, 2021. Citado 6 vezes nas páginas 13, 14, 18, 19, 24 e 52.

NEWMAN, M. E.; GIRVAN, M. Finding and evaluating community structure in networks. *Physical review E*, APS, v. 69, n. 2, p. 026113, 2004. Citado na página 24.

PANDEY, N.; PAL, A. et al. Impact of digital surge during covid-19 pandemic: A viewpoint on research and practice. *International journal of information management*, Elsevier, v. 55, p. 102171, 2020. Citado 2 vezes nas páginas 25 e 26.

PARÉS, F. et al. Fluid communities: A competitive, scalable and diverse community detection algorithm. In: SPRINGER. *International Conference on Complex Networks and their Applications*. [S.l.], 2017. p. 229–240. Citado 4 vezes nas páginas 7, 19, 22 e 23.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>. Citado na página 40.

RAGHAVAN, U. N.; ALBERT, R.; KUMARA, S. Near linear time algorithm to detect community structures in large-scale networks. *Physical review E*, APS, v. 76, n. 3, p. 036106, 2007. Citado 5 vezes nas páginas 19, 20, 21, 23 e 24.

RAMASWAMI, R.; SIVARAJAN, K.; SASAKI, G. *Optical networks: a practical perspective*. [S.l.]: Morgan Kaufmann, 2009. Citado 2 vezes nas páginas 26 e 27.

ROCCO, C. M. et al. Community detection and resilience in multi-source, multi-terminal networks. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, SAGE Publications Sage UK: London, England, v. 232, n. 6, p. 616–626, 2018. Citado 4 vezes nas páginas 15, 30, 31 e 32.

SHUN, J. et al. Parallel local graph clustering. *arXiv preprint arXiv:1604.07515*, 2016. Citado na página 18.

STROGATZ, S. H. Exploring complex networks. *nature*, Nature Publishing Group, v. 410, n. 6825, p. 268–276, 2001. Citado na página 19.

TOPOLOGY-ZOO. *ARNES (The Academic and Research Network of Slovenia)*. <<http://www.topology-zoo.org/maps/Arnes.jpg>>. Acesso em 30/10/2021. Citado 2 vezes nas páginas 7 e 36.

TOPOLOGY-ZOO. *CARNET: Croatian Academic and Research Network*. <<http://www.topology-zoo.org/maps/Carnet.jpg>>. Acesso em 19/07/2021. Citado 2 vezes nas páginas 7 e 14.

TOPOLOGY-ZOO. *Kentucky Datalink*. <<http://www.topology-zoo.org/maps/Kdl.jpg>>. Acesso em 19/07/2021. Citado 2 vezes nas páginas 7 e 15.

VISUALIZATION python. *Folium*. 2020. Disponível em: <<https://python-visualization.github.io/folium/>>. Citado na página 40.

ZACHARY, W. W. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, University of New Mexico, v. 33, n. 4, p. 452–473, 1977. Citado na página 13.

ZENG, X. et al. A consensus community-based particle swarm optimization for dynamic community detection. *IEEE transactions on cybernetics*, IEEE, v. 50, n. 6, p. 2502–2513, 2019. Citado na página 19.

ZHU, K.; MUKHERJEE, B. Traffic grooming in an optical wdm mesh network. *IEEE Journal on selected areas in communications*, IEEE, v. 20, n. 1, p. 122–133, 2002. Citado na página 26.