UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

CENTRO TECNOLÓGICO

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

PROJETO DE GRADUAÇÃO

LEONARDO LEAL REBLIN

ANÁLISE DE ESTRUTURAS DE COMUNIDADE EM REDES ÓPTICAS DE COMUNICAÇÃO

VITÓRIA
2019

LEONARDO LEAL REBLIN

ANÁLISE DE ESTRUTURAS DE COMUNIDADE EM REDES ÓPTICAS DE COMUNICAÇÃO

Parte manuscrita do Projeto de Graduação do aluno **Leonardo Leal Reblin**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito para obtenção do grau de Engenheiro Eletricista.

Orientadora: Profa. Dra. Marcia Helena Moreira Paiva

VITÓRIA
2019

LEONARDO LEAL REBLIN

ANÁLISE DE ESTRUTURAS DE COMUNIDADE EM REDES ÓPTICAS DE COMUNICAÇÃO

Parte manuscrita do Projeto de Graduação do aluno **Leonardo Leal Reblin**, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito para obtenção do grau de Engenheiro Eletricista.

Aprovada em 16 de dezembro de 2019.

COMISSÃO EXAMINADORA:

Profa. Dra. Marcia Helena Moreira Paiva
Universidade Federal do Espírito Santo
Orientadora

Profa. Dra. Daniela Bertolini Depizzol
Instituto Federal do Espírito Santo
Examinadora

Prof. Dr. Moisés Renato Nunes Ribeiro
Universidade Federal do Espírito Santo
Examinador

Dedico este trabalho à minha família, aos meus amigos e a Deus.

RESUMO

Projetar uma rede óptica eficiente costuma ser um grande desafio, pois requer um amplo conhecimento da área e uma definição clara do “comportamento” que se deseja obter da rede. Focando, por exemplo, na topologia da rede, seria desejável descobrir quais os parâmetros topológicos corresponderiam a características e a comportamentos de interesse. Porém não é tão trivial a escolha de tais parâmetros. Em redes ópticas, vários sinais são transportados em uma mesma fibra, cada um com seu comprimento de onda. Uma das formas eficientes e desejadas de se transportar esses sinais é usando o número mínimo de diferentes comprimentos de onda, que pode ser influenciado por parâmetros topológicos estabelecidos no projeto da rede. Entretanto, não é simples determinar exatamente qual é a relação entre parâmetros topológicos e o número de comprimentos de onda (λ) em redes ópticas. Uma possível modelagem para redes ópticas se dá por meio de grafos, que são estruturas compostas por um número de vértices (n) e de arestas (m), a qual é usada no presente trabalho. No presente trabalho são aplicados conceitos e técnicas de Ciência de Dados, pois será analisado um conjunto muito grande de dados (aproximadamente $2,2 \times 10^6$ grafos). Utilizando ferramentas como banco de dados e linguagens de programação - como a linguagem R, por exemplo -, este estudo visa obter algum conhecimento sobre esses dados brutos e atingir os melhores resultados. Baseando-se na hipótese de que o valor de λ é influenciado pela topologia da rede, este trabalho propõe aplicar nove métodos de detecção de comunidades no conjunto de aproximadamente $2,2 \times 10^6$ grafos (com valores de λ já calculados). Para cada divisão de comunidades de cada grafo desse conjunto, calcula-se o índice de modularidade. Logo, cada grafo fica com nove valores de modularidade, os quais são analisados em conjunto com λ. Com isso, busca-se identificar o tipo de relação que existe entre o número de comprimentos de onda e a divisão de um grafo em comunidades. Para uma maior fidedignidade dos resultados, o conjunto de grafos usado é uma amostra de redes aleatórias que imitam redes reais. Como resultado, destaca-se uma melhor precisão na escolha de topologias de rede via grafos que minimizam o número de comprimentos de onda, através do parâmetro modularidade, obtido pela aplicação dos métodos de detecção de comunidades que apresentaram maior correlação entre seus valores de modularidade e o número de comprimentos de onda.

Palavras-chave: Redes ópticas. Topologia. Comprimento de onda. Comunidades. Grafo. Ciência de Dados.

ABSTRACT

Design an efficient optical network is a great challenge, because it requires an extensive knowledge in this area and a clear definition of the desired “behavior” from the network. For example, focusing on the network topology, it would be desirable to find which topological parameters would satisfy the characteristics and the behaviors of interest. However, it is not so trivial to choose those parameters. In optical networks, multiple signals are transported on the same fiber, each with its wavelength. An efficient and desired way to transport those signals is using the minimum number of different wavelengths. But it’s not simple to determine, exactly, what is the relationship between topological parameters and the number of wavelengths (λ) in optical networks. A possible modeling of optical networks is given by graphs, which are structures composed by a number of vertices (n) and a number of edges (m), which is used on this study. In this project, concepts and techniques of Data Science are used beacause of the big data volume (nearly $2,2 \times 10^6$ graphs). Said that, using computational tools like data bases and program languages, this study aims to get some knowledge about this data set and to obtain the best results. Based on the hypothesis that the value of λ is influenced by the network topology, this study aims to apply nine different community detection methods on a set of nearly $2,2 \times 10^6$ graphs (with λ values already known). For each community division of each graph, the modularity index is calculated. So each graph will have nine values of modularity, which will be analysed with λ. Said that, it aims to identify a relationship between the number of wavelengths and the division of a graph into communities. For a better trustworthiness, the set of graphs used is a sample of random networks that imitate real ones. As a result, there is better accuracy in choosing network topologies represented by graphs that minimize the number of wavelengths through modularity, applying the community detection methods that showed the best relationships between their modularity and number of wavelengths.

Keywords: Optical networks. Topology. Wavelength. Communities. Graph. Data Science.

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE ABREVIATURAS E SIGLAS

CNM	*Clauset-Newman-Moore*
EON	*Elastic Optical Networks*
GN	*Girvan-Newman*
JSON	*JavaScript Object Notation*
OOFDM	*Optical Orthogonal Frequency Division Multiplexing*
RWA	*Routing and Wavelength Assignment*
WDM	*Wavelength Division Multiplexing*

SUMÁRIO

1 INTRODUÇÃO

1.1 Apresentação

As Redes Ópticas de Comunicação são, atualmente, a base para a comunicação digital mundial devido a vários fatores, como: alta velocidade de propagação, capacidade de tráfego e alcance. Além desses fatores, esse tipo de rede traz como vantagem o compartilhamento de vários canais independentes em uma mesma fibra óptica, ou seja, vários sinais podem ser transportados em uma mesma fibra simultaneamente, usando diferentes comprimentos de onda. O número de comprimentos de onda necessários para estabelecer comunicação *all-to-all* (todos querendo se comunicar com todos) é representado por λ. Esse número é uma variável de interesse no projeto de redes ópticas de comunicação (DEPIZZOL et al., 2018).

Uma forma intuitiva de representar uma rede óptica de comunicação é através de grafos, como pode ser visto na Seção 2.2. Um grafo é denotado por G(V,E), ou apenas G, onde V é um conjunto de vértices e E é um conjunto de arestas (DIESTEL, 2016). Os vértices correspondem aos nós da rede, e as arestas aos *links* (conexões), então as arestas conectam os vértices da mesma forma que os *links* conectam os nós. O número de vértices é a ordem do grafo, denotado por $n = |V|$. O número de arestas é o tamanho do grafo, denotado por $m = |E|$ (DEPIZZOL et al., 2018).

Usando grafos, é possível calcular os invariantes dos grafos, ou invariantes topológicos, os quais são parâmetros numéricos que não mudam, independentemente de como são rotulados os vértices e as arestas. Os invariantes são importantes porque eles representam os parâmetros topológicos do grafo e, consequentemente, da rede óptica de comunicação. Tendo isso, neste estudo esse tipo de parâmetro é calculado para entender como a topologia das redes influencia o número de comprimentos de onda (λ) (DEPIZZOL et al., 2018).

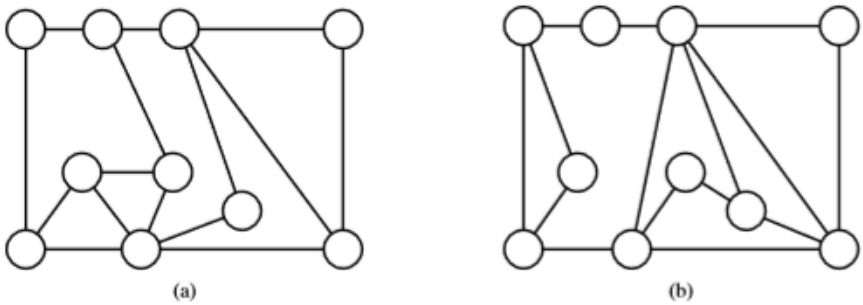
O presente estudo propõe uma análise topológica de redes ópticas de comunicação. Em particular, será feita um estudo da estrutura de comunidades dessas redes. Serão apresentados, no próximo capítulo, alguns métodos de detecção de comunidades (ou

métodos de *clusterização*), que serão aplicados a essas redes.

Após a aplicação dos métodos, é possível verificar como cada rede foi “clusterizada”, por meio do cálculo da modularidade e, com ela verificar se a maneira como foram “clusterizadas” influencia em questões de interesse da rede. Como métrica de análise, destaca-se o número mínimo de comprimentos de onda necessários para uma comunicação *all-to-all*, no caso particular do presente trabalho.

Para ilustrar como o número de comprimentos de onda depende da topologia da rede, a Figura 1 mostra duas topologias diferentes com o mesmo número de nós e de arestas, mas com números de comprimentos de onda diferentes.

Figura 1 – Duas topologias diferentes com o mesmo número de nós ($n = 10$) e arestas ($m = 15$), sendo que a topologia (a) requer 8 comprimentos de onda e a topologia (b) requer 12 comprimentos de onda



Fonte: Depizzol e outros (2018).

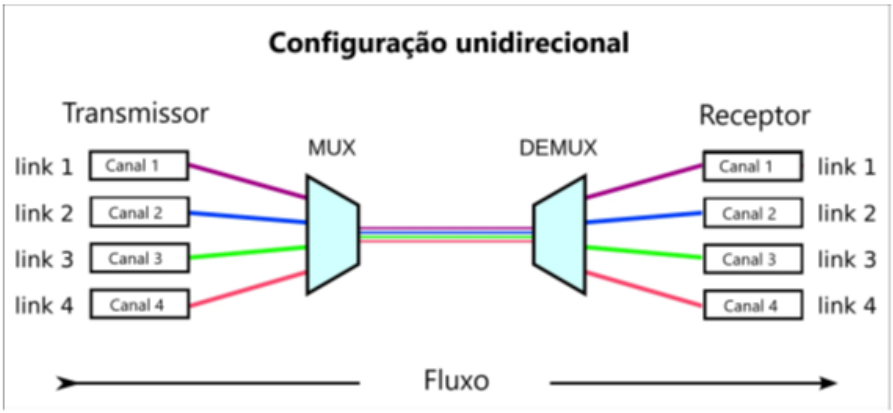
1.2 Multiplexação por Divisão de Comprimento de Onda (WDM)

A tecnologia que permite o roteamento da rede óptica por comprimentos de onda é a Multiplexação por Divisão de Comprimento de Onda (WDM, do inglês *Wavelength Division Multiplexing*). A WDM é uma técnica onde sinais ópticos em diferentes comprimentos de onda são combinados, transmitidos juntos em uma mesma fibra e, por fim, separados novamente. Há duas configurações básicas de sistemas de transmissão WDM: transmissão unidirecional e bidirecional, como mostrado na Figura 2. Ambas requerem uma única fibra óptica, várias fontes de luz e fotodetectores, e dispositivos multi/demultiplexadores ópticos. O sistema de transmissão unidirecional requer um multiplexador (MUX) e um demultiplexador (DEMUX), enquanto que o bidirecional requer um multi/demultiplexador (MUX/DEMUX) em cada ponta da fibra (ISHIO; MINOWA;

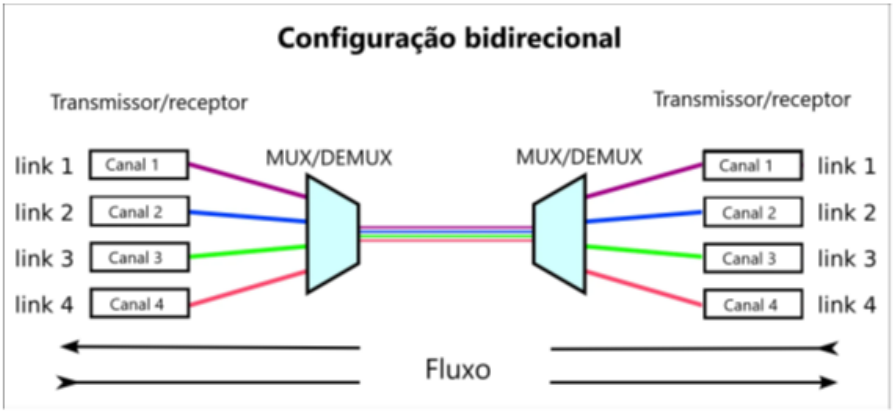
NOSU, 1984).

As vantagens dos sistemas WDM são: aumento da capacidade de transmissão por fibra, redução de custos do sistema, transmissão simultânea de diferentes sinais e capacidade de expansão do canal de serviço após a instalação da fibra. Essa é a razão pela qual a tecnologia WDM é amplamente aplicada a sistemas em vários campos de comunicação (ISHIO; MINOWA; NOSU, 1984).

Figura 2 – (a) Configuração do sistema de transmissão WDM unidirecional. (b) como em (a), porém bidirecional



(a)



(b)

Fonte: Produção do próprio autor.

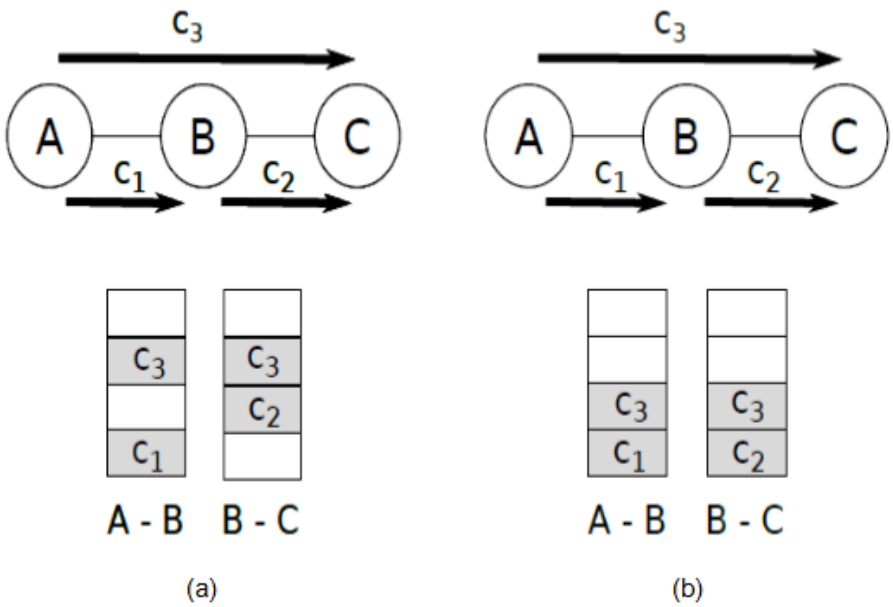
Nesse contexto, surge o problema de Roteamento e Atribuição de Comprimentos de Onda (RWA). Há várias abordagens para solucionar tal problema, mas a que separa claramente o RWA em dois subproblemas é predominante: roteamento de demanda, seguido da alocação de comprimento de onda para os canais ópticos, com o objetivo de minimizar o número de comprimentos de onda para atender a demanda pelo roteamento atribuído (DEPIZZOL et al., 2018).

1.3 Número de comprimentos de onda

Um dos objetivos do projeto de uma rede óptica é minimizar o número de comprimentos de onda (λ) necessários para estabelecer as demandas de tráfego, que está diretamente relacionado ao custo e à capacidade das redes.

No presente estudo, λ é calculado usando o modelo de programação inteira, proposto por (COUSINEAU et al., 2015) para resolver o RWA. A Figura 3 explicita o conceito de número de comprimentos de onda. Na Figura 3(a) três canais são transmitidos usando três comprimentos de onda, e na Figura 3(b) os mesmos três canais são transmitidos com apenas dois comprimentos de onda, que é a menor quantidade possível de comprimentos de onda que consegue atender a demanda de tráfego, nesse caso, $\lambda = 2$ (DEPIZZOL, 2018).

Figura 3 – Em (a): três canais são transmitidos com $\lambda = 3$; em (b): três canais são transmitidos com $\lambda = 2$



Fonte: Depizzol (2018).

O número de comprimentos de onda usados para atender a demanda do tráfego é um fator de custo dominante no dimensionamento de rede, pois minimizar λ minimiza a fragmentação espacial, logo maximiza a eficiência espectral da rede, ou seja, maximiza a faixa de espectro não utilizada. Portanto, um projeto de rede que visa minimizar λ pode acomodar melhor o tráfego e evitar a sobrecarga de tráfego em enlaces específicos,

reduzindo, assim, a possibilidade de *contentions*[1] (DEPIZZOL, 2018).

1.4 Objetivos

1.4.1 Objetivo geral

Foi desenvolvida uma análise em redes ópticas de comunicação, aplicando 9 métodos diferentes de detecção de comunidades, observando como suas topologias influenciaram o número de comprimentos de onda de uma rede.

1.4.2 Objetivos Específicos

- Aplicou-se os métodos de detecção de comunidades descritos na Seção 2.6 em todas as redes do conjunto de $2,2 \times 10^6$ de redes consideradas para análise e calculou-se a modularidade para cada método aplicado.
- Analisou-se a relação entre os valores de modularidade e número de comprimentos de onda por meio de gráficos de dispersão.
- Analisou-se quais os grafos que minimizam λ e comparou-se a modularidade desses grafos com o número de arestas, via diagramas de dispersão.

Para poder alcançar esses objetivos, foram aplicados um conjunto de métodos, técnicas e ferramentas, melhor explicitados no Capítulo 3.

[1] Quando dois ou mais sinais usam o mesmo comprimento de onda.

2 MÉTODOS DE DETECÇÃO DE COMUNIDADES

2.1 Introdução

Este capítulo tem por finalidade estabelecer os conceitos teóricos necessários usados neste trabalho. Portanto, o capítulo inicia-se com a definição de comunidades e como elas são caracterizadas. Em seguida, é apresentada uma exposição de suas estruturas e como são classificadas. Finalmente são explicadas as técnicas de detecção de comunidades (ou *clusters*) que serão utilizadas ao longo do projeto, iniciando com uma explicação geral sobre o assunto, suas definições matemáticas e finalizando com algumas aplicações.

2.2 Grafos

Antes de iniciar o estudo sobre comunidades é preciso definir o que são grafos. A palavra “grafo” é um neologismo derivado da palavra *graph*, em inglês. Ela foi usada pela primeira vez, no sentido de interesse deste trabalho, pelo matemático inglês James Joseph Sylvester (1817–1987) (FURTADO, 1973).

Para qualquer conjunto V, será denotado por $V^{(2)}$ o conjunto de todos os pares não-ordenados de elementos de V. Se V tem n elementos, então $V^{(2)}$ tem $\binom{n}{2} := \frac{n(n-1)}{2}$ elementos. Os elementos de $V^{(2)}$ serão identificados com os subconjuntos de V que têm cardinalidade 2. Assim, cada elemento de $V^{(2)}$ terá a forma $\{v,w\}$, sendo v e w dois elementos distintos de V (FURTADO, 1973).

Um grafo é um par (V,E) em que V é um conjunto arbitrário e E é um subconjunto de $V^{(2)}$. Os elementos de V são chamados de vértices e os de E são chamados de arestas (FURTADO, 1973).

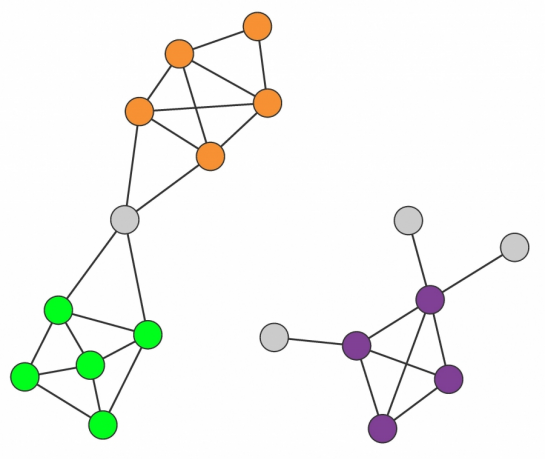
2.3 Comunidades

Neste estudo, considera-se como conceito de comunidade um grupo de vértices de uma rede (grafo) que têm maior conectividade (relação em comum, afinidade, vínculo) entre si a vértices de grupos alheios.

Há uma hipótese importante que, de certa forma, precede a definição de comunidade no presente estudo: a hipótese da Conectividade e da Densidade, a qual diz que "comunidades são subgrafos conectados localmente densos em uma rede" (BARABÁSI, 2016). Em outras palavras, todos os vértices de uma comunidade conseguem ser alcançados através de outros vértices pertencentes a mesma comunidade (conectividade), ou seja, há maior probabilidade de nós que pertencem a uma certa comunidade se conectarem a membros da mesma comunidade, do que a nós de outras comunidades (densidade). Embora esta hipótese tende a definir uma comunidade, ela não a define de forma exclusiva.

Em algumas aplicações, quando usados métodos diversos de detecção de comunidades, o que pode ocorrer é a não correlação de vértices, conforme o esperado. Portanto, a escolha do método e dos parâmetros se torna imprescindível para a detecção de comunidades. Na Seção 2.6 serão detalhados alguns métodos de detecção de comunidades.

Figura 4 – Representação de comunidades em grafos



Fonte: Barabási (2016).

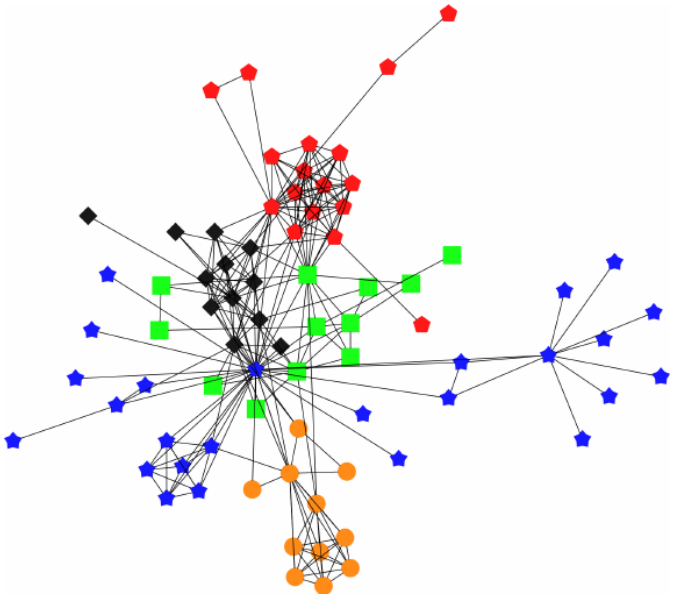
2.4 Classificação das comunidades

2.4.1 Cliques

Uma das primeiras publicações sobre estruturas de comunidade definiu que uma comunidade é um grupo de indivíduos, no qual todos os membros se conhecem. Em teoria de grafos isso significa que uma comunidade pode ser considerada um subgrafo completo, chamado *clique* (BARABÁSI, 2016). Uma clique é um subgrafo que possui todos os nós ligados uns aos outros. No entanto, enxergar comunidades como cliques traz algumas desvantagens:

- Definir uma comunidade como um subgrafo completo pode ser algo bem restritivo, fazendo com que outras comunidades legítimas não sejam consideradas comunidades (BARABÁSI, 2016). A Figura 5 ilustra um exemplo de comunidades legítimas, das quais uma delas é um subgrafo não completo, que está representado pelos pontos laranja.
- A frequente existência de triângulos em uma rede torna cliques largos algo mais raro (BARABÁSI, 2016).

Figura 5 – Estrutura de comunidades em um grafo



Fonte: Griechisch e Pluhár (2011).

2.4.2 Comunidade Forte e Comunidade Fraca

Fugindo da ideia de cliques, considerando um subgrafo conexo C de N_C nós em uma rede. O grau interno k_i^{int} do nó i é o número de ligações que conectam i aos outros nós em C. O grau externo k_C^{ext} é o número de ligações que conectam i ao resto da rede. Se $k_C^{ext} = 0$, cada vizinho de i está contido em C, consequentemente, C é uma boa comunidade para o nó i. Se $k_i^{int} = 0$, então i deve ser atribuído a uma comunidade diferente (BARABÁSI, 2016).

- C é uma **comunidade forte** se cada nó contido em C tem mais ligações com nós da própria comunidade do que com os demais nós do grafo (FLAKE; LAWRENCE; GILES, 2000; RADICCHI et al., 2004). Especificamente, um subgrafo C se torna uma comunidade forte se para cada nó $i \in C$ a inequação (2.1) for satisfeita.

$$k_i^{int} > k_i^{ext} \tag{2.1}$$

- C é uma **comunidade fraca** se a soma dos graus internos de um subgrafo excede a soma dos graus externos (RADICCHI et al., 2004). Especificamente, um subgrafo C forma uma comunidade fraca se for satisfeita a inequação (2.2).

$$\sum_{i \in C} k_i^{int}(C) > \sum_{i \in C} k_i^{ext}(C) \tag{2.2}$$

Com isso, percebe-se que toda clique é uma comunidade forte, e que cada comunidade forte é uma comunidade fraca, mas nem toda comunidade fraca é uma comunidade forte.

A Figura 6(a) explicita comunidades segundo as definições de clique, comunidade forte (Figura 6(b)) e comunidade fraca (Figura 6(c)), em laranja, roxo e verde, respectivamente.

2.5 Modularidade

A *modularidade* (Q) é uma medida proposta por Girvan e Newman. É um parâmetro que indica o quão boa é a divisão das comunidades. Como é dada por Depizzol, Paiva

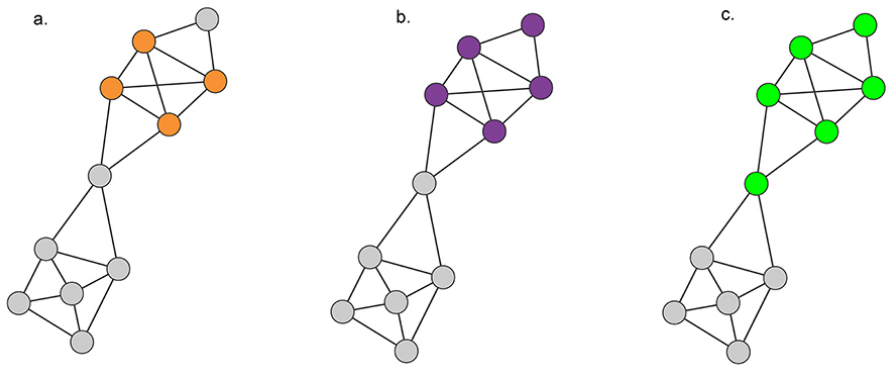
e Segatto (2017), seu cálculo se dá da seguinte forma:

$$Q = \frac{1}{2m} \sum_{i,j=1}^{n} (A_{i,j} - \frac{k_i k_j}{2m}) \delta(i,j) \tag{2.3}$$

Na equação (2.3), tem-se:

- n e m são números de vértices e de arestas, respectivamente.
- i e j representam vértices do grafo.
- A é a matriz de adjacências, que é uma matriz composta por 1's e 0's, onde cada elemento A_{ij} é igual a 1 se o vértice i faz par com o vértice j, ou A_{ij} é igual a 0, caso contrário.
- k_i e k_i são os graus dos vértices i e j, respectivamente.
- $\delta(i,j)$ é uma função, onde: $\delta(i,j)$ = 1 se i e j pertencem à mesma comunidade, e $\delta(i,j)$ = 0 caso contrário.

Figura 6 – Exemplos de cliques e de comunidades forte e fraca



Fonte: Barabási (2016).

A modularidade Q assume um valor entre $-1/2$ e 1, onde quanto mais próximo de 1, mais fortes são as estruturas de comunidades encontradas (BRANDES et al., 2007).

2.6 Métodos de detecção de comunidades

No artigo de Depizzol, Paiva e Segatto (2017) foram utilizadas várias redes com topologias diferentes, nas quais foram aplicados nove métodos de detecção de comunidades

que serão apresentados na presente seção. A escolha desses métodos é dada pelo critério de acessibilidade, ou seja, alguns dos mais conhecidos e já implementados métodos foram escolhidos a partir do fato de eles serem os mais usados em vários estudos.

No APÊNDICE A, estão reunidas as ilustrações das topologias das redes usadas no artigo de Depizzol, Paiva e Segatto (2017).

2.6.1 *Edge betweenness*

O método *Edge betweenness* se baseia no cálculo do menor caminho (menor distância) entre todos os pares de vértices de uma rede. Pode-se dizer também que o *betweenness* de uma aresta é calculado através do número de caminhos mínimos entre pares de vértices que utilizam aquela aresta. Mais formalmente, o cálculo do *betweenness* é feito tomando cada par de nós do grafo e contando quantas vezes um nó pode interromper os caminhos mais curtos (distância geodésica) entre o par de nós tomado. Para a detecção de comunidades, à medida que são retiradas as arestas de maior número de caminhos entre os vértices, as comunidades vão sendo reveladas (NEWMAN; GIRVAN, 2004).

Seja x_{ij} o número de caminhos mais curtos que passam pela aresta (i, j). As arestas que conectam comunidades distintas tendem a ter grande x_{ij}, enquanto arestas dentro de uma comunidade tendem a ter x_{ij} menor (BARABÁSI, 2016).

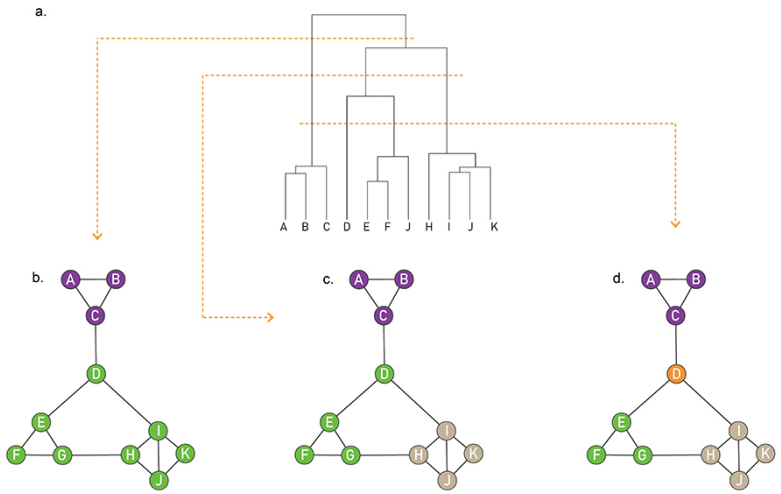
Os passos do algoritmo para a detecção de comunidades via *Edge Betweenness* são dados a seguir:

(i) Calcular o maior *betweenness* para todas as arestas da rede.
(ii) Retirar a aresta de maior *betweenness*.
(iii) Recalcular o *betweenness* para as demais arestas que sobraram.
(iv) Voltar ao passo 2 até não existir mais aresta.

Como resultado do algoritmo, tem-se a formação de uma árvore hierárquica (ou dendrograma).

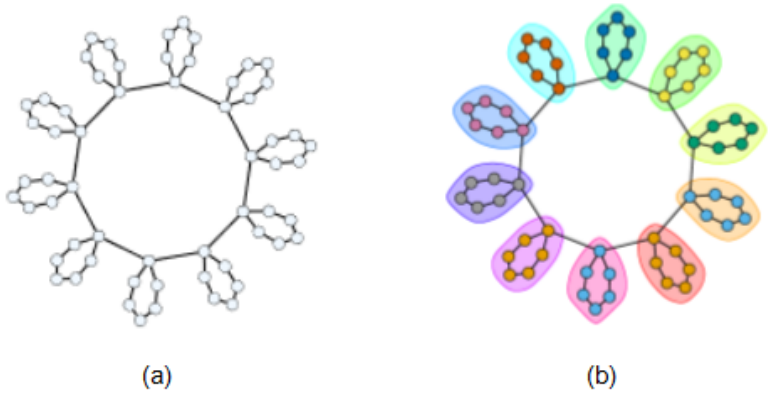
A Figura 7 mostra uma árvore hierárquica gerada por este método, onde é explicitada a formação de estruturas de comunidades ao longo dos passos do algoritmo de GN. Um exemplo de aplicação do método está ilustrado na Figura 8.

Figura 7 – Formação de comunidades utilizando o algoritmo de GN



Fonte: Barabási (2016).

Figura 8 – Aplicação do método *Edge betweennes* em uma rede com 10 anéis de 6 nós cada. Em (a): rede com 10 anéis de 6 nós cada; em (b): resultado do método *Edge betweenness* aplicado à rede em (a)



Fonte: Depizzol, Paiva e Segatto (2017).

2.6.2 *Multilevel Modularity* – Estratégia Multinível

Este método consiste em utilizar um algoritmo de otimização multinível de modularidade. Trata-se de um método aglomerativo, isto é, a cada iteração, os nós são agrupados em comunidades. O algoritmo deste método funciona da seguinte maneira: de início, cada vértice é considerado uma comunidade, ou seja, numa rede de n vértices, há n comunidades. As comunidades são unidas em pares repetidamente, escolhendo a

cada passo a junção que resulta no maior ganho ou na menor perda de Q (AIRES; NAKAMURA, 2017).

A Figura 9 ilustra uma aplicação deste método em uma das redes estudadas no trabalho de Depizzol, Paiva e Segatto (2017).

Figura 9 – Aplicação do método *Multilevel modularity* em uma rede com 4 anéis de 15 nós cada. Em (a): rede com 4 anéis de 15 nós cada; em (b): resultado do método *Multilevel modularity* aplicado à rede em (a)

Fonte: Depizzol, Paiva e Segatto (2017).

2.6.3 *Optimal modularity* – o método de Louvain

O método de Louvain é uma das abordagens mais populares em maximização da modularidade devido a sua simplicidade, eficiência e fácil implementação. O algoritmo de Louvain procura otimizar localmente as comunidades até que a modularidade global não consiga mais ser melhorada.

O algoritmo depende de um procedimento um tanto ambicioso: a partir de qualquer partição dos vértices (normalmente a partição em *singletons*[1]), o algoritmo tenta aumentar o valor da modularidade, movendo vértices de sua comunidade para qualquer outro vértice vizinho. Em outras palavras, o algoritmo computa o ganho da modularidade obtido adicionando o vértice i à comunidade C (DUGUÉ; PEREZ, 2015).

[1] *Singleton* é um conjunto que contém exatamente um elemento.

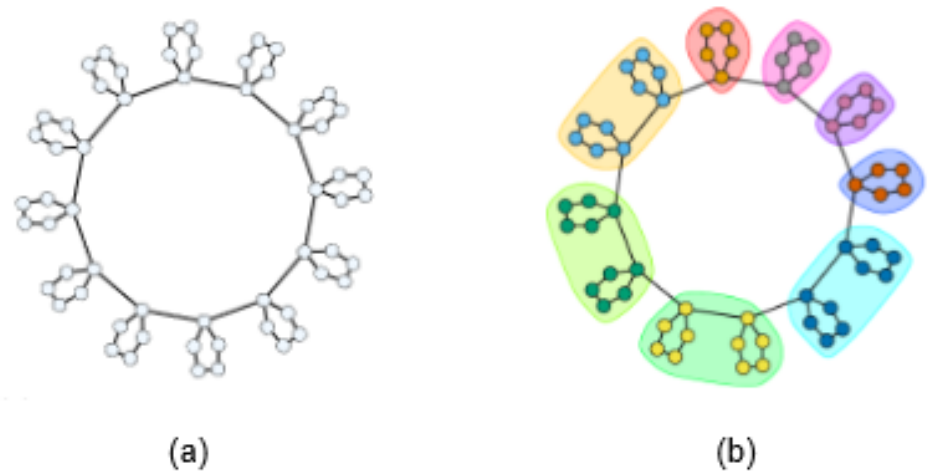
O algoritmo do método de *Louvain* é dado pela equação (2.4).

$$\begin{aligned}\Delta_Q &= \left[\frac{\sum in + d_i^C}{2m} - \left(\frac{\sum in + d_i}{2m}\right)^2\right] - \left[\frac{\sum in}{2m} - \left(\frac{\sum tot}{2m}\right)^2 - \left(\frac{d_i}{2m}\right)^2\right] = \\ &= \frac{d_i^C}{2m} - \frac{\sum tot \cdot d_i}{2m^2}\end{aligned} \tag{2.4}$$

Da equação (2.4), d_i^C denota o grau do nó i inserido na comunidade C; $\sum in$ o número de arestas contidas na comunidade C, e $\sum tot$ o número total de arestas ligadas à comunidade C. O algoritmo continua até não haver mais um movimento que melhore o valor da modularidade (DUGUÉ; PEREZ, 2015).

A Figura 10 explicita o resultado deste método aplicado a uma determinada rede.

Figura 10 – Aplicação do método *Optimal modularity* em uma rede com 12 anéis de 5 nós cada, sendo (a) rede com 12 anéis de 5 nós cada, e (b) esultado do método *Optimal modularity* aplicado à rede em (a)



Fonte: Depizzol, Paiva e Segatto (2017).

2.6.4 *Walktrap*

Este método consiste em procurar comunidades através de *random walks* (caminhos aleatórios). A ideia usada aqui é que *random walks* pequenos tendem a mostrar vértices que pertencem à mesma comunidade.

A partir dessa ideia, o algoritmo propõe uma medida que calcula uma distância entre

vértices, e/ou grupo de vértices, baseado no caminho percorrido pelo *random walker* entre os dois componentes. Essa medida é usada como cálculo de similaridade num algoritmo de detecção de comunidades hierárquico (PONS; LATAPY, 2005).

O *random walker* se move de um vértice a outro a cada passo do algoritmo. O vértice para o qual será movido é escolhido de forma aleatória dentre os vizinhos do vértice em que ele se encontra. A sequência, ou caminho, de vértices percorridos pelo *walker* na rede é uma cadeia de Markov, isto é, um processo estocástico com a propriedade de que a distribuição de probabilidade do próximo estado depende apenas do estado atual e não da sequência de eventos precedentes (KARLIN, 2014), sendo que os estados da cadeia são os vértices visitados. A cada passo, a probabilidade de transação de um vértice i para um vértice j é $P_{ij} = \frac{A_{ij}}{d(i)}$ (PONS; LATAPY, 2005), onde:

- A *matriz de adjacências* A é uma das formas de se representar um grafo. Dado um grafo de n vértices, pode-se representá-lo por uma matriz $A_{n \times n}$ ou simplesmente A. As entradas dessa matriz variam de acordo com as propriedades do grafo.
- A_{ij} corresponde aos elementos da matriz de adjacências A, onde: $A_{ij} = 1$ se os vértices i e j estão conectados, e $A_{ij} = 0$, caso contrário.
- $d(i) = \sum_j A_{ij}$ é o grau do vértice i que corresponde ao número de seus vizinhos (incluindo ele mesmo).
- P é a matriz de transição e descreve o processo do *random walker*.

Para calcular a probabilidade do *walker* ir do vértice i para o vértice j em t passos, basta calcular P^t e observar o elemento $(P^t)_{ij}$.

A medida da distância entre 2 vértices, pelo algoritmo, é calculada a partir da matriz de transição P e é descrita pela equação (2.5) (PONS; LATAPY, 2005).

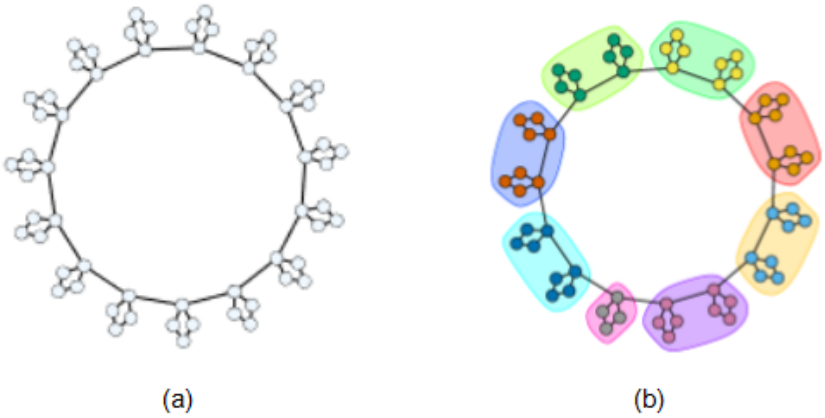
$$r_{C1C2} = \sqrt{\sum_{k=1}^{n} \frac{(P^t_{C1k} - P^t_{C2k})^2}{d(k)}} \tag{2.5}$$

Onde C_1 e C_2 são os componentes (vértices ou subgrafos), os quais está sendo medida a distância entre eles, e P^t_{ij} é uma simplificação de $(P^t)_{ij}$.

Essa distância é calculada entre todas as comunidades que são adjacentes, ou seja, que tenham pelo menos uma aresta ligando-as, e as duas comunidades que obtiverem a maior similaridade serão unidas numa única comunidade.

A Figura 11 explicita o resultado deste método aplicado a uma determinada rede.

Figura 11 – Aplicação do método *Walktrap* em uma rede com 15 anéis de 4 nós cada, sendo (a) rede com 15 anéis de 4 nós cada, e (b) resultado do método *Walktrap* aplicado à rede em (a)



Fonte: Depizzol, Paiva e Segatto (2017).

2.6.5 *Spin-glass*

Este método procura encontrar estruturas de comunidade usando o modelo *spin-glass*. Um *spin-glass* ou vidro de *spin* é um sistema magnético no qual, no conjunto, os acoplamentos entre os momentos magnéticos dos distintos átomos são aleatórios nas interações de troca de sinal variável, tanto ferromagnético como anti-ferromagnético e apresenta um forte grau de frustração aumentado por desordem estocástica. Esta desordem magnética é semelhante a ordenação posicional de um vidro químico convencional (ALMEIDA; THOULESS, 1978).

A energia do sistema de spin é equivalente à configuração de spins, e minimizar essa energia resulta em um estado fundamental estável. Reichardt mostrou que o estado fundamental do modelo de Ising[2] corresponde a dividir um grafo em duas comunidades naturais (ALMEIDA; THOULESS, 1978).

[2]O modelo Ising é uma coleção de dois estados de rotação em uma configuração (ALMEIDA; THOULESS, 1978).

O algoritmo para a detecção de comunidades, baseado no modelo *sping-glass*, segue então o seguinte procedimento:

(i) Classificar os nós de acordo com um certo grau de relevância.
(ii) Seleção e ajuste de nós influentes fazem os nós selecionados como um potencial líder da comunidade local.
(iii) Usar o líder como o ponto de partida para expandir a comunidade local baseada no modelo *spin-glass*.
(iv) Voltar ao passo 2 até que cada nó seja atribuído a pelo menos uma comunidade local.

2.6.6 *Fastgreedy modularity*

Este método detecta estruturas de comunidades via otimização direta de valores de índice de modularidade.

Inicialmente, cada vértice compõe uma única comunidade e as comunidades são mescladas iterativamente, de forma que cada fusão é localmente ótima (ou seja, produz o maior aumento no valor atual da modularidade). O algoritmo para quando não é mais possível aumentar a modularidade, por isso ele fornece um agrupamento, mais conhecido como dendrograma. O método é rápido e este é o método que geralmente é tentado como uma primeira aproximação porque não possui parâmetros para ajustar. No entanto, sabe-se que sofre de um limite de resolução, isto é, comunidades abaixo de um determinado limite de tamanho (dependendo do número de nós e bordas) serão sempre fundidas com comunidades vizinhas (NEWMAN; GIRVAN, 2004).

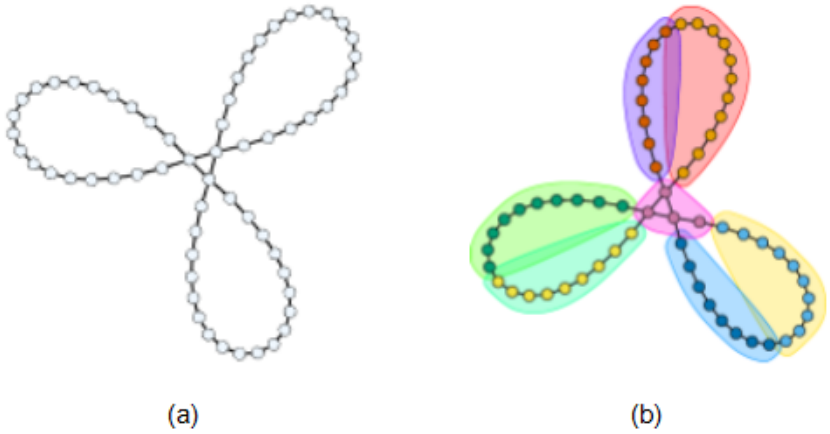
A Figura 12 explicita o resultado deste método aplicado a uma determinada rede.

2.6.7 *Infomap* – método baseado em fluxo

Métodos baseados em fluxos operam sobre a dinâmica da rede e não sobre sua estrutura topológica em si. A lógica é que a função principal das redes é capturar o fluxo entre os componentes dos sistemas reais que eles representam. Consequentemente, as comunidades consistem em nós, entre os quais o fluxo persiste por um longo tempo,

uma vez inserido (BOHLIN et al., 2014). Nesta seção, é apresentada uma descrição de um dos métodos baseados em fluxo, conhecido como *equação do mapa*.

Figura 12 – Aplicação do método *Fastgreedy modularity* em uma rede com 3 anéis de 20 nós cada, sendo (a) rede com 3 anéis de 20 nós cada, e (b) resultado do método *Fastgreedy modularity* aplicado à rede em (a)



Fonte: Depizzol, Paiva e Segatto (2017).

Para uma dada partição de rede, a equação do mapa especifica qual o limite teórico do quão concisamente pode-se descrever a trajetória de um caminho aleatório em uma rede. Minimizar a equação do mapa sobre todas as partições possíveis da rede revela aspectos importantes da estrutura da rede em relação a sua dinâmica (ROSVALL; AXELSSON; BERGSTROM, 2009). Para encontrar uma partição ideal da rede, é suficiente calcular o limite para diferentes partições da rede e escolher o que descreve menor caminho.

Para uma partição de módulo M de n nós $\alpha = 1,2,\ldots,n$ em m módulos $i = 1,2,\ldots,m$, define-se $L(M)$ como o limite inferior no comprimento do código (ROSVALL; AXELSSON; BERGSTROM, 2009). Para calcular L para uma partição arbitrária, deve-se invocar o teorema de codificação de origem de Shannon (BONFIM, 2014), que implica que quando usados n palavras de código para descrever os n estados de uma variável X que ocorrem com uma frequência p_i, o comprimento médio de uma palavra código não pode ser menor que a entropia da própria variável aleatória X: $H(X) = \sum_1^n p_i \log(p_i)$ (mede-se o comprimento do código em bits e o logaritmo tomado é na base 2) (ROSVALL; AXELSSON; BERGSTROM, 2009).

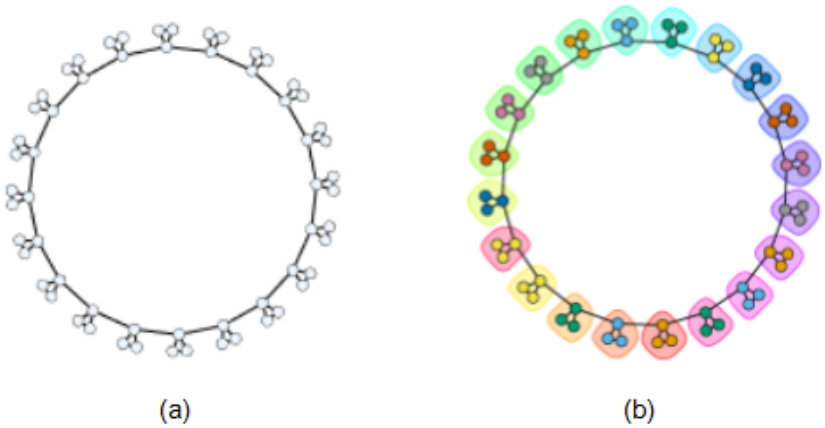
A equação do mapa é dada pela equação (2.6).

$$L(M) = q_{\curvearrowright} H(\mathscr{Q}) + \sum_{i=1}^{m} p^{i}_{\circlearrowright} H(\mathscr{P}^{i}) \tag{2.6}$$

Da equação (2.6), tem-se: $H(\mathscr{Q})$ é o comprimento médio ponderado por frequência da palavra-código e $H(\mathscr{P}^{i})$ é o comprimento médio ponderado por frequência de palavras de código no livro de códigos de módulo i; $q_{\curvearrowright}$ é a probabilidade do *random walker* de mudar de módulo em um determinado passo; e $p^{i}_{\circlearrowright}$ é a fração de tempo que o caminho aleatório gasta no módulo i, mais a probabilidade de que ele saia do módulo e a mensagem de saída seja usada.

A Figura 13 explicita o resultado deste método aplicado a uma determinada rede.

Figura 13 – Aplicação do método *Infomap* em uma rede com 20 anéis de 3 nós cada, sendo (a) rede com 20 anéis de 3 nós cada, e (b) resultado do método *Infomap* aplicado à rede em (a)



Fonte: Depizzol, Paiva e Segatto (2017).

2.6.8 *Label propagation*

Neste método, na condição inicial, os nós carregam um rótulo que denota a comunidade a que pertencem. A associação a uma comunidade muda com base nos rótulos que os nós vizinhos possuem. Essa alteração está sujeita ao número máximo de rótulos dentro de um grau dos nós. Cada nó é inicializado com um rótulo exclusivo e, em seguida, os rótulos são difundidos pela rede. Consequentemente, grupos densamente conectados alcançam um rótulo comum rapidamente. Quando muitos grupos densos são criados em toda a rede, eles continuam a se expandir para fora até que seja impossível fazê-lo (DEPIZZOL; PAIVA; SEGATTO, 2017; RAGHAVAN; ALBERT; KUMARA, 2007).

Este processo consiste em 5 etapas:

(i) Atribuir rótulos a todos os nós da rede. Para um dado nó x, $C_x(0) = x$, supondo que x possui vizinhos $x_1, x_2, \ldots, x_k$.

(ii) Tomar $t = 1$, onde t é o número de iterações.

(iii) Organizar os nós na rede numa ordem aleatória, configurando-a em um conjunto X.

(iv) Para cada nó $x \in X$ escolhido nessa ordem específica, fazer:

$$C_x(t) = f(C_{x_{i1}}(t), \ldots, C_{x_{im}}(t), C_{x_{i(m+1)}}(t-1), \ldots, C_{x_{ik}}(t-1)) \tag{2.7}$$

onde $C_{x_{i1}}(t), \ldots, C_{x_{im}}(t), C_{x_{i(m+1)}}(t-1), \ldots, C_{x_{ik}}(t-1)$ e $x_{i1}, \ldots, x_{im}$ são vizinhos de x que já foram atualizados na iteração, enquanto que $x_{i(m+1)}, \ldots, x_{ik}$ são vizinhos que ainda não foram atualizados na iteração atual. A função f aqui retorna o rótulo que ocorre com maior frequência entre os vizinhos e as ligações são divididas uniforme e aleatoriamente.

(v) Se cada nó tiver um rótulo o qual o número máximo de vizinhos tenha, pare o algoritmo. Senão, faça $t = t + 1$ e volte à etapa 3.

A Figura 14 explicita o resultado deste método aplicado a uma determinada rede.

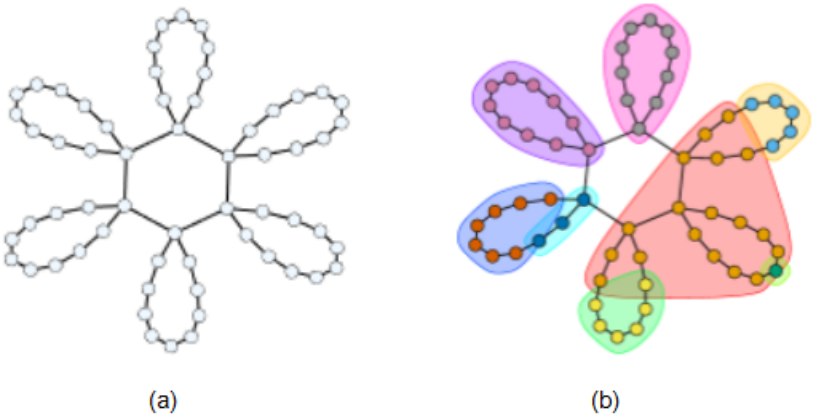
2.6.9 *Eingenvector modularity*

Este método consiste em encontrar comunidades calculando o principal autovetor não-negativo da matriz de modularidade do grafo (DEPIZZOL; PAIVA; SEGATTO, 2017), B, a qual $B = A - P$, onde A é a matriz de adjacências da rede e P contém a probabilidade de certas arestas estarem presentes de acordo com o modelo de configuração. Em outras palavras,um elemento P_{ij} de P é a probabilidade de haver uma aresta entre os vértices i e j em uma rede aleatória na qual os graus de todos os vértices são os mesmos que os do grafo de entrada (NEWMAN, 2006).

Dessa forma, o método calcula o autovetor da matriz de modularidade B para o maior autovalor positivo e depois separa os vértices em duas comunidades com base no sinal do elemento correspondente no autovetor. Se todos os elementos no autovetor forem de mesmo sinal, isso significa que a rede não possui estrutura de comunidade

subjacente.

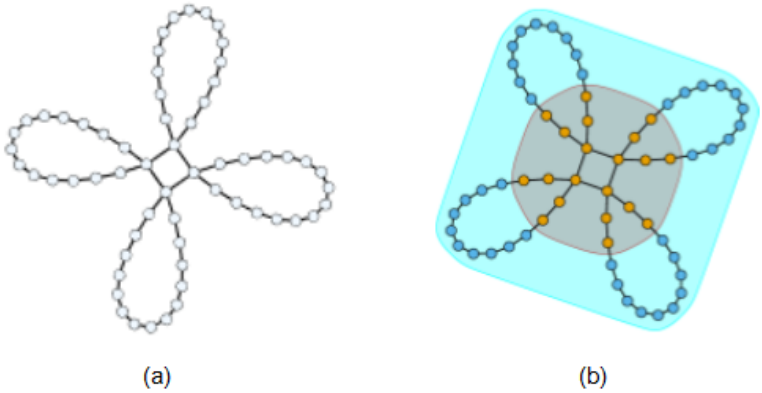
Figura 14 – Aplicação do método *Label propagation* em uma rede com 6 anéis de 10 nós cada, sendo (a) rede com 6 anéis de 10 nós cada, e (b) resultado do método *Label propagation* aplicado à rede em (a)



Fonte: Depizzol, Paiva e Segatto (2017).

A Figura 15 explicita o resultado deste método aplicado a uma determinada rede.

Figura 15 – Aplicação do método *Eigenvector modularity* em uma rede com 4 anéis de 15 nós cada, sendo (a) rede com 4 anéis de 15 nós cada, e (b) resultado do método *Eigenvector modularity* aplicado à rede em (a)



Fonte: Depizzol, Paiva e Segatto (2017).

3 METODOLOGIA

3.1 Introdução

Esta pequisa investiga como os aspectos topológicos de uma rede podem influenciar o número de comprimentos de onda, que é considerada uma métrica de desempenho de rede. Para isso, serão aplicados os métodos de detecção de comunidade abordados no capítulo anterior. Cada topologia de rede pode apresentar diferentes estruturas de comunidade, de acordo com o método de detecção utilizado. Serão analisadas estas estruturas para cada topologia de rede considerada.

Todos os métodos dados na Seção 2.6 serão aplicados para uma quantidade de aproximadamente $2,2$ milhões de redes. Este conjunto contém grafos com topologias 2-conexas com número de vértices de 10 a 20, e que satisfazem $0,1 < \alpha < 0,4$, onde α é a densidade do grafo (DEPIZZOL et al., 2018). A densidade de um grafo é dada em função da relação entre sua ordem e seu tamanho, ou seja, representa a proporção entre a quantidade de arestas e vértices, podendo ser calculada pela equação (3.1).

$$\alpha = \frac{2m}{n(n-1)} \tag{3.1}$$

Da equação (3.1), α é a densidade, m é o número de arestas e n é o número de vértices.

Pode-se dizer que este é um trabalho sobre Ciência de Dados, visto que há uma análise de um grande volume de dados, como já citado, além do fato de serem usadas técnicas e ferramentas para tratamento, validação e armazenamento de dados, as quais serão detalhadas ainda neste capítulo.

As redes que serão estudadas foram geradas de forma aleatória, porém com alguns critérios, fazendo com que suas topologias “imitassem” as do mundo real, especificamente, redes ópticas *backbone*, que é o tipo de rede que transporta dados coletados de redes menores que se conectam com ela.

Uma rede *backbone* se comporta como uma espinha dorsal para o transporte de sinais

advindos das redes menores, ou seja, é por onde redes locais ou regionais se conectam para estabelecerem uma interconexão de longa distância.

3.2 Cálculo das métricas

Os grafos analisados foram representados por *edge lists*, que é uma estrutura de dados usada para representar um grafo como uma lista de suas arestas. Uma aresta é definida pelo seu vértice inicial e final, portanto, cada aresta pode ser representada por dois números.

O cálculo de todas as métricas topológicas no presente estudo (exceto o cálculo do número mínimo de comprimentos de onda) foi feito pelo autor deste trabalho, utilizando uma máquina de processador Intel Xeon Processor E5-2430 v2 com memória RAM de 96 GB DDR3. Utilizou-se *scripts* de linguagem R (IHAKA; GENTLEMAN, 1996) e Julia (BEZANSON et al., 2017) para ler todos os *edge lists*, efetuar os cálculos de todas as métricas e tratar os dados gerados. Todos os dados das propriedades calculadas foram salvos em arquivos no formato JSON (*Javascript Object Notation*), que é uma formatação leve de troca de dados (para seres humanos, é fácil de ler e entender; para máquinas, é fácil de interpretar e gerar) (SEVERANCE, 2012).

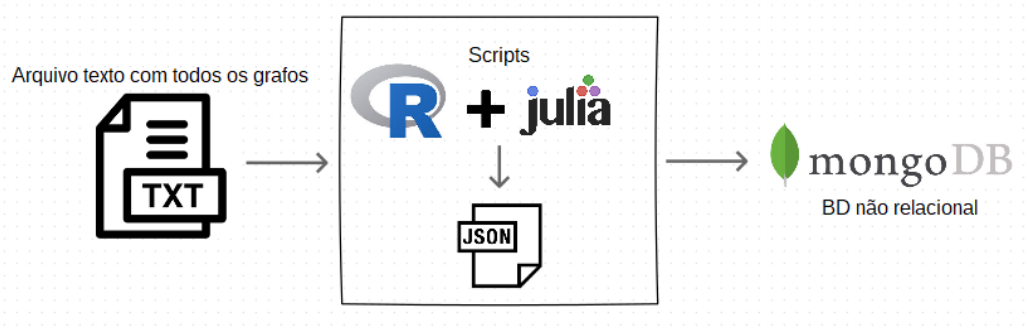
3.3 Armazenamento e extração dos dados

Por último, para armazenar os arquivos JSON, foi criado um banco de dados MongoDB (BANKER, 2011), que é um banco de dados não relacional orientado a documentos.

Todas as informações necessárias para a realização deste trabalho foram extraídas diretamente do banco, através de consultas em *MongoDB's Query Language*. O conjunto de dados foi dividido em vários subconjuntos de grafos que possuíam o mesmo número de vértices que satisfazem $10 \leq n \leq 20$. Cada subconjunto foi exportado para um arquivo CSV (*Comma Separated Value*). Com os dados desses arquivos foi possível gerar gráficos de dispersão para a realização das análises da correlação dos invariantes topológicos.

A Figura 16 explicita o diagrama do processo descrito neste capítulo: leitura, cálculo das métricas e armazenamento dos dados.

Figura 16 – Diagrama do processo de leitura, cálculo e armazenamento dos dados



Fonte: Produção do próprio autor.

3.4 Geração dos gráficos de dispersão

Para cada conjunto de grafos de mesmo número de vértices, utilizando linguagem R, foi gerado um gráfico de dispersão relacionando algumas métricas de interesse: modularidade calculada para cada método de detecção de comunidade da Seção 2.6 *versus* número de comprimentos de onda e, tomando redes com o menor λ, para cada número de vértices, modularidade *versus* número de arestas.

O objetivo da geração desses gráficos foi analisar e identificar uma relação entre essas métricas: observar se o gráfico apresentou algum tipo de tendência, seja ela crescente ou decrescente e, com isso, realizar as análises e obter resultados que serão discutidos no próximo capítulo.

4 RESULTADOS

4.1 Relação entre Modularidade e Número de Comprimentos de Onda (λ)

Antes da análise dos gráficos de dispersão, alguns questionamentos foram feitos: há alguma relação de dependência entre a modularidade e o número de comprimentos de onda? Se sim, quanto maior é a modularidade, o número de comprimentos de onda é maior ou menor? A partir desses questionamentos, pôde-se chegar à grande pergunta deste estudo: um valor alto de modularidade (grafo que possui comunidades teoricamente bem definidas) implica um número de comprimentos de onda alto? Uma resposta positiva a essa pergunta leva à seguinte conclusão: uma rede com fortes divisões de comunidades possui um custo elevado, pois teria um número de comprimentos de onda alto, o que proporcionaria, para um projeto de uma rede de custo alto, projetá-la com uma modularidade alta.

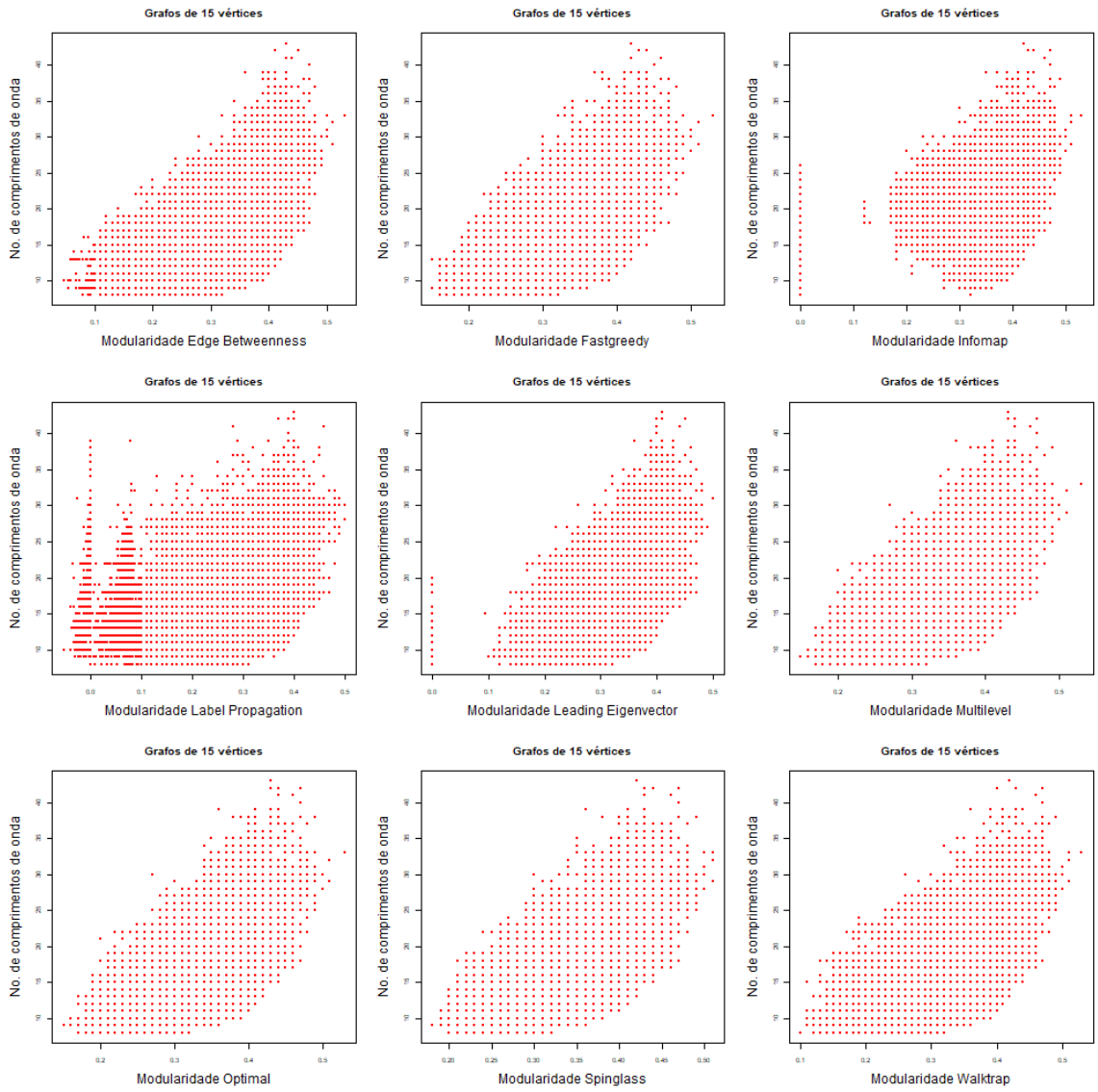
A partir da análise dos gráficos de modularidade *versus* número de comprimentos de onda, observou-se que, para alguns dos nove métodos de detecção de comunidades aplicados, o número de comprimentos de onda cresce à medida que a modularidade também cresce. Os resultados deste capítulo foram baseados no conjunto de grafos de 15 vértices do conjunto, frisando que as análises foram feitas para o conjunto de aproximadamente 2,2 milhões de redes. A Figura 17 explicita os gráficos de dispersão gerados para n = 15, relacionando modularidade e número de comprimentos de onda para cada método de detecção de comunidades considerado. Os demais resultados são apresentados nas Figuras 24 a 34, no APÊNDICE B.

Pela Figura 17, os gráficos dos métodos *Edge Betweenness*, *Fastgreedy*, *Label Propagation*, *Leading Eigenvector*, *Multilevel Modularity*, *Optimal Modularity*, *Spinglass* e *Walktrap* apresentaram uma certa tendência. Entretanto, para o método *Leading Eigenvector*, há uma indefinição de número de comprimentos de onda quando a modularidade é 0, portanto, não se pode dizer que, quando aplicado esse método, existe sempre uma relação entre modularidade e λ.

É válido ressaltar que as tendências mostradas na Figura 17 se repetiram para todos

os conjuntos de grafos divididos por número de nós, como mostram as Figuras 24 a 34, constantes no APÊNDICE B. O Quadro 1 informa o tipo de tendência observada (crescente, decrescente ou nenhuma) de cada gráfico modularidade *versus* λ gerado.

Figura 17 – Todos os gráficos de dispersão de modularidade *versus* λ, para o conjunto de 15 vértices



Fonte: Produção do próprio autor.

Conforme apresentado no Quadro 1, nenhum método de detecção de comunidades apresentou uma tendência decrescente, ou seja, apenas analisando o gráfico de dispersão entre as duas métricas (modularidade e λ), não se pode inferir que quanto mais forte é a divisão de uma rede em comunidades, menor é seu número de comprimentos de onda. A partir dessa primeira observação, foi feito um estudo mais minucioso, olhando para o conjunto de grafos que apresentou o menor λ.

Quadro 1 – Tendências apresentadas nos gráficos de dispersão entre modularidade e λ para o conjunto de grafos de 15 vértices

Métodos de detecção de comunidades	**Tendência**
Edge Betweenness	Crescente
Fastgreedy	Crescente
Infomap	Nenhuma
Label Propagation	Nenhuma
Leading Eigenvector	Nenhuma
Multilevel	Crescente
Optimal	Crescente
Spinglass	Crescente
Walktrap	Crescente

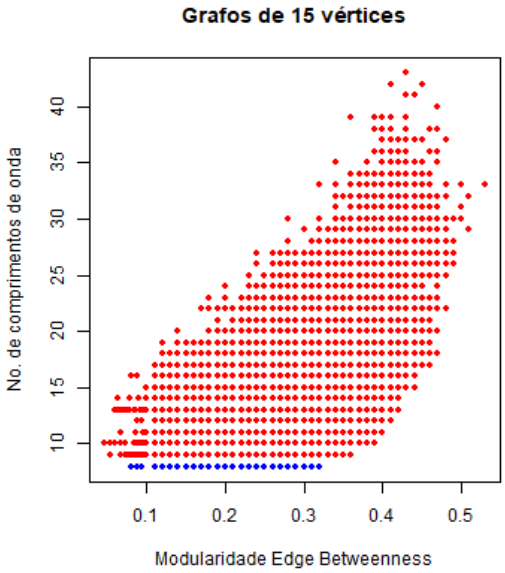
Fonte: Produção do próprio autor.

4.2 Análise do conjunto de grafos que minimizam o número de comprimentos de onda

Após analisar os gráficos que apresentaram uma tendência crescente, conforme discutido na Seção 4.1, foram destacados os grafos com o menor número de comprimentos de onda. A partir disso, observou-se que esse conjunto de redes abrange um intervalo de valores de modularidade consideravelmente amplo, onde foi possível inferir que, naquele intervalo, a modularidade não influencia o número de comprimentos de onda. Tomando como exemplo o método *Edge Betweenness*, devido ao fato de seu valores de modularidade terem correlação com os valores de λ, a Figura 18 mostra, na cor azul, os grafos de menor número de comprimentos de onda. Os pontos ocupam uma faixa de modularidade que varia entre, aproximadamente, 0,08 a pouco mais de 0,32, para o conjunto de grafos de 15 vértices; para os conjuntos com n = 10, 11, 12, 13, 14, 16, 17 e 18, essa faixa teve amplitude similar, já para os conjuntos com n = 19 e n = 20, essa amplitude de valores de modularidade foi reduzida, como está explicitado na Figura 19.

Pela Figura 19, percebe-se que nos conjuntos de 19 e 20 vértices, o intervalo de valores de modularidade para o menor λ é menor em relação aos outros conjuntos. Isso se repete quando aplicados os outros métodos de detecção de comunidades que apresentaram gráficos de tendência crescente, como pode ser visto nas Figuras 24 a 34 do APÊNDICE B.

Figura 18 – Destaque dos grafos de 15 vértices de menor λ, utilizando a modularidade *Edge Betweenness*



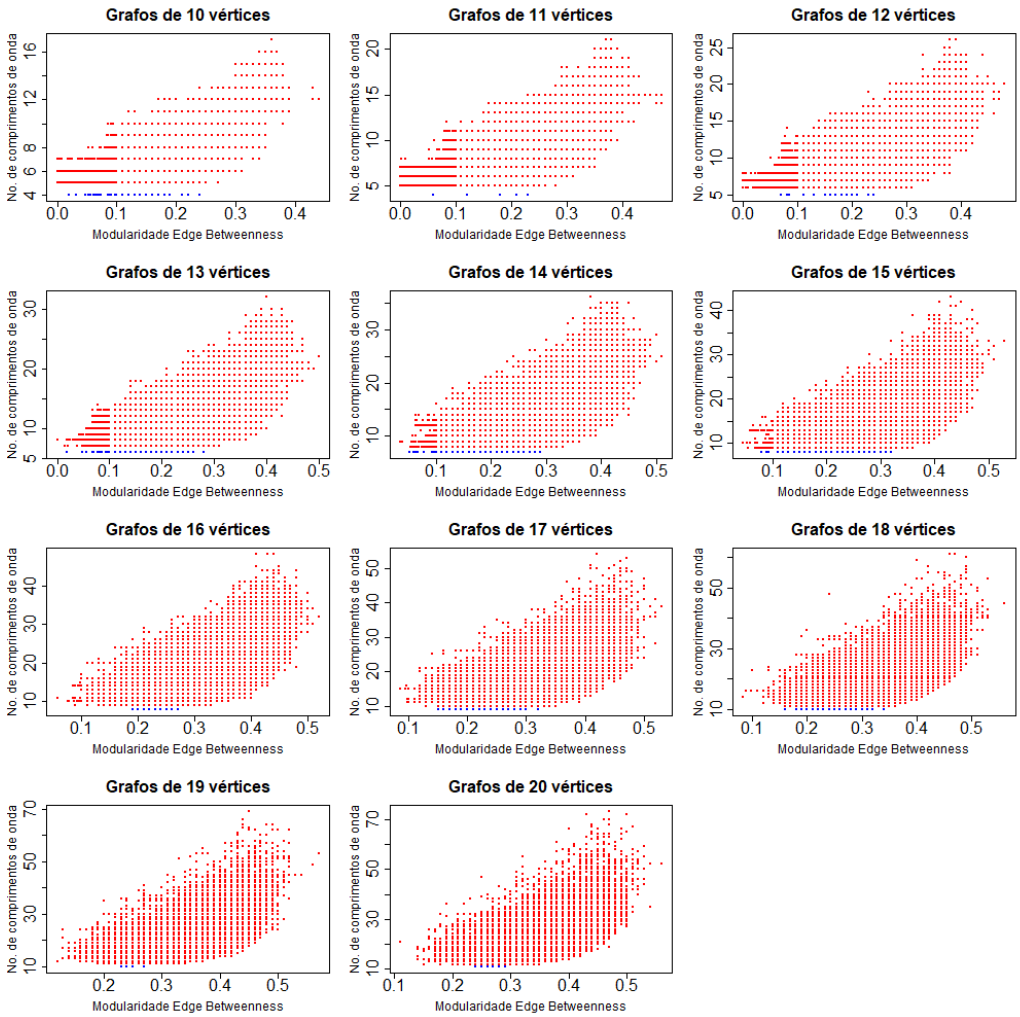
Fonte: Produção do próprio autor.

Mesmo com a faixa de valores reduzida, para os casos de 19 e de 20 vértices, não foi possível determinar um valor exato de modularidade (para todos os casos) que implica um menor número de comprimentos de onda. Entretanto, foi possível obter uma melhor precisão na escolha de valores de modularidade que podem levar aos menores números de comprimentos de onda. Voltando ao gráfico da Figura 18, percebe-se que o eixo de modularidade contempla valores que variam entre 0 e 0,55, e, como já dito, a faixa de valores de modularidade que possui o menor valor de λ está entre 0,08 a 0,32, aproximadamente. Visualmente, os valores de λ que estão na faixa inteira do eixo de modularidade variam entre 4 e quase 50; já os valores de λ correspondentes aos valores de modularidade da faixa menor variam de 4 a pouco mais de 30. Ou seja, é possível ter uma melhor precisão na escolha de grafos com menores números de comprimentos de onda, reduzindo a incerteza na busca pelos grafos com as características desejadas. Essa análise vale para todos os gráficos gerados de todos os métodos de detecção de comunidade que tiveram tendência crescente, os quais podem ser consultados nas Figuras 24 a 34 do APÊNDICE B, citadas na Seção 4.1.

O Quadro 2 mostra quantos grafos com o menor λ tem cada conjunto de grafos

separados por número de vértices. Por exemplo, pode-se notar que as amostras de grafos com 10 e 11 vértices consideradas possuem o mesmo λ mínimo, no caso $\lambda = 4$. No entanto, enquanto a amostra de grafos de 10 vértices apresenta 99 grafos que minimizam λ, na amostra de grafos de 11 vértices há apenas 5 grafos que atingem o λ mínimo. Esse fato pode ter uma aplicação prática no que diz respeito a escalabilidade de uma rede. Quando aumentada em 1 vértice, a rede se mantém com o mesmo número de comprimentos de onda mínimo. Um caso análogo a esse pode ser observado nas amostras de grafos com λ mínimo iguais a 8 e 10.

Figura 19 – Em azul, estão destacados os grafos de menor λ de cada conjunto separado por número de vértices, utilizando a modularidade *Edge Betweenness*



Fonte: Produção do próprio autor.

Já nas Figuras 21 e 22, são mostrados os histogramas dos grafos separados por número de vértices, relacionando a quantidade de grafos por número de comprimentos de onda. Os histogramas servem para mostrar a diferença grande que existe entre a

quantidade de grafos que minimizam λ e a quantidade dos demais grafos. Percebe-se que o conjunto dos grafos correlacionados ao menor λ é relativamente pequeno.

Quadro 2 – Quantidade de grafos com menor λ de cada conjunto de grafos separados por número de vértices, onde N_λ é a quantidade de grafos que minimizam o número de comprimentos de onda

	Número de vértices (n)										
	10	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
λ	4	4	5	6	7	8	8	9	10	10	11
N_λ	99	5	20	498	1086	2201	22	191	362	5	17

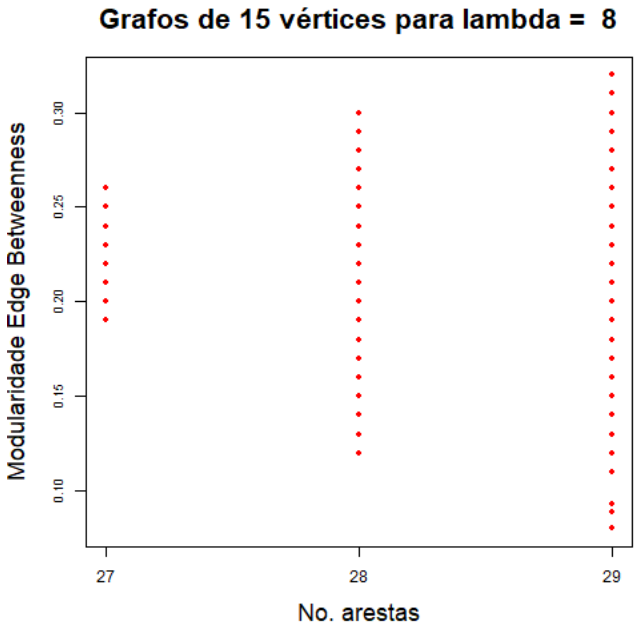
Fonte: Produção do próprio autor.

4.3 Relação entre a Modularidade e o Número de Arestas em grafos que minimizam o número de comprimentos de onda

A partir do estudo da Seção 4.1, tomou-se o conjunto de grafos que minimizam o número de comprimentos de onda para cada n e verificou-se a relação entre a modularidade de cada método de detecção de comunidade e o número de arestas (m). Essa análise de números de arestas teve como objetivo verificar se cada um desses grafos possuem o mesmo número de arestas ou se há número de arestas diferente entre eles, já que a quantidade de arestas também é um invariante de interesse da rede: quanto menos arestas, menor é seu custo. A Figura 20 mostra o gráfico que compara o número de arestas com a modularidade, tomando como exemplo a modularidade calculada pelo método Edge Betweenness. No APÊNDICE C, estão reunidos nas Figuras 35 a 45 os gráficos que relacionam as modularidades calculadas para todos os métodos de detecção de comunidade com o número de arestas, fixando o menor λ, para todos os conjuntos separados por número de vértices.

Observando a Figura 20, vê-se que os grafos com o menor λ têm valores de m = 27, assim como m = 28 e m = 29. Dentre eles, destacam-se os de menor custo os grafos que possuem o menor m, ou seja, m = 27. Os valores de modularidade correspondentes ao menor valor de m estão em um intervalo ainda menor em relação ao intervalo de valores de modularidade dos grafos que minimizam o λ. Isso mostra que a escolha das redes de menor custo (menor m), por meio da métrica modularidade, torna-se ainda mais precisa.

Figura 20 – Gráfico de dispersão entre modularidade *Edge Betweenness* e m, para grafos de 15 vértices

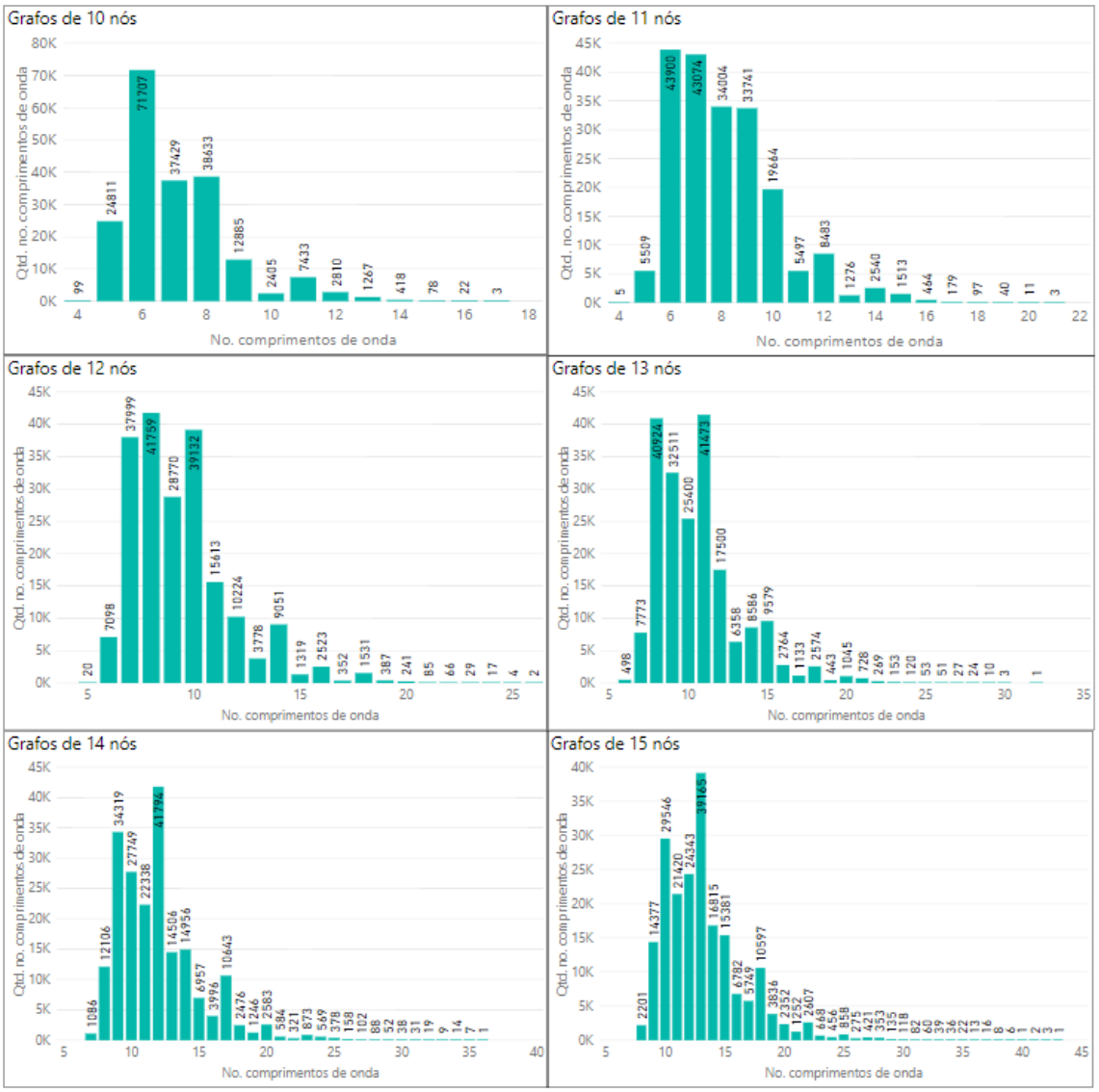


Fonte: Produção do próprio autor.

Entretanto, percebeu-se que os números de arestas dos grafos com o menor λ são os maiores em relação aos demais grafos do conjunto. Isso mostra que uma rede precisa de um número relativamente alto de arestas para conseguir minimizar o λ. O Quadro 3 mostra os números de arestas mínimos de cada conjunto de grafos que minimizam λ, como também reúne todos os valores mínimos e máximos de m (do conjunto inteiro) dos grafos separados por números de nós.

Por outro lado, em nenhum caso o número de arestas necessárias para minimizar o λ ficou condicionado ao uso do maior número possível de arestas do conjunto. Portanto, pode-se concluir que, apesar do número de arestas necessários para um grafo minimizar o λ ser grande, não é necessário sempre usar m máximo do conjunto.

Figura 21 – Histogramas de grafos com n entre 10 e 15, relacionando a quantidade de grafos por λ



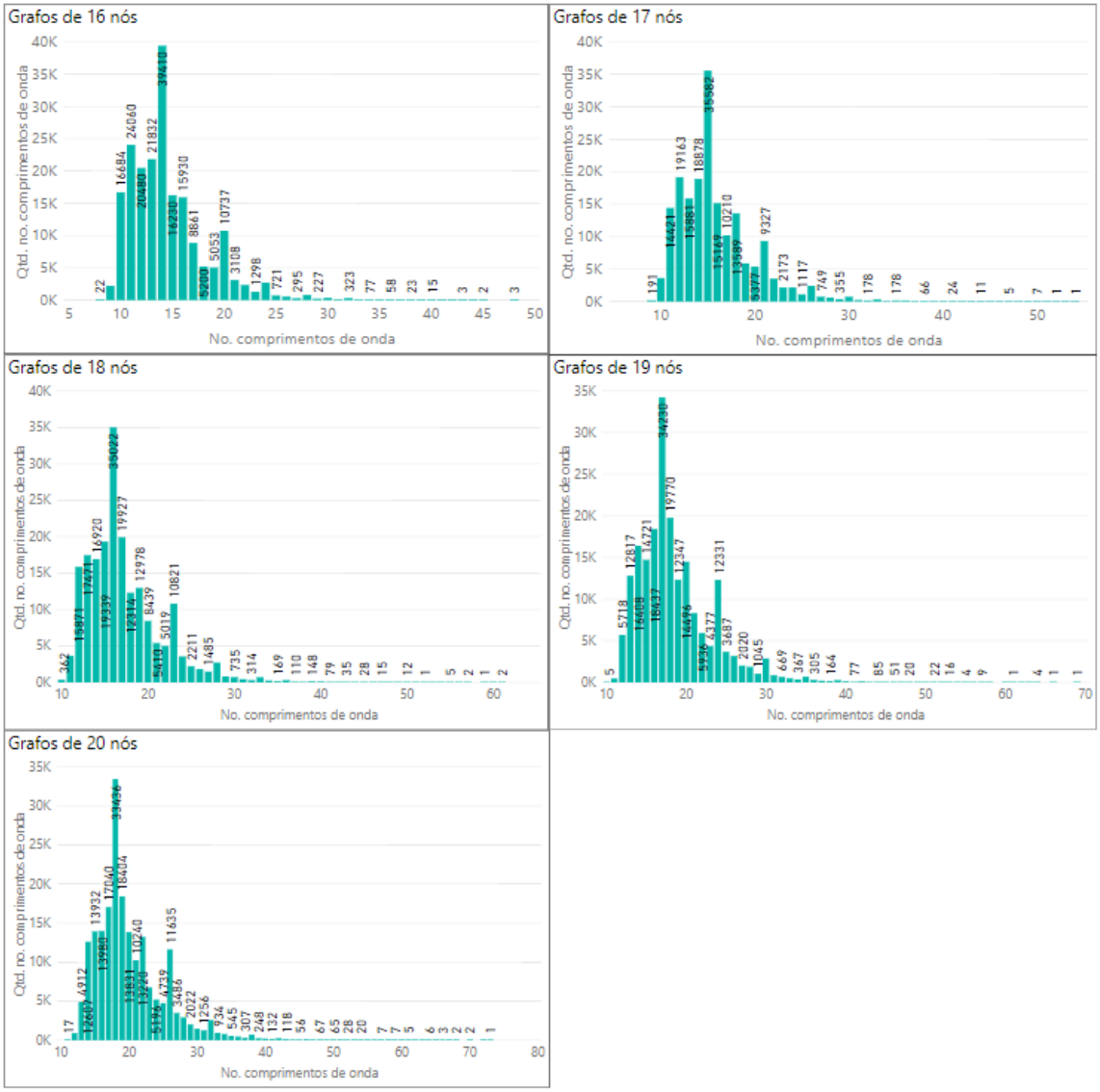
Fonte: Produção do próprio autor.

Quadro 3 – Comparativo entre o número mínimo de arestas dos grafos que minimizam λ ($m_{min(\lambda)}$), o número de arestas mínimo e o número de arestas máximo dos conjuntos separados por n

	n										
	10	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Mínimo $m_{min(\lambda)}$	17	20	22	24	26	27	30	32	34	37	38
Mínimo m	11	12	13	14	16	16	18	19	22	22	23
Máximo m	19	21	23	25	27	29	31	33	35	37	39

Fonte: Produção do próprio autor.

Figura 22 – Histogramas de grafos com n entre 16 e 20, relacionando a quantidade de grafos por λ



Fonte: Produção do próprio autor.

5 CONCLUSÕES E TRABALHOS FUTUROS

5.1 Conclusões

O presente trabalho busca encontrar parâmetros topológicos de redes ópticas de comunicação - modeladas por meio de grafos - que se correlacionam com o número de comprimentos de onda. O processo se inicia analisando um conjunto de aproximadamente 2,2 milhões de redes, cujas topologias foram modeladas utilizando grafos. A modularidade é uma das principais métricas deste estudo, juntamente com a métrica de número de comprimentos de onda (λ). Essa métrica foi calculada conforme detalhado no Capítulo 3. Em seguida, os grafos foram separados usando como critério o número de vértices (n) e, então, foram comparadas as métricas Modularidade - obtidas de cada um dos nove métodos aplicados - com λ, através de gráficos de dispersão. Após obter um conjunto de grafos que minimizam o λ, fixou-se o λ mínimo e realizou-se uma nova comparação de métricas com gráficos de dispersão: Modularidade e Número de Arestas (m).

Primeiramente, quando comparadas as métricas modularidade e λ, foi possível identificar que nem todos os métodos de detecção de comunidades aplicados tiveram valores de modularidade que se correlacionaram com λ, ou seja, os gráficos de dispersão para alguns métodos apresentaram uma correlação fraca. Já para outros métodos, essa tendência foi visível e, a partir deles, foi feita a análise das duas métricas. Dessa análise, concluiu-se que à medida que se aumenta a modularidade, o valor de λ tende a aumentar. Entretanto, o conjunto de grafos que minimizam o λ abrange uma faixa de valores de modularidade consideravelmente ampla, ou seja, a modularidade nesse caso se torna indiferente. Porém, foi possível obter uma melhor precisão na escolha desses grafos com λ mínimo, pois essa faixa corresponde a um percentual menor da faixa total dos valores de modularidade.

A partir da conclusão anterior, os grafos que minimizam λ foram analisados fixando-se o λ mínimo, comparando, agora, a modularidade com o número de arestas destes grafos. Essa comparação foi necessária para verificar se o número de arestas influenciava

no número de comprimentos de onda. A conclusão foi que os grafos que possuem o menor λ têm seu número de arestas grande. Dentre os valores de m apresentados desses grafos, foram escolhidos os menores, devido ao fato de que um número de arestas relativamente grande acarreta em um custo maior da rede.

5.2 Trabalhos Futuros

Como se pode observar na Tabela 2, os números mínimos de comprimentos de onda dos grafos com n = 10 e n = 11 são iguais, com λ = 4 para ambos os casos. Isso se repete para os grafos com n = 15, 16, 18 e 19 (para 15 e 16 nós, λ = 8 e para 18 e 19 nós, λ = 10). Com isso, é proposto como trabalho futuro analisar a escalabilidade de uma rede, pois, por esses casos observados, se a rede é aumentada em 1 vértice, ela pode conseguir se manter com o mesmo número de comprimentos de onda. Observa-se que o quando ocorre esse aumento, o número de grafos com o mesmo λ diminui significativamente. Por exemplo, de n = 10 para n = 11, o conjunto reduziu de 99 para 5 grafos. Com isso, a ideia seria estudar esse conjunto pequeno de grafos e descobrir quais os parâmetros topológicos que fazem esses grafos manterem seus números de comprimentos de onda iguais ao do conjunto com n - 1.

REFERÊNCIAS BIBLIOGRÁFICAS

AIRES, V. P.; NAKAMURA, F. G. Detecção de Comunidades em Redes Sociais: Relacionando o Método LFouvain a Medidas de Centralidade. *In*: SOCIEDADE BRASILEIRA DE COMPUTAÇÃO (SBC), 2017, São Paulo. **Anais do XXXVI Concurso de Trabalhos de Iniciação Científica da SBC**. São Paulo, 2017.

ALMEIDA, J. D.; THOULESS, D. J. Stability of the sherrington-kirkpatrick solution of a spin glass model. **Journal of Physics A: Mathematical and General**, IOP Publishing, v. 11, n. 5, p. 983, 1978.

BANKER, K. **MongoDB in action**. USA: Manning Publications Co., 2011.

BARABÁSI, A.-L. **Network science**. Cambridge: Cambridge University Press, 2016.

BEZANSON, J.; EDELMAN, A.; KARPINSKI, S.; SHAH, V. B. Julia: A fresh approach to numerical computing. **SIAM review**, SIAM, v. 59, n. 1, p. 65–98, 2017.

BOHLIN, L.; EDLER, D.; LANCICHINETTI, A.; ROSVALL, M. Community detection and visualization of networks with the map equation framework. In: **Measuring scholarly impact**. Cham: Springer, 2014. p. 3–34.

BONFIM, J. F. **Codificação de fonte**. 2014. Dissertação (Mestrado em Matemática) – Programa de Pós-Graduação em Matemática, Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista (UNESP), São Paulo, 2014.

BRANDES, U.; DELLING, D.; GAERTLER, M.; GÖRKE, R.; HOEFER, M.; NIKOLOSKI, Z.; WAGNER, D. On finding graph clusterings with maximum modularity. *In:* INTERNATIONAL WORKSHOP ON GRAPH-THEORETIC CONCEPTS IN COMPUTER SCIENCE, 33., 2007, Berlin. **Proceedings of the 33rd International Workshop on Graph-Theoretic Concepts in Computer Science**. Berlim: Lecture Notes in Computer Science, 2007. p. 121–132.

COUSINEAU, M.; PERRON, S.; CAPOROSSI, G.; PAIVA, M.; SEGATTO, M. RWA problem with geodesics in realistic OTN topologies. **Optical Switching and Networking**, Elsevier, v. 15, p. 18–28, 2015.

DEPIZZOL, D. B. **Uma Abordagem Estatística para o Projeto de Topologias Físicas de Redes Ópticas**. 2018. Tese (Doutorado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica, Centro Tecnológico, Universidade Federal do Espírito Santo, Vitória, 2018.

DEPIZZOL, D. B.; MONTALVÃO, J.; LIMA, F. O.; PAIVA, M. H. M.; SEGATTO, M. E. V. Feature selection for optical network design via a new mutual information estimator. **Expert Systems with Applications**, Elsevier, v. 107, p. 72–88, 2018.

DEPIZZOL, D. B.; PAIVA, M. H. M.; SEGATTO, M. E. V. Evaluating community detection methods in a controlled experiment. *In*: 2017 CHILEAN CONFERENCE ON ELECTRICAL, ELECTRONICS ENGINEERING, INFORMATION AND COMMUNICATION TECHNOLOGIES (CHILECON), 2017, Pucon. **Proceedings of the 2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)**. Pucon: [s.n.], 2017. p. 1–6.

DIESTEL, R. **Directions in Infinite Graph Theory and Combinatorics: With an introduction by C. St. JA Nash-Williams**. Amsterdam: Elsevier, 2016. v. 3.

DUGUÉ, N.; PEREZ, A. **Directed Louvain: maximizing modularity in directed networks**. Orléans, 2015. (Relatório Técnico). Disponível em: https://hal.archives-ouvertes.fr/hal-01231784. Acesso em: 31 jul. 2019.

FLAKE, G. W.; LAWRENCE, S.; GILES, C. L. Efficient identification of web communities. *In*: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 2000, Nova Iorque. **Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.** Nova Iorque: [s.n.], 2000. p. 150–160.

FURTADO, A. **Teoria dos grafos: algoritmos**. Rio de Janeiro: Livros Técnicos e Científicos, 1973. (Ciência da Computação).

GRIECHISCH, E.; PLUHÁR, A. Community detection by using the extended modularity. **Acta cybernetica**, v. 20, n. 1, p. 69–85, 2011.

IHAKA, R.; GENTLEMAN, R. R: a language for data analysis and graphics. **Journal of computational and graphical statistics**, Taylor & Francis Group, v. 5, n. 3, p. 299–314, 1996.

ISHIO, H.; MINOWA, J.; NOSU, K. Review and status of wavelength-division-multiplexing technology and its application. **Journal of Lightwave Technology**, v. 2, p. 448 – 463, 1984.

KARLIN, S. **A first course in stochastic processes**. Nova Iorque: Academic Press, 2014.

NEWMAN, M. E.; GIRVAN, M. Finding and evaluating community structure in networks. **Physical Review E**, American Physical Society, v. 69, n. 2, p. 026113, 2004.

NEWMAN, M. E. J. Finding community structure in networks using the eigenvectors of matrices. **Physical Review E**, American Physical Society, v. 74, p. 036104, 2006.

PONS, P.; LATAPY, M. Computing communities in large networks using random walks. *In*: INTERNATIONAL SYMPOSIUM ON COMPUTER AND INFORMATION SCIENCES, 2005, Istambul. **Proceedings of the International Symposium on Computer and Information Sciences.** Istambul: [s.n.], 2005. p. 284–293.

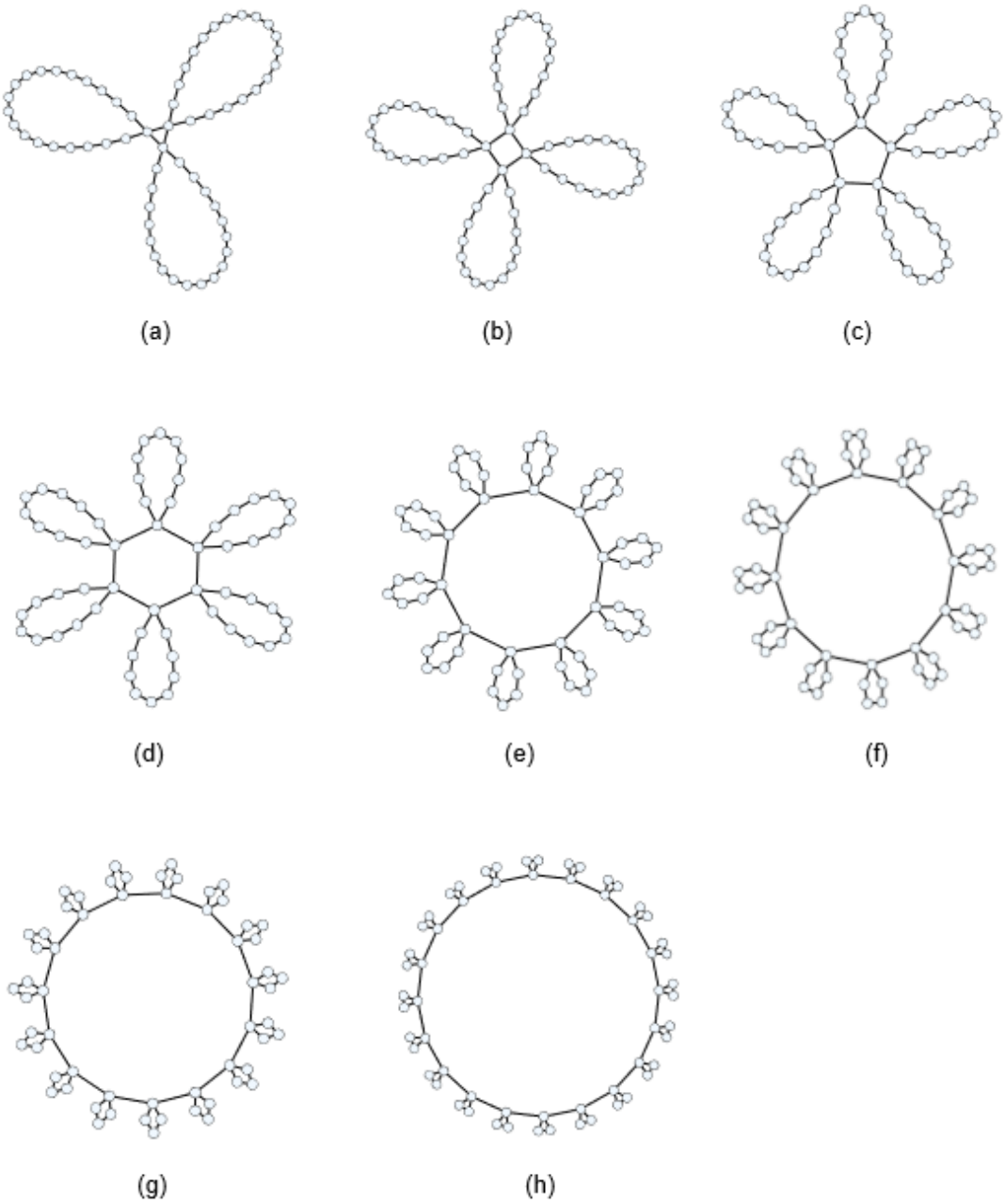
RADICCHI, F.; CASTELLANO, C.; CECCONI, F.; LORETO, V.; PARISI, D. Defining and identifying communities in networks. **Proceedings of the National Academy of Sciences of the United States of America**, National Acad. Sciences, v. 101, n. 9, p. 2658–2663, 2004.

ROSVALL, M.; AXELSSON, D.; BERGSTROM, C. T. The map equation. **The European Physical Journal Special Topics**, Springer, v. 178, n. 1, p. 13–23, 2009. ISSN 1951-6355.

SEVERANCE, C. Discovering javascript object notation. **Computer**, IEEE, v. 45, n. 4, p. 6–8, 2012.

APÊNDICE A – TOPOLOGIAS DE REDES DE 60 NÓS

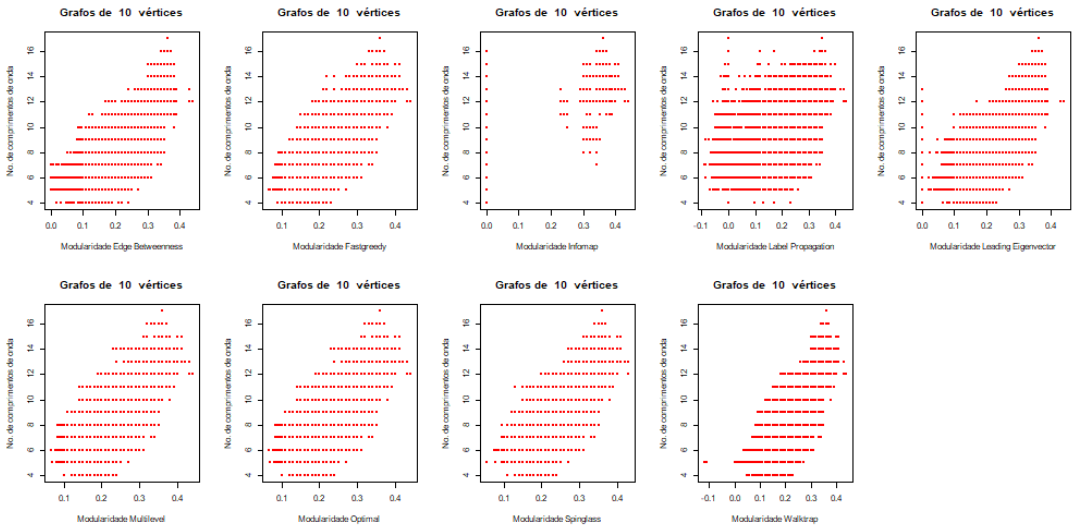
Figura 23 – Redes de anéis hierárquicos de 60 nós divididos de forma homogênea, sendo (a): rede com 3 anéis de 20 nós cada; (b): rede com 4 anéis de 15 nós cada; (c): rede com 5 anéis de 12 nós cada; (d): rede com 6 anéis de 10 nós cada; (e): rede com 10 anéis de 6 nós cada; (f): rede com 12 anéis de 5 nós cada; (g): rede com 15 anéis de 4 nós cada, e (h): rede com 20 anéis de 3 nós cada



Fonte: Depizzol, Paiva e Segatto (2017).

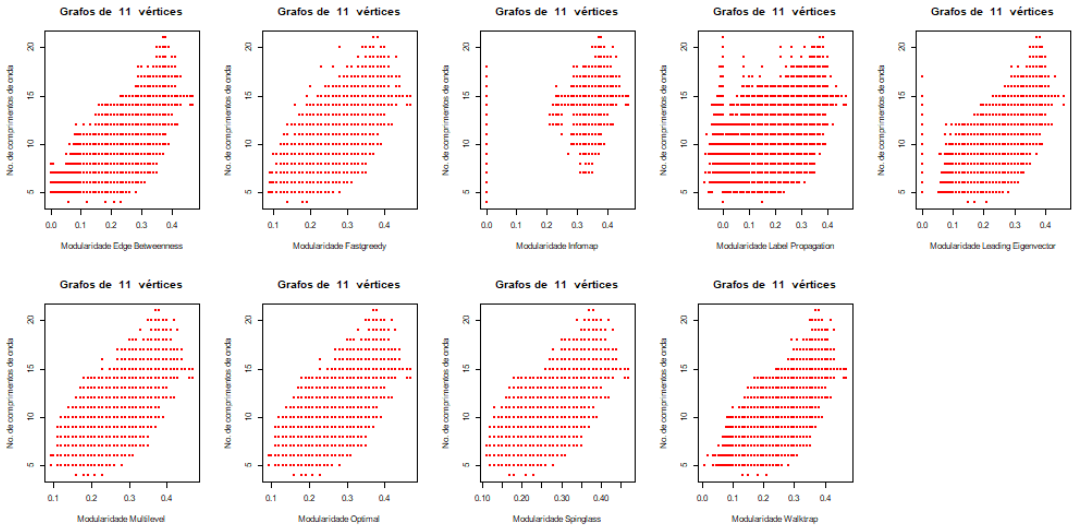
APÊNDICE B – GRÁFICOS DE DISPERSÃO DE MODULARIDADE *VERSUS* λ DE TODOS OS CONJUNTOS DE GRAFOS DIVIDIDOS POR NÚMERO DE NÓS

Figura 24 – Gráficos de dispersão do conjunto de grafos $n = 10$, comparando modularidade com λ



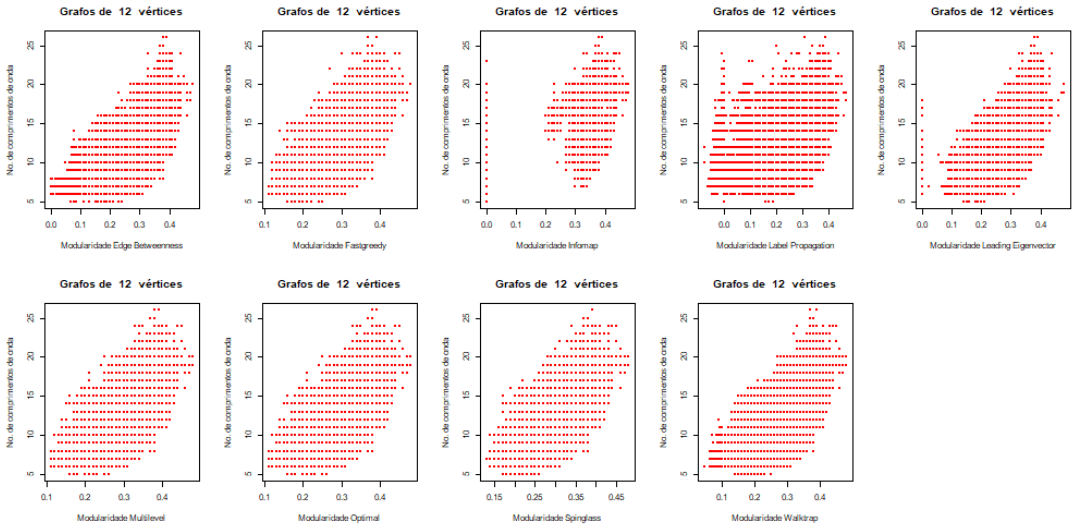
Fonte: Produção do próprio autor.

Figura 25 – Gráficos de dispersão do conjunto de grafos $n = 11$, comparando modularidade com λ



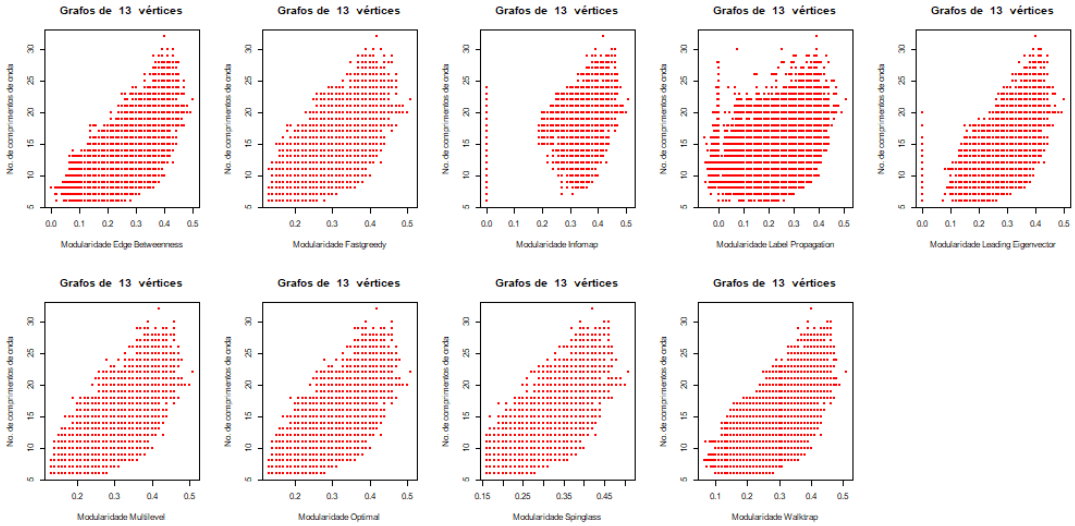
Fonte: Produção do próprio autor.

Figura 26 – Gráficos de dispersão do conjunto de grafos $n = 12$, comparando modularidade com λ



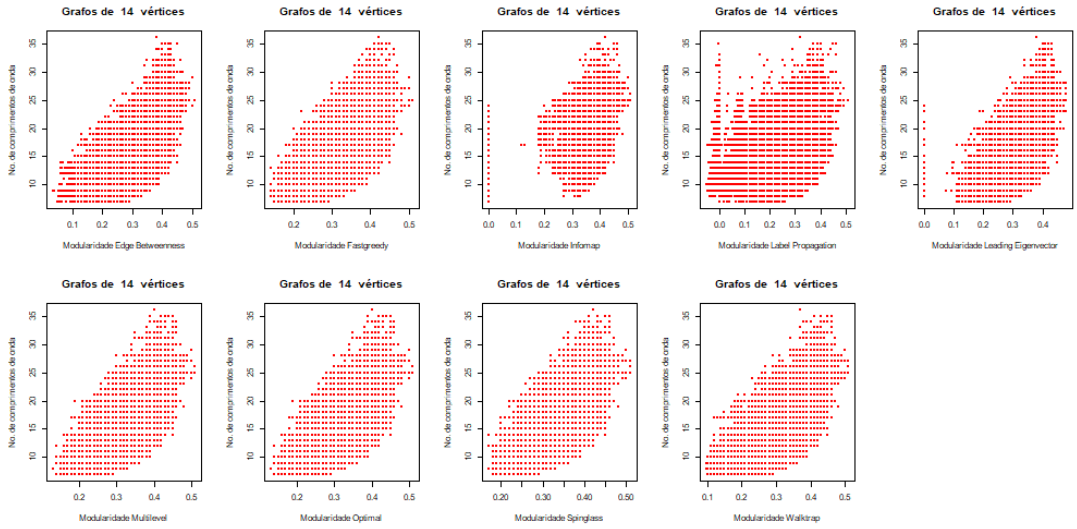
Fonte: Produção do próprio autor.

Figura 27 – Gráficos de dispersão do conjunto de grafos $n = 13$, comparando modularidade com λ



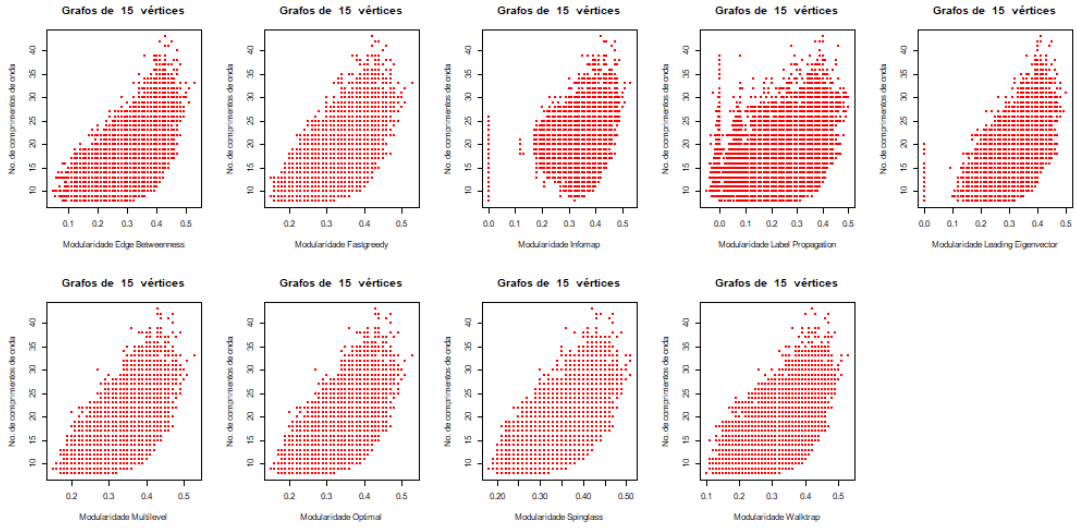
Fonte: Produção do próprio autor.

Figura 28 – Gráficos de dispersão do conjunto de grafos $n = 14$, comparando modularidade com λ



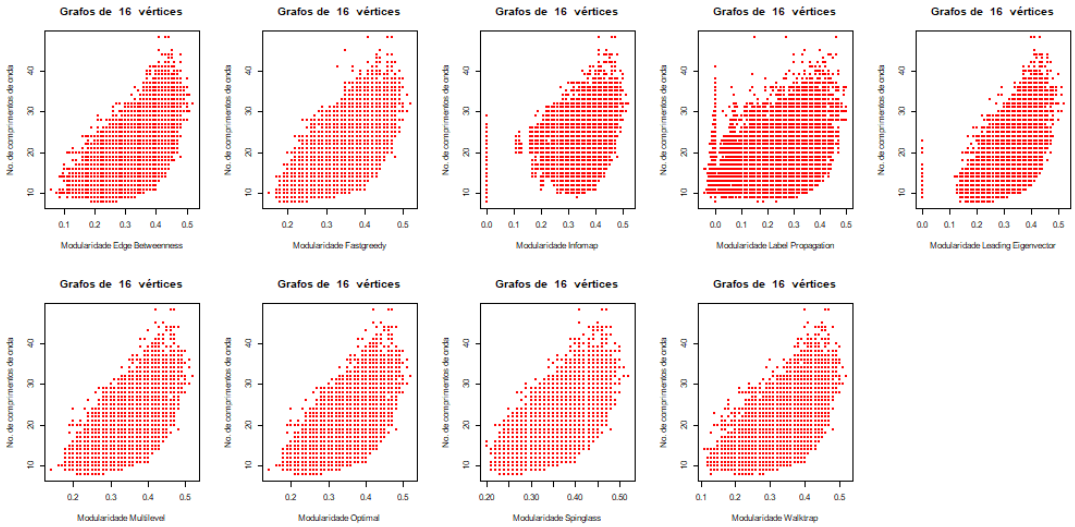
Fonte: Produção do próprio autor.

Figura 29 – Gráficos de dispersão do conjunto de grafos $n = 15$, comparando modularidade com λ



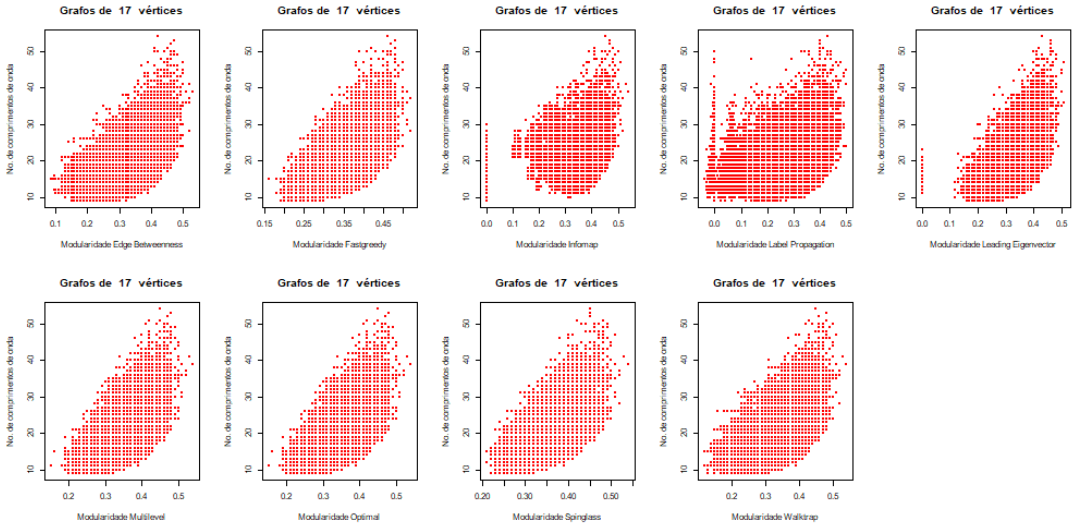
Fonte: Produção do próprio autor.

Figura 30 – Gráficos de dispersão do conjunto de grafos $n = 16$, comparando modularidade com λ



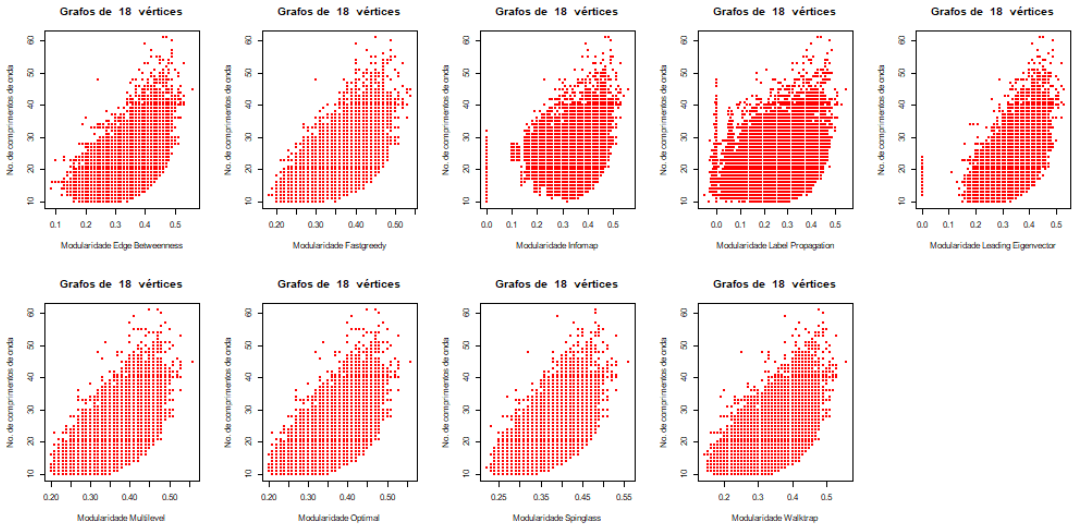
Fonte: Produção do próprio autor.

Figura 31 – Gráficos de dispersão do conjunto de grafos $n = 17$, comparando modularidade com λ



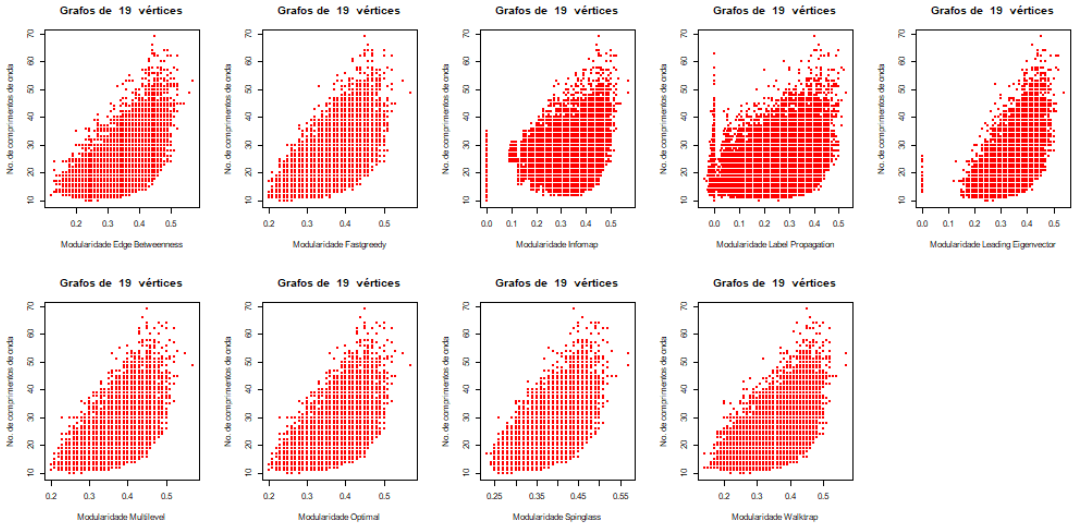
Fonte: Produção do próprio autor.

Figura 32 – Gráficos de dispersão do conjunto de grafos $n = 18$, comparando modularidade com λ



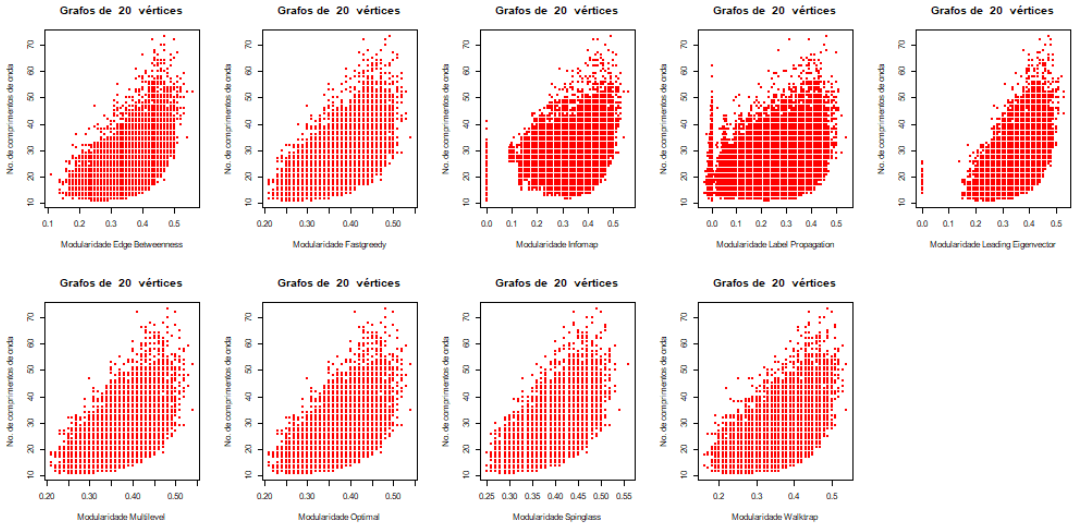
Fonte: Produção do próprio autor.

Figura 33 – Gráficos de dispersão do conjunto de grafos $n = 19$, comparando modularidade com λ



Fonte: Produção do próprio autor.

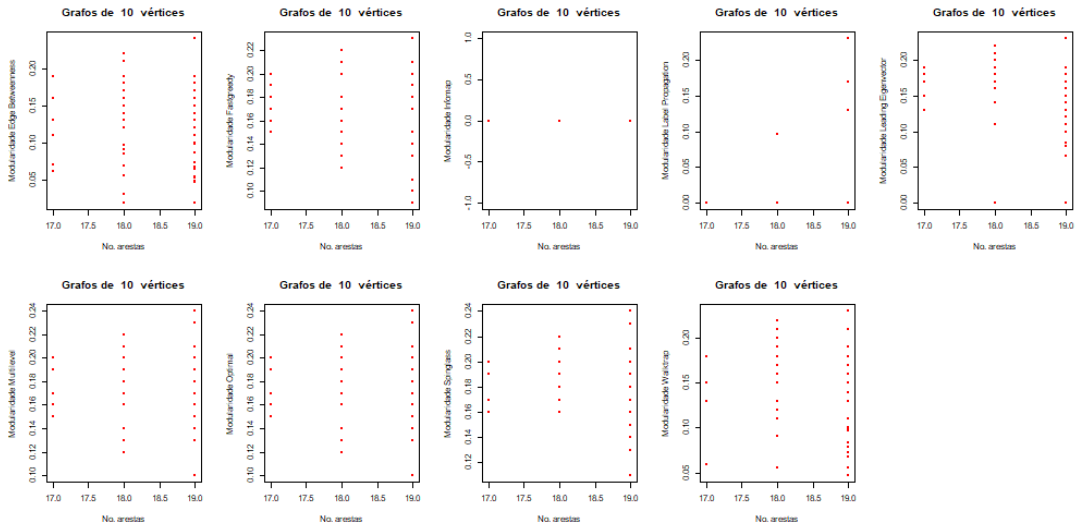
Figura 34 – Gráficos de dispersão do conjunto de grafos $n = 20$, comparando modularidade com λ



Fonte: Produção do próprio autor.

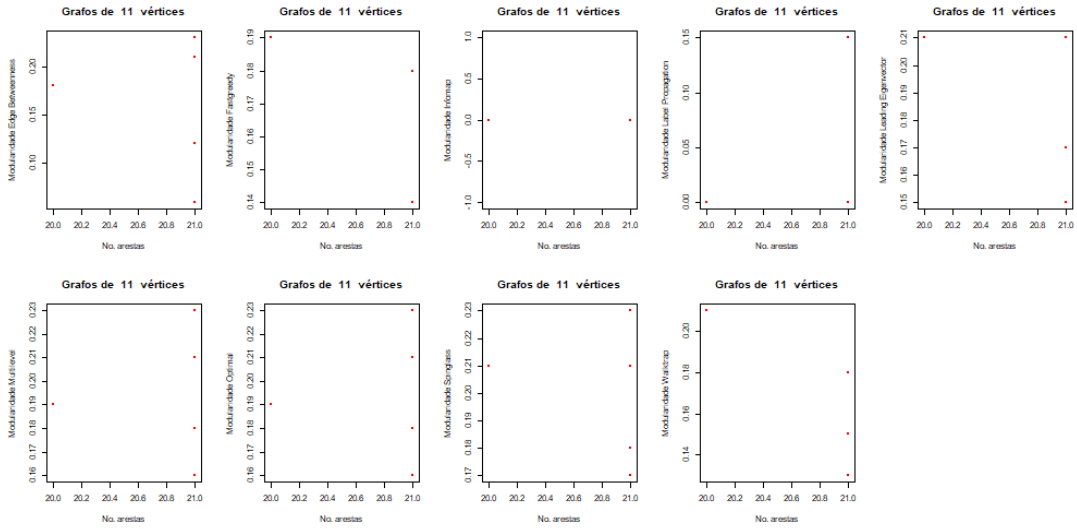
APÊNDICE C – GRÁFICOS DE DISPERSÃO DE MODULARIDADE *VERSUS* NÚMERO DE ARESTAS DE TODOS OS CONJUNTOS DE GRAFOS DIVIDIDOS POR NÚMERO DE NÓS

Figura 35 – Gráficos de dispersão do conjunto de grafos $n = 10$, comparando modularidade com número de arestas



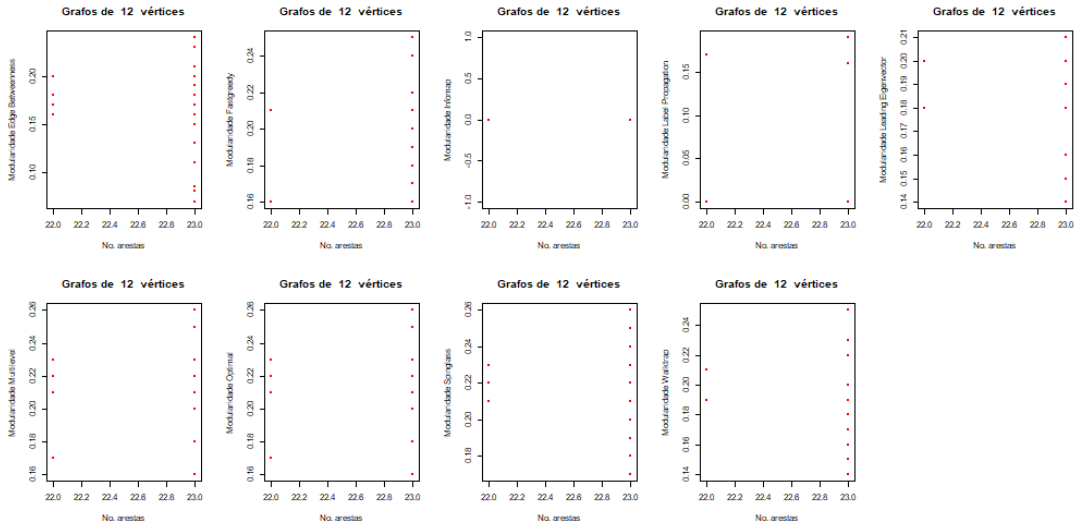
Fonte: Produção do próprio autor.

Figura 36 – Gráficos de dispersão do conjunto de grafos $n = 11$, comparando modularidade com número de arestas



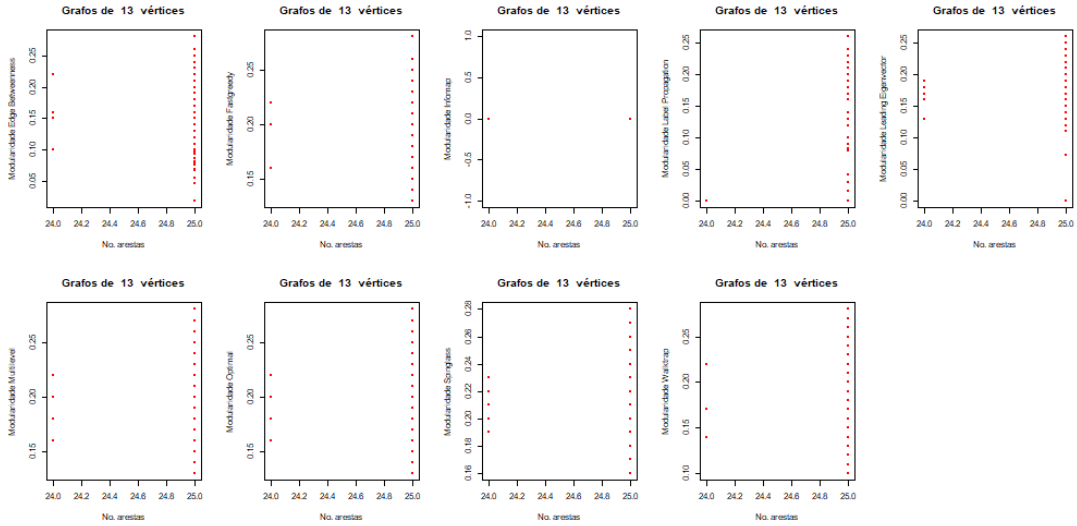
Fonte: Produção do próprio autor.

Figura 37 – Gráficos de dispersão do conjunto de grafos $n = 12$, comparando modularidade com número de arestas



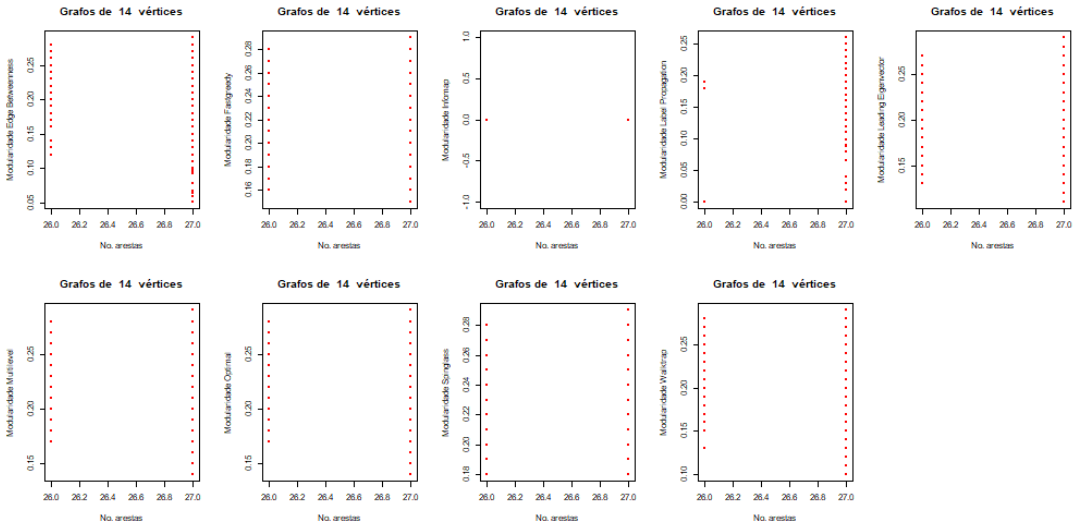
Fonte: Produção do próprio autor.

Figura 38 – Gráficos de dispersão do conjunto de grafos $n = 13$, comparando modularidade com número de arestas



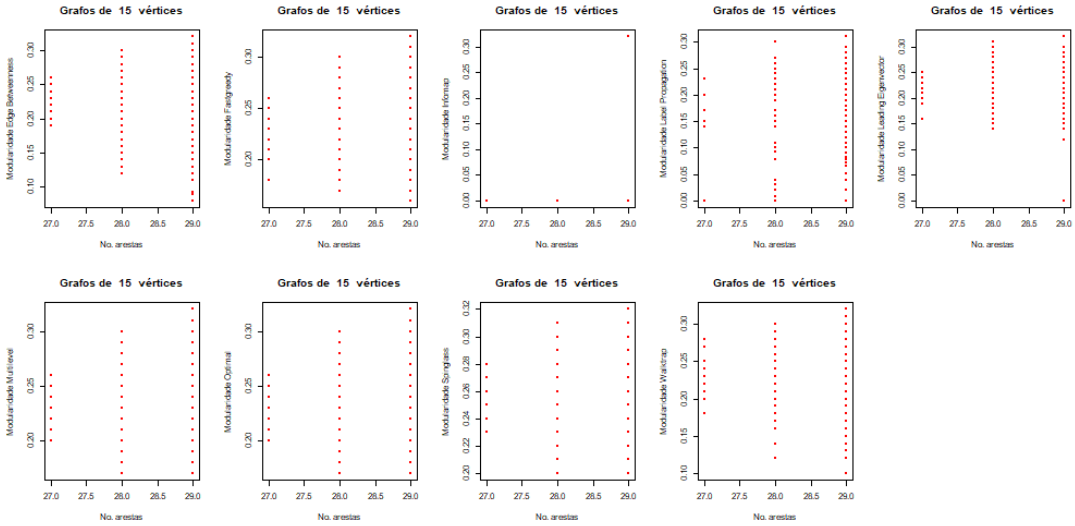
Fonte: Produção do próprio autor.

Figura 39 – Gráficos de dispersão do conjunto de grafos $n = 14$, comparando modularidade com número de arestas



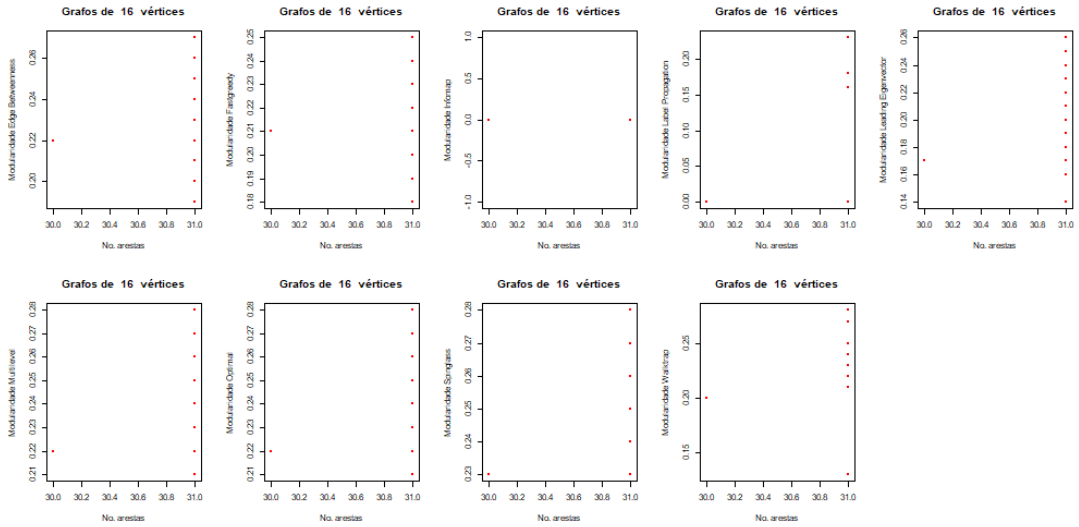
Fonte: Produção do próprio autor.

Figura 40 – Gráficos de dispersão do conjunto de grafos $n = 15$, comparando modularidade com número de arestas



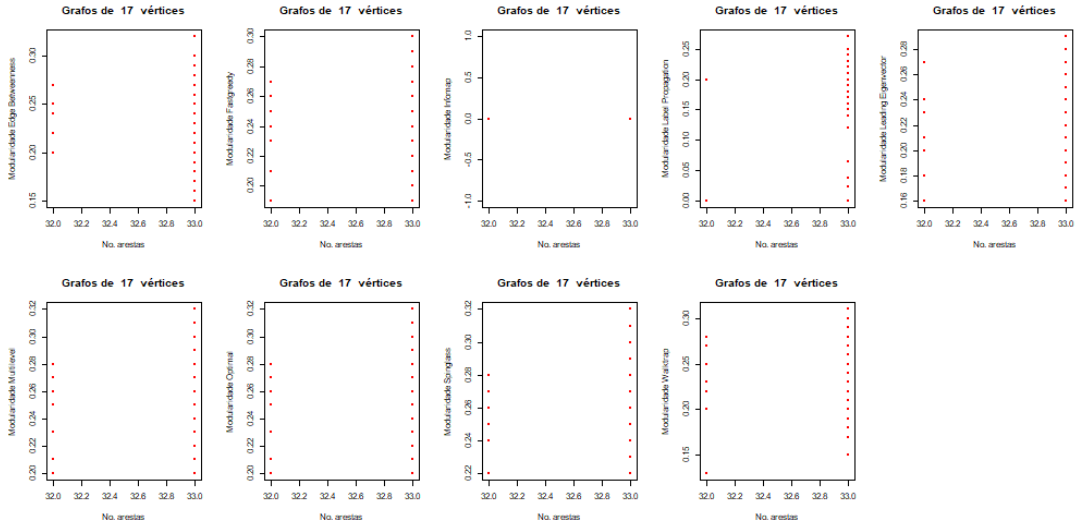
Fonte: Produção do próprio autor.

Figura 41 – Gráficos de dispersão do conjunto de grafos $n = 16$, comparando modularidade com número de arestas



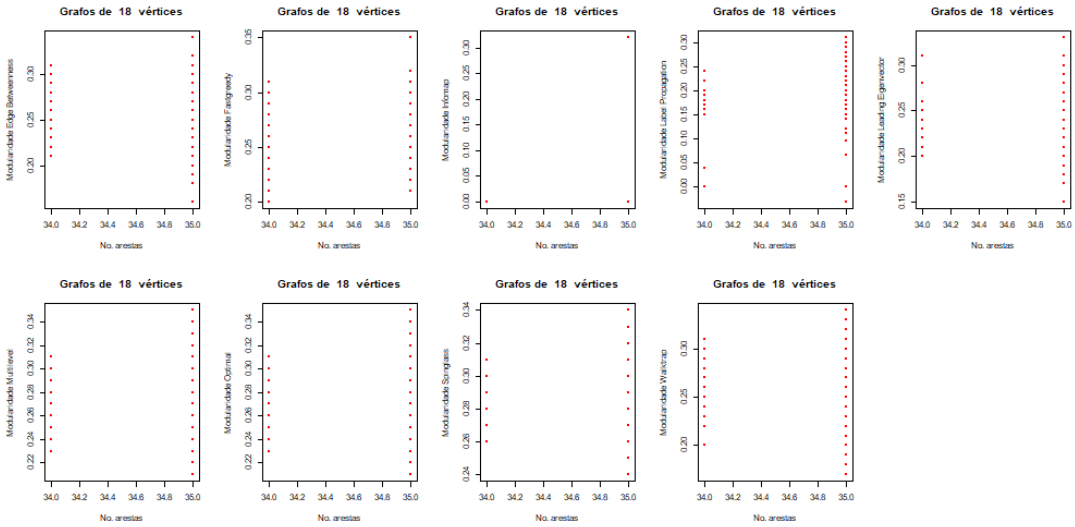
Fonte: Produção do próprio autor.

Figura 42 – Gráficos de dispersão do conjunto de grafos $n = 17$, comparando modularidade com número de arestas



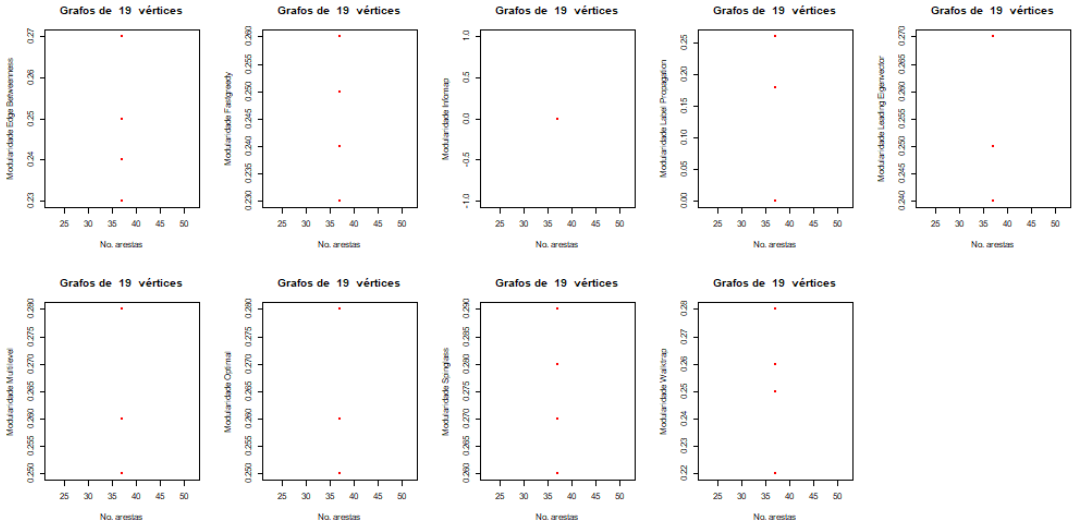
Fonte: Produção do próprio autor.

Figura 43 – Gráficos de dispersão do conjunto de grafos $n = 18$, comparando modularidade com número de arestas



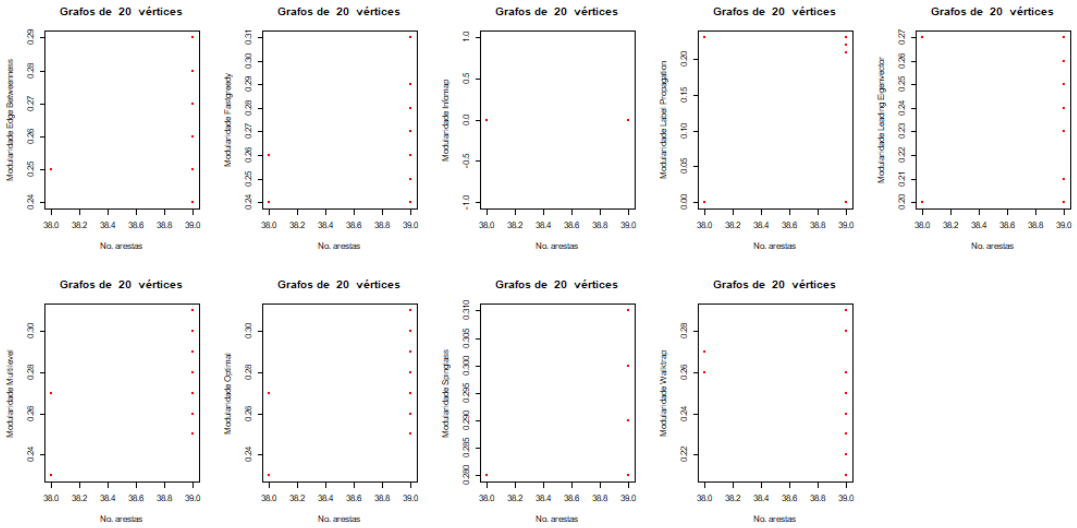
Fonte: Produção do próprio autor.

Figura 44 – Gráficos de dispersão do conjunto de grafos $n = 19$, comparando modularidade com número de arestas



Fonte: Produção do próprio autor.

Figura 45 – Gráficos de dispersão do conjunto de grafos $n = 20$, comparando modularidade com número de arestas



Fonte: Produção do próprio autor.