

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ - UTFPR
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL
PROFMAT**

MARCÉLIA PICOLOTTO DA SILVA TEIXEIRA

DETECÇÃO DE COMUNIDADES EM GRAFOS

CURITIBA

2019

MARCÉLIA PICOLOTTO DA SILVA TEIXEIRA

DETECÇÃO DE COMUNIDADES EM GRAFOS

Dissertação apresentada ao Mestrado Profissional em Matemática em Rede Nacional da Universidade Tecnológica Federal do Paraná em Curitiba - PROFMAT-UTCT como requisito parcial para obtenção do grau de Mestre.

Orientador: João Luis Gonçalves

CURITIBA

2019

Dados Internacionais de Catalogação na Publicação

T266d Teixeira, Marcélia Picolloto da Silva
Detecção de comunidades em grafos [recurso eletrônico] /
Marcélia Picolloto da Silva Teixeira.-- 2019.
1 arquivo texto (53 f.) : PDF : 1,26 MB.

Modo de acesso: World Wide Web.

Texto em português com resumo em inglês.

Dissertação (Mestrado) - Universidade Tecnológica
Federal do Paraná. Programa de Mestrado Profissional em
Matemática em Rede Nacional. Área de Concentração: Análise
Numérica, Curitiba, 2019.

Bibliografia: f. 43-44.

1. Matemática - Dissertações. 2. Teoria dos grafos. 3. Álgebra
linear. 4. Matrizes (Matemática). 5. Autovalores. 6. Autovetores.
7. Redes sociais on-line. 8. Professores de matemática -
Formação. 9. Prática de ensino. 10. Matemática - Estudo e ensino
(Ensino médio). I. Gonçalves, João Luis, orient. II. Universidade
Tecnológica Federal do Paraná. Programa de Mestrado
Profissional em Matemática em Rede Nacional. III. Título.

CDD: Ed. 23 -- 510

Biblioteca Central do Câmpus Curitiba – UTFPR
Bibliotecária: Luiza Aquemi Matsumoto CRB-9/794

TERMO DE APROVAÇÃO DE DISSERTAÇÃO Nº 64

A Dissertação de Mestrado intitulada “Detecção de Comunidades em Grafos”, defendida em sessão pública pela candidata Marcélia Picolloto da Silva Teixeira, no dia 22 de abril de 2019, foi julgada para a obtenção do título de Mestre, área de concentração Análise Numérica, e aprovada em sua forma final, pelo Programa de Pós-Graduação em Matemática em Rede Nacional – PROFMAT.

BANCA EXAMINADORA:

Prof. Dr. João Luis Gonçalves - Presidente – UTFPR

Prof. Dr. Roy Wilhelm Probst - UTFPR

Prof. Dr. José Carlos Corrêa Eidam - UFPR

A via original deste documento encontra-se arquivada na Secretaria do Programa, contendo a assinatura da Coordenação após a entrega da versão corrigida do trabalho.

Curitiba, 22 de abril de 2019.

Carimbo e Assinatura do(a) Coordenador(a) do Programa

AGRADECIMENTOS

Primeiramente agradeço a Deus por ter estado ao meu lado, por me agraciar com serenidade e sabedoria, guiando minhas escolhas para o bem. Por me dar força para suportar o desafio de me manter focada no estudo por esses três anos, conseguindo conciliar a segunda licenciatura em Física com esse mestrado profissional.

Aos meus pais por me apoiarem todas as vezes que me mantive ausente para estudar. Em especial, agradeço aos meus filhos que de certa forma não aceitaram, mas respeitaram a minha decisão de cursar esse mestrado. Tainá, por todas as traduções dos livros, artigos e textos e Eduardo por não ter cumprido a ameaça de queimar meus cadernos. E deixo aqui registrado as minhas desculpas pelas viagens que não fizemos, pelos filmes que não assistimos, pelas trilhas que não desbravamos, pelos piqueniques que não comemos e pelas conversas que não tivemos.

Ao amor da minha vida que, mesmo sem imaginar, seria o grande incentivo dessa jornada.

À CAPES pela recomendação do PROFMAT por meio do parecer do Conselho Técnico Científico da Educação Superior e pelo incentivo financeiro.

À Sociedade Brasileira de Matemática que na busca da melhoria do ensino de Matemática na Educação Básica viabilizou a implementação do PROFMAT.

Aos meus professores do PROFMAT que grandemente contribuíram para que a minha prática em sala de aula fosse enriquecida, através dos conhecimentos por eles empregados.

Ao meu orientador pela sua insistência, por não ter me deixado desistir. Por todas as conceitos que me ensinou, entre eles o da humildade. Agradeço o tempo que disponibilizou para esse projeto e por ter me apresentado um ramo da matemática fabuloso. Por ter despertado minha curiosidade e por não ter medido esforços para que esse trabalho chegasse ao fim.

RESUMO

PICOLOTTO DA SILVA TEIXEIRA, Marcélia. **Detecção de Comunidades em Grafos**. 54 f. Dissertação - Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, Universidade Tecnológica Federal do Paraná. Curitiba, 2019.

Neste trabalho apresentamos o método do corte mínimo para a detecção de comunidades em grafos, ilustramos o método com exemplos e propomos um roteiro de trabalho, para a aplicação desse método pelo professor de matemática. O objetivo desse roteiro é que o professor seja capaz de implementar uma sequência de aulas com conteúdos abordados de forma não usual no Ensino Médio, em que caberá aos estudantes detectar comunidades em grafos que representem redes de seu interesse. No método e nas atividades propostas o conteúdo de matrizes é tratado de uma forma diferente da usual, sem associação a resolução de sistemas lineares, usando autovalores e autovetores.

Palavras-chave: Grafos. Autovalores. Detecção de Comunidades.

ABSTRACT

PICOLOTTO DA SILVA TEIXEIRA, Marcélia. **Detection of Communities in Graphs**. 54 pg. Dissertation - Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, Universidade Tecnológica Federal do Paraná. Curitiba, 2019.

In this work we present the minimum cut method for the detection of communities in graphs, we illustrate the method with examples and we propose an work routine such that the math teacher could apply communities detection activity for high school students, where it is up to them to detect communities in graphs that represent networks of their daily lives. In the proposed method and activities the matrix subject is treated in a unusual way, not associated with the resolution of linear systems, applying eigenvalues and eigenvectors.

Keywords: Graph. Eigenvalues. Communities Detection.

LISTA DE ILUSTRAÇÕES

Figura 1 – Grafo com 5 vértices 6 arestas.	12
Figura 2 – Grafo da família Silva	13
Figura 3 – Grafo ponderado.	14
Figura 4 – Diagrama das pontes de Konigsberg no documento original (Euler, 1741) .	21
Figura 5 – 7 comunidades apresentadas em (BEVERIDGE; SHAN, 2016)	29
Figura 6 – Comunidades detectadas pelo corte mínimo	30
Figura 7 – Charge de Simon Kneebone sobre as pontes de Konigsberg (à esquerda), representação gráfica do problema (centro) e grafo modelo gerado (à direita)	34
Figura 8 – Topologia de redes de comunicação (à esquerda) e bases nitrogenadas (à direita)	37
Figura 9 – Grafo de uma rede social	37
Figura 10 – Ícones de redes sociais	38
Figura 11 – Modelo para planilha de adjacência.	39

SUMÁRIO

	INTRODUÇÃO	10
1	CONCEITOS FUNDAMENTAIS DA TEORIA DE GRAFOS	12
2	COMUNIDADES EM GRAFOS	20
2.1	Aspectos Históricos da Detecção de Comunidades	20
2.2	Redes Sociais	22
2.3	Corte Mínimo	22
2.4	CORTE MÍNIMO EM GRAFOS PONDERADOS	26
3	EXPERIMENTO COMPUTACIONAL	28
3.1	Comunidades do Livro <i>Storm of Swords</i>	28
4	PROPOSTA DE APLICAÇÃO EM NÍVEL MÉDIO	32
4.1	Embasamento teórico para o professor	32
4.2	Proposta de aula sobre conceitos fundamentais	33
4.2.1	Aula sobre teoria de grafos	34
4.2.2	Aula sobre autovalor e autovetor	34
4.2.3	Aulas sobre álgebra linear aplicada à teoria de grafos	35
4.3	Proposta de roteiro de aula sobre comunidades e redes sociais	36
4.3.1	Aula sobre conceito de rede e comunidade	36
4.4	Proposta de atividade prática para a detecção de comunidades	38
4.4.1	Aula questionário para detecção de comunidades	38
4.4.2	Aula sobre corte mínimo e detecção de comunidades no R	40
4.5	Estudo de caso	41
5	CONCLUSÃO	43
	REFERÊNCIAS	44
	ANEXOS	46
	ANEXO A – CÓDIGO EM R PARA O MÉTODO DO CORTE MÍNIMO	47
	ANEXO B – CÓDIGO EM R PARA DETECÇÃO DE COMUNIDADE USANDO A MODULARIDADE E O MÉTODO DE LOU- VAIN	54

INTRODUÇÃO

Os grafos são objetos matemáticos simples e particularmente eficientes para descrever diversos fenômenos. Além disso, a teoria por trás da amigável modelagem com grafos é bastante robusta e suportada em outras áreas da matemática. Portanto, ao usar grafos descrevemos bem o problema e também abrimos um grande leque de ferramentas matemáticas para a resolução desses problemas.

O problema incipiente da teoria de grafos foi o Problema das Sete Pontes de Königsberg. Em (EULER, 1741), a cidade de Königsberg era cortada por um rio de forma que duas ilhas são formadas e ligando as margens do rio existiam sete pontes. O problema em questão é se seria possível atravessar todas as sete pontes passando apenas uma vez por cada uma. Euler resolveu o problema em 1741, usando grafos, mas a aplicabilidade da teoria de grafos na ciência não parecia relevante. Em 1847, Kirchhoff, que coincidentemente é natural de Königsberg, usou a ideia de grafos para estudar circuitos elétricos, gerando a teoria de árvores, que chamou a atenção de outros cientistas para a teoria de grafos.

Muitos problemas atuais estão sendo modelados usando a teoria de grafos, por exemplo problemas associados a redes sociais, design de redes de transportes, identificação de complexos de proteínas, mapeamento de atividades neurais e pareamento entre grupos. Além desses, alguns problemas mais abstratos e intrínsecos à teoria de grafos, com grande potencial de aplicação são comuns, como os de coloração, planificação, caminho mínimo, caminhos eulerianos e caminhos hamiltonianos.

Várias dissertações, inclusive do PROFMAT, já trataram do tema teoria de grafos, veja por exemplo (ASSIS, 2016), (COSTA, 2017), (MALTA, 2008) e (NOGUEIRA, 2015). De forma geral, esses trabalhos apresentam uma introdução à teoria de grafos e tratam um problema com uma abordagem voltada a aplicação no Ensino Médio. Embora não conste na grade curricular, a teoria de grafos tem interseções com diversos de seus conteúdos, como matrizes e determinantes, que são tratados nesse trabalho, e também funções, análise combinatória e probabilidade.

Nossos objetivos nesse trabalho são apresentar uma breve introdução à teoria de grafos e a alguns conceitos de álgebra linear, tratar o problema de detecção de comunidades e propor uma atividade compatível com o nível médio sobre grafos e detecção de comunidade.

O trabalho está estruturado da seguinte maneira: no Capítulo 1 apresentamos conceitos da teoria de grafos e de álgebra linear que serão fundamentais para os problemas tratados. No Capítulo 2, além da parte histórica e conceitos sobre redes apresentamos uma abordagem para o problema de detectar comunidades, no qual exploramos o método de corte mínimo, contido em (NEWMAN, 2006). No Capítulo 3 apresentamos a implementação do método discutido no capítulo anterior e os resultados obtidos no experimento computacional e comparamos esses

resultados com o mesmo experimento, apresentado em (BEVERIDGE; SHAN, 2016). E no Capítulo 4 apresentamos uma proposta de atividade de nível médio envolvendo a detecção de comunidades na qual conteúdos de matrizes e teoria de grafos deverão ser utilizados. Envolvemos, ainda, nesse capítulo, a utilização de um software computacional, ver (SANTANA, 2017). Os códigos para a implementação do método implementado estão disponíveis nos anexos deste trabalho.

1 CONCEITOS FUNDAMENTAIS DA TEORIA DE GRAFOS

Neste capítulo, apresentamos, de forma breve, os conceitos e resultados da teoria de grafos e da álgebra linear que são fundamentais ao tema que trataremos nos próximos capítulos, a detecção de comunidades. Os conceitos e demonstrações apresentadas nesse capítulo foram retiradas, essencialmente de (NETTO; JURKIEWICZ, 2009), (DIESTEL, 2005), (ORTIZ; MOREIRA, 2016) e (FEOLILOFF P.; WAKABAYASHI, 2012).

O primeiro conceito é justamente o de **grafo**, visto em (FEOLILOFF P.; WAKABAYASHI, 2012).

Definição 1.1. Um **Grafo** G é definido como um par de conjuntos, o conjunto dos vértices, V , e o das arestas, E , sendo os elementos de E pares de elementos de V . Denotamos G por $G(V, E)$.

Exemplo 1.2. Seja o grafo $G(V, E)$, com o conjunto de vértices $V = \{Ana, Edu, Isa, Tai, Fer\}$ e o conjunto de arestas

$$E = \{(Ana, Edu), (Edu, Isa), (Ana, Tai), (Tai, Fer), (Ana, Fer), (Edu, Fer)\}.$$

Este grafo pode ser representado como na Figura 1, mas outras representações também são possíveis.

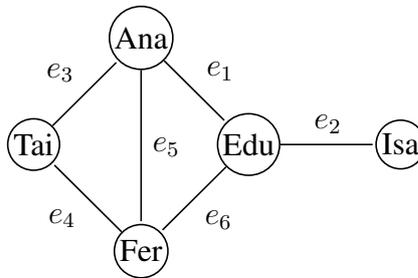


Figura 1 – Grafo com 5 vértices 6 arestas.

Neste exemplo, os vértices são pessoas e as arestas representam, por exemplo, a relação de amizade entre os vértices, isto é, $E = \{(v, w) | v \text{ é amigo de } w\}$. Como consideramos a amizade uma relação simétrica, pois se v é amigo de w então, w é amigo de v , as arestas que ligam os vértices não possuem orientação.

Assim chamamos $G(V, E)$ no Exemplo 1.2 de um **grafo não direcionado**, ou seja, um grafo em que as arestas são definidas como pares não ordenados de vértices.

Existem situações em que o sentido das ligações entre os vértices é importante. Nesse caso, as arestas são definidas como pares ordenados de vértices, sendo o primeiro vértice do par, o vértice de saída e o segundo o vértice, de chegada. Nesse contexto, as arestas possuem

um sentido que é representado por uma seta, que aponta do vértice de saída para o vértice de chegada. Segundo (NETTO; JURKIEWICZ, 2009), grafos com essa característica são chamados **grafos direcionados**.

Exemplo 1.3. *Seja o grafo $G(V, E)$, com o conjunto de vértices, $V = \{ \text{João Santos, Maria Santos Silva, Paulo Silva, Antônio Santos, Antônia Silva, José Silva} \}$, membros de uma mesma família, e o conjunto de arestas, $E = \{ (\text{João}, \text{Antônio}), (\text{Maria}, \text{Antônio}), (\text{Maria}, \text{Antônia}), (\text{Paulo}, \text{Antônia}), (\text{Antônia}, \text{José}) \}$. Esse grafo pode ser representado pela Figura 2.*

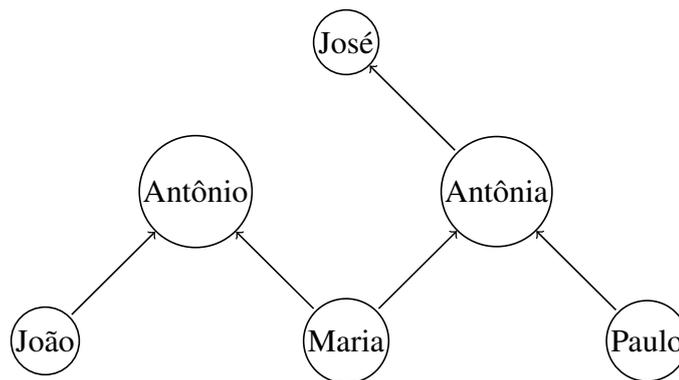


Figura 2 – Grafo da família Silva

O conjunto das arestas, nesse exemplo, é definido como $E = \{ (v, w) | v \text{ é pai/mãe de } w \}$. A relação definida por E não é simétrica pois se v é pai ou mãe de w , não implica que w é pai ou mãe de v . Há, portanto, uma orientação na relação, com um correspondente efeito na representação gráfica de G , portanto esse grafo é direcionado.

A cardinalidade dos conjuntos que definem o grafo são propriedades importantes e são definidas como a seguir.

Definição 1.4. *Define-se $|V|$, a **ordem** de G , como o número de vértices do grafo $G(V, E)$ e, analogamente, define-se $|E|$, o **tamanho** de G , como o número de arestas de $G(V, E)$.*

Então, usando a Definição 1.4 e a Figura 1, temos que a ordem do grafo do Exemplo 1.2 é 5 e o seu tamanho é 6. Já para o grafo do Exemplo 1.3, da Figura 2 temos que, a ordem é 6 e o tamanho é 5.

Em algumas situações queremos representar, usando grafos, as relações entre diferentes vértices, que têm entre si intensidades diferentes. Para contemplar essas diferentes intensidades, introduzimos a definição de **grafo ponderado**.

Definição 1.5. *Um grafo é dito **ponderado** quando suas arestas possuem pesos. Usamos a notação c_{ij} para denotar o peso da aresta entre os vértices i e j .*

Considere representar o número de interações entre cada par de amigos em uma rede social. Neste caso, podemos escolher o peso c_{ij} de cada aresta (i, j) , como o número de interações entre os vértices i e j , por exemplo considere o grafo da Figura 3.

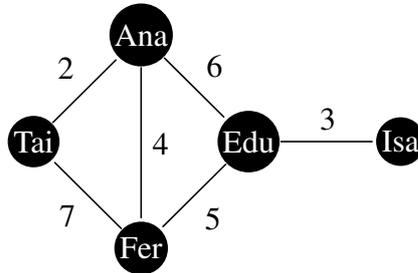


Figura 3 – Grafo ponderado.

Analisando a Figura 3, fica visualmente claro que que Tai e Fer interagiram 7 vezes, na rede social, enquanto Edu e Isa interagiram por 3 vezes.

Uma vez que um vértice não necessariamente tem arestas com todos os demais vértices, o número de vértices com que cada vértice está conectado também é uma característica importante do grafo. Essa característica está no **grau** de cada vértice e para todo o grafo na **matriz de graus**, conforme definidos a seguir.

Definição 1.6. *Define-se grau de um vértice v , denotado por $d(v)$, como o número de arestas que incidem em v . Caso o grafo seja ponderado, $d(v)$ é a soma dos pesos das arestas que incidem em v .*

Logo na Figura 1, do Exemplo 1.2, podemos observar que o grau de Ana é 3, o grau de Tai é 2 e o de Isa é 1.

Definição 1.7. *A matriz de graus $D(G)$ de G é a matriz diagonal de ordem n cujas entradas D_{ij} são:*

$$D_{ij} = \begin{cases} d(v_i) & \text{se } i = j \\ 0 & \text{se } i \neq j. \end{cases}$$

Ou seja, a matriz de graus possui as entradas da diagonal principal como sendo o grau de cada vértice.

Para o grafo do Exemplo 1.2, a matriz de graus é:

$$D = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}$$

Para o grafo ponderado apresentado na Figura 3, a matriz de graus ponderada, denotada por D_p , é:

$$D_p = \begin{bmatrix} 12 & 0 & 0 & 0 & 0 \\ 0 & 14 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 & 16 \end{bmatrix}$$

Em um grafo não direcionado, dois vértices v e w são adjacentes se há uma aresta $a = (v, w)$ em E . Por exemplo, os vértices Ana e Edu no grafo da Figura 1 são adjacentes.

A informação sobre as adjacências entre todos os vértices do grafo é sintetizada na matriz de adjacência, definida a seguir.

Definição 1.8. *Seja $G = G(V, E)$ um grafo com n vértices. A **matriz adjacência** A_{ij} de G é a matriz quadrada de ordem n de entradas:*

$$A_{ij} = \begin{cases} 1, & \text{se } (i, j) \in E \text{ para } i, j \in V, \\ 0, & \text{caso contrário.} \end{cases} \quad (1.1)$$

Se o grafo for ponderado, a **matriz de adjacência** é definida como:

$$A_{p_{ij}} = \begin{cases} c_{ij}, & \text{se } (i, j) \in E \text{ para } i, j \in V, \\ 0, & \text{caso contrário.} \end{cases} \quad (1.2)$$

O grafo representado na Figura 1, por exemplo, tem como matriz adjacência

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

Para o grafo ponderado da Figura 3 tem seguinte matriz de adjacência ponderada

$$A_p = \begin{bmatrix} 0 & 6 & 0 & 2 & 4 \\ 6 & 0 & 3 & 0 & 5 \\ 0 & 3 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 7 \\ 4 & 5 & 0 & 7 & 0 \end{bmatrix}.$$

Definição 1.9. *Sejam D a matriz diagonal dos graus dos vértices de um grafo G e A a matriz de adjacência de G . A **matriz laplaciana** do grafo G é a matriz L , definida por*

$$L = D - A. \quad (1.3)$$

A matriz laplaciana é relevante em diversos contextos da teoria de grafos, pois reúne em si toda a informação sobre a conectividade do grafo.

Em especial, o segundo menor autovalor da matriz laplaciana, denominado conectividade algébrica, desempenha um papel relevante em diversas aplicações, por exemplo os problemas de fluxo de redes.

Um grande número de propriedades relacionadas ao espectro da matriz laplaciana é apresentado em (ROCHA, 2015). Neste trabalho utilizaremos apenas o autovetor de Fiedler, autovetor associado ao segundo menor autovalor da matriz laplaciana, para a detecção de comunidades.

Aplicando a Definição 1.9, no Exemplo 1.2, obtemos

$$L = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & -1 & 0 & -1 & -1 \\ -1 & 3 & -1 & 0 & -1 \\ 0 & -1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 2 & -1 \\ -1 & -1 & 0 & -1 & 3 \end{bmatrix} \quad (1.4)$$

Essa mesma definição pode ser aplicada para grafos ponderados, como o grafo apresentado na Figura 3, cuja matriz laplaciana ponderada é:

$$L_p = \begin{bmatrix} 12 & 0 & 0 & 0 & 0 \\ 0 & 14 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 & 16 \end{bmatrix} - \begin{bmatrix} 0 & 6 & 0 & 2 & 4 \\ 6 & 0 & 3 & 0 & 5 \\ 0 & 3 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 7 \\ 4 & 5 & 0 & 7 & 0 \end{bmatrix} = \begin{bmatrix} 12 & -6 & 0 & -2 & -4 \\ -6 & 14 & -3 & 0 & -5 \\ 0 & -3 & 3 & 0 & 0 \\ -2 & 0 & 0 & 9 & -7 \\ -4 & -5 & 0 & -7 & 16 \end{bmatrix} \quad (1.5)$$

A matriz laplaciana é fonte de grande número de propriedades do grafo associado a ela, como já mencionamos. No entanto, para os fins desta dissertação apenas algumas propriedades da matriz laplaciana são relevantes.

Uma delas, é o fato de que a soma dos termos da i -ésima coluna (ou linha) da matriz laplaciana é zero, ou seja,

$$\sum_j L_{ij} = 0. \quad (1.6)$$

Por exemplo, para a matriz da Equação 1.4, o somatório das entradas da segunda linha $-1+3-1+0-1 = 0$, assim como o somatório das entradas da quarta coluna, $-1+0+0+2-1 = 0$. O mesmo acontece na matriz da Equação 1.5, para qualquer linha ou coluna.

A justificativa para essa propriedade é que as entradas da i -ésima linha, fora da diagonal representam a existência de arestas adjacentes ao vértice i , logo a soma dessas entradas resulta no grau do vértice i , que é justamente o valor que consta na i -ésima entrada da diagonal, com sinal oposto. Portanto a soma de todas as entradas será zero.

Considerando a definição matriz laplaciana, essa propriedade continua sendo válida também para a matriz laplaciana ponderada.

Essa propriedade da matriz laplaciana é útil no Teorema 1.12. Mas antes de enunciá-lo, apresentamos as definições de autovalor e autovetor de uma matriz.

Definição 1.10. *Seja $A \in M_n(\mathbb{R})$, uma matriz $n \times n$. Um par (λ, \vec{v}) , sendo $\lambda \in \mathbb{R}$ e \vec{v} um vetor do \mathbb{R}^n , com $\vec{v} \neq \vec{0}$, é um par de **autovalor** e de **autovetor** de A se*

$$Av = \lambda \vec{v}.$$

Os autovalores e autovetores podem ser determinados como a solução para o sistema $(A - \lambda I)\vec{v} = \vec{0}$. Para matrizes pequenas a resolução desse sistema é simples, porém para matrizes grandes faz-se necessário o uso de métodos computacionais que em geral oferecem uma aproximação para os autovalores e autovetores de uma matriz.

Vejam no Exemplo 1.11 como determinar os autovalores e os autovetores associados a uma matriz.

Exemplo 1.11. Seja a matriz $A = \begin{bmatrix} 4 & 3 \\ 2 & -1 \end{bmatrix}$, aplicando a Definição 1.10,

$$Av = \lambda v \implies Av - \lambda v = 0 \implies Av - \lambda I_2 v = 0 \implies (A - \lambda I)v = 0,$$

então

$$\begin{aligned} \left(\begin{bmatrix} 4 & 3 \\ 2 & -1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right) \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} 4 - \lambda & 3 \\ 2 & -1 - \lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{aligned} \tag{1.7}$$

Resolvendo o determinante obtemos

$$(4 - \lambda)(-1 - \lambda) - 6 = 0 \implies \lambda^2 - 3\lambda - 10 = 0 \implies \begin{cases} \lambda_1 = 5 \\ \lambda_2 = -2 \end{cases}$$

Os autovetores associados aos autovalores, λ_1 e λ_2 , são determinados com a Equação 1.7.

Para $\lambda_1 = 5$

$$\begin{aligned} \begin{bmatrix} 4 - \lambda & 3 \\ 2 & -1 - \lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} &\implies \begin{bmatrix} -1 & 3 \\ 2 & -6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &\implies \begin{cases} -x + 2y = 0 \\ 3x - 6y = 0 \end{cases} \implies x = 2y \end{aligned}$$

Logo, para o autovalor $\lambda_1 = 5$, está associado o autovetor $v = (2y, y)^T$, com $y \neq 0$.

Para $\lambda_1 = -2$

$$\begin{aligned} \begin{bmatrix} 4 - \lambda & 3 \\ 2 & -1 - \lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} &\implies \begin{bmatrix} 6 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &\implies \begin{cases} 6x + 2y = 0 \\ 3x + y = 0 \end{cases} \implies -3x = y \end{aligned}$$

Logo, para o autovalor $\lambda_1 = -2$, está associado o autovetor $v = (x, -3x)^T$, com $x \neq 0$.

Teorema 1.12. *Seja G um grafo com n vértices e L sua matriz laplaciana. Sejam $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ os autovalores de L . Então:*

- (i) $\lambda_1 = 0$ e o vetor $\vec{1}$ é o autovetor associado.
- (ii) G é conexo se, e somente se, $\lambda_2 > 0$.

As demonstrações desses resultados estão disponíveis em (ROCHA, 2015). No item (i), 0 ser um autovalor é consequência de $\sum_j L_{ij} = 0$, pois para $\vec{v}_1 = \vec{1} = (1, 1, \dots, 1)$ temos $L\vec{v}_1 = \vec{0} = 0\vec{v}_1$.

Como todos os autovalores de L são não negativos, $\lambda_1 = 0$ é o menor autovalor e seu vetor normalizado se dá por $\vec{v}_1 = \frac{1}{\sqrt{n}}(1, 1, \dots)$.

Os grafos que consideraremos nos próximos capítulos são grafos não direcionados e portanto suas matrizes de adjacências, de graus e laplacianas, ponderadas ou não, serão simétricas. Matrizes simétricas têm boas propriedades, principalmente quanto a seus autovalores e autovetores.

Apresentamos a seguir os resultados para matrizes simétricas que são mais relevantes para essa dissertação. Esses e outros resultados pertinentes podem ser encontrados em (DIESTEL, 2005).

Teorema 1.13. *Se A é uma matriz real simétrica, então seus autovalores são reais.*

Demonstração. Seja $z = a + bi$ um número complexo, chamamos de conjugado de z o número $\bar{z} = a - bi$. Se A_{ij} é a matriz com entradas complexas, o conjugado de A_{ij} é denotado por \bar{A} . Se A é uma matriz real $A = \bar{A}$ se deve ao fato de que o conjugado de um número real é o próprio número.

Seja λ autovalor de A com autovetor correspondente v . Então $Av = \lambda v$ e, usando conjugados complexos, temos $\overline{Av} = \overline{\lambda v}$. Segue que,

$$A\bar{v} = \overline{Av} = \overline{\lambda v} = \bar{\lambda}v = \bar{\lambda}\bar{v}$$

pois A é real. Usando o fato de que A é simétrica e tomando as transpostas, temos

$$\bar{v}^T A = \bar{v}^T A^T = (A\bar{v})^T = (\bar{\lambda}\bar{v})^T = \bar{\lambda}\bar{v}^T$$

logo,

$$\lambda(\bar{v}^T v) = \bar{v}^T(\lambda v) = \bar{v}^T(A\lambda) = v(\bar{v}^T A) = v(\bar{\lambda}\bar{v}^T) = \bar{\lambda}(\bar{v}^T v)$$

assim,

$$\lambda(\bar{v}^T v) - \bar{\lambda}(\bar{v}^T v) = 0 \implies (\lambda - \bar{\lambda})(\bar{v}^T v) = 0$$

. Se

$$v = \begin{bmatrix} a_1 + b_1 i \\ \vdots \\ a_n + b_n i \end{bmatrix}$$

então,

$$\bar{v} = \begin{bmatrix} a_1 - b_1 i \\ \vdots \\ a_n - b_n i \end{bmatrix}$$

e assim,

$$\bar{v}^T v = (a_1^2 + b_1^2) + \dots + (a_n^2 + b_n^2) \neq 0$$

. Já que $v \neq 0$, por ser autovetor. Concluimos que

$$\lambda - \bar{\lambda} = 0 \implies \lambda = \bar{\lambda}$$

. Logo λ é real. □

Outro resultado importante diz sobre a ortogonalidade dos autovetores de uma matriz simétrica.

Teorema 1.14. *Autovetores, associados a autovalores distintos, de matrizes reais simétricas são ortogonais.*

Demonstração. Sejam A uma matriz real simétrica, v e w autovetores de A associados aos autovalores λ_1 e λ_2 , respectivamente, com $\lambda_1 \neq \lambda_2$. Temos que $Av = \lambda_1 v$ e $Aw = \lambda_2 w$. Segue então que:

$$w^T(\lambda_1 v) = \lambda_1 w^T v = \lambda_1 v^T w = (\lambda_1 v)^T w = (Av)^T w = A^T v^T w = v^T(Aw) = v^T(\lambda_2 w),$$

e portanto

$$\lambda_1(v^T w) = \lambda_2(v^T w).$$

Como $\lambda_1 \neq \lambda_2$, temos que $v^T w = 0$ portanto v e w são ortogonais. □

2 COMUNIDADES EM GRAFOS

As redes sociais virtuais, devido a grande adesão, são uma fonte de informações que podem impulsionar mercados já existentes e mesmo criar novos mercados, ou mesmo fomentar política públicas que melhorem a vida das pessoas. Entretanto, essa quantidade enorme de informação precisa de tratamento para que se torne um conhecimento útil. Alguns dos tratamentos mais relevantes para dados tão complexos e tão grandes é o agrupamento em comunidades, que dependendo do contexto pode ajudar a resolver a diferentes problemas. Assim, detectar comunidades tornou-se um problema muito interessante e suas aplicações vão além de redes sociais.

Comumente grafos representam redes, sendo os vértices representantes dos objetos da rede e as arestas caracterizando as conexões entre esses objetos.

Uma comunidade em uma rede ou grafo pode ser caracterizada de diferentes formas. Consideraremos apenas uma caracterização de comunidades, a caracterização via corte mínimo, apresentada em (NEWMAN, 2006). Nessa caracterização cada comunidade é um subconjunto de vértices e o grafo é dividido em comunidades disjuntas e não vazias de forma que o número de arestas que conectam vértices de diferentes comunidades seja o menor possível. Outras condições podem ser impostas como o número de comunidades e o número de vértices em cada comunidade.

Observemos que as comunidades impõem uma partição a um grafo, mas uma partição não necessariamente representa as comunidades de um grafo.

Apresentaremos nas seções a seguir aspectos históricos e conceituais acerca da detecção de comunidades, bem como um método para realizar essa detecção em grafos.

2.1 ASPECTOS HISTÓRICOS DA DETECÇÃO DE COMUNIDADES

Um dos primeiros problemas matemáticos que foi preciso resolver recorrendo à teoria dos grafos é o conhecido problema das pontes de Königsberg.

Segundo (RODRIGUES, 2009), a cidade de Königsberg, conhecida hoje, por Kalinegrado, na Rússia, fica nas margens do rio Pregel. No meio desse rio existem duas ilhas e sete pontes unem as quatro massas de terra que o permeiam. O problema que se punha era o de saber se seria possível percorrer todas as sete pontes num caminho que não as repetisse.

Euler resolveu esse problema em 1741, provando a inexistência de uma solução. Ele tomou as quatro massas de terra como vértices e as sete pontes como arestas, então o caminho desejado equivale a percorrer todas as arestas do grafo passando uma única vez por cada aresta. Porém tal caminho só é possível se zero ou dois vértices estejam conectados a um número par de

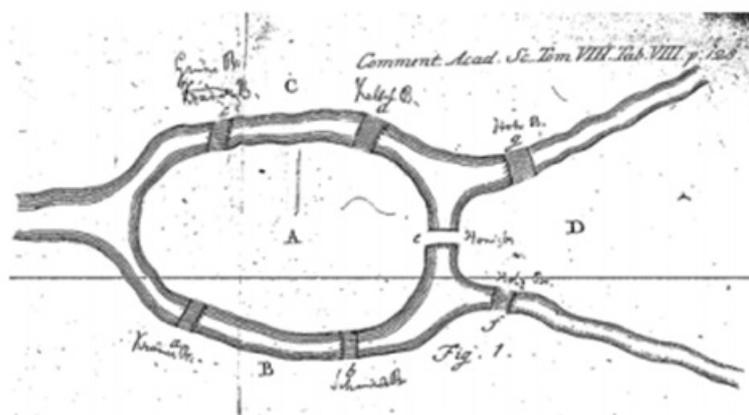


Figura 4 – Diagrama das pontes de Königsberg no documento original (Euler, 1741)

arestas. Como o grafo em questão tem seus 4 vértices com grau ímpar, não existe o caminho desejado. Uma discussão mais detalhada pode ser encontrada em (RODRIGUES, 2009).

De Sola Pool e Kochen escreveram em 1958 um artigo que só viria a ser publicado em 1978, ver (POOL; KOCHEN, 1978) em que questionam:

- Quantos indivíduos conhece cada um dos membros de uma rede? Ou em termos da teoria de grafos, qual o grau de cada pessoa na rede?
- Qual é a distribuição desses conhecimentos ou grau? Qual o valor médio e quais os valores maiores e menores?
- Que tipos de indivíduos tem o maior número de contatos? São estes os indivíduos mais influentes na rede social?
- Como é que os contatos se organizam? Qual é efetivamente a estrutura da rede?

Essas questões são precursoras do conceito de comunidade.

Esse artigo, mesmo antes de ser publicado, serviu de inspiração para Stanley Milgram que detectou pequenas comunidades em uma rede. Em parceria com Jeffrey Travers, em 1969, Stanley Milgram, escreveu um artigo com os detalhes desse experimento, conhecido por teoria dos seis graus de separação, veja em (MILGRAN; TRAVERS, 1969).

A partir daí vários modelos para a detecção de redes foram estudados. Os trabalhos mais recentes nas áreas de *Machine Learning* e Inteligência Artificial em alguns casos também realizam detecções de comunidades e por serem áreas em franco desenvolvimento e com grande potencial, a detecção de comunidades tornou-se um problema de destaque.

O campo da Teoria de Grafos, hoje em dia é vasto e muito formalismo matemático é necessário para compreender essa teoria. Nessa dissertação, estudaremos uma forma simples de detecção de comunidades, que seja possível de aplicar em nível médio e que motive os estudantes

prossequirem estudando temas relacionados. Para um conhecimento mais profundo sugerimos o livro (DIESTEL, 2005).

2.2 REDES SOCIAIS

Grafos estão inerentes em diversas áreas, mas são muitas vezes chamados de redes ou outros termos similares. Por exemplo as redes biológicas, redes de comunicação, redes de informação, redes de negócio, redes de mercados econômicos, redes sociais, entre outras.

As redes que usaremos para a proposta didática de aplicação da detecção de comunidades serão as redes sociais.

Em (LÉVI-STRAUSS, 1987), Claude Lévi-Strauss diz que cada um de nós é uma espécie de encruzilhada onde acontecem coisas. O Artigo (MAIA L. F., 2014) define rede social como uma rede na qual o conjunto de objetos são as pessoas e as conexões entre elas são relações sociais, ou seja, é uma estrutura composta por pessoas ou organizações, conectadas por diversos tipos de relações.

Atualmente, vários pesquisadores tem se dedicado a compreender quais são as propriedades dos grafos com o objetivo de conhecer como uma rede se comporta, gerando conhecimento útil a partir de um grafo.

Podemos perceber, nos últimos tempos, um forte crescimento nas redes sociais e nelas observamos vários modelos de interação. Dentre as redes sociais atuais aquelas que mais agradam os brasileiros, segundo o sistema de pesquisa Social Media Trends 2018, veja em (CONTENT, 2019), são: Instagram, Facebook, Pinterest, Twitter, WhatsApp e YouTube. Essas redes sociais mostram a atual tendência de organização e comunicação entre as pessoas.

2.3 CORTE MÍNIMO

Para detectar comunidades de redes, ou seja grafos que têm muitos vértices, algoritmos eficientes precisam ser criados. Aqui iremos abordar o método corte mínimo, em que o conjunto de vértices é dividido em subconjuntos disjuntos e não vazios de forma que o número de arestas que conectam vértices de diferentes grupos é o menor possível.

Um dos primeiros conceitos acerca do particionamento de grafos foi levantado por Kernighan e Lin (1970), segundo eles, o problema do particionamento de um grafo $G = (V,E)$ consiste em dividir o grafo em k subconjuntos de vértices. Para essa divisão são desconsideradas algumas arestas, aquelas que são adjacentes a vértices que não estão no mesmo grupo. O número de arestas desconsideradas para criar o particionamento é chamado de corte.

O método consiste portanto na partição dos vértices de um grafo em dois subconjuntos disjuntos de tal modo que o corte seja minimizado. Caso a partição obtida não represente satisfa-

toriamente as comunidades que procuramos, podemos particionar novamente as comunidades obtidas.

O essencial para esse método é modelar o corte em termos das comunidades.

Iniciamos definindo o corte, denotado por R , para um grafo G que tem matriz de adjacência A e matriz laplaciana L . O corte está associado a uma partição em dois grupos G_1 e G_2 e é definido como:

$$R = \frac{1}{2} \sum_{\substack{i \in G_1 \\ j \in G_2}} A_{ij}. \quad (2.1)$$

Na Equação 2.1 o fator $\frac{1}{2}$ compensa a contagem duplicada das arestas que compõe o corte. Essa definição só é aplicável se os grupos/comunidades são conhecidos e essa é uma limitação quando desejamos justamente encontrar esses grupos/comunidades.

Para reescrever o corte, R , de forma mais conveniente, consideramos um vetor de escolha, s , de forma que a i -ésima entrada é definida por:

$$s_i = \begin{cases} +1 & \text{se o vértice pertence a } G_1 \\ -1 & \text{se o vértice pertence a } G_2. \end{cases} \quad (2.2)$$

Note que s satisfaz a condição de $s^T s = n$ e ainda,

$$\frac{1}{2} (1 - s_i s_j) = \begin{cases} 1 & \text{se } i \text{ e } j \text{ são de grupos diferentes} \\ 0 & \text{se } i \text{ e } j \text{ são do mesmo grupo.} \end{cases} \quad (2.3)$$

Para o vetor de escolha definido dessa forma, podemos reescrever a Equação 2.1 como:

$$\begin{aligned} R &= \frac{1}{2} \sum_{i,j} A_{ij} \\ &= \frac{1}{2} \sum_{i,j} \frac{1}{2} (1 - s_i s_j) A_{ij} \\ &= \frac{1}{4} \sum_{i,j} (1 - s_i s_j) A_{ij}. \end{aligned} \quad (2.4)$$

O número de arestas conectadas ao vértice i , ou seja o grau do vértice i , k_i , pode ser representado como

$$k_i = \sum_j A_{ij}, \quad (2.5)$$

e usando a Equação 2.5, temos

$$\sum_{i,j} A_{ij} = \sum_i k_i = \sum_i s_i^2 k_i = \sum_{i,j} s_i s_j k_i \delta_{ij}, \quad (2.6)$$

na qual $s_i^2 = 1$, já que $s_i = \pm 1$ e

$$\delta_i = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{caso contrário.} \end{cases} \quad (2.7)$$

Portanto, o corte pode ser descrito como

$$R = \frac{1}{4} \sum_{i,j} s_i s_j (k_i \delta_{i,j} - A_{i,j}). \quad (2.8)$$

Observando que, $(k_i \delta_{i,j} - A_{i,j}) = L_{i,j}$, onde L é a matriz laplaciana do grafo, podemos escrever o corte R da seguinte forma

$$R = \frac{1}{4} \mathbf{s}^T L \mathbf{s}. \quad (2.9)$$

Nosso objetivo é encontrar \mathbf{s} que minimiza R e da Equação 2.9 temos que R está fortemente relacionado com a matriz laplaciana L .

Como L tem boas propriedades, simétrica, real e não negativa definida, é vantajoso usar o espectro de L . Usando o Teorema 1.14 para a matriz L , temos que os autovalores de L são não negativos, os autovetores associados a autovalores distintos são ortogonais e existe uma base ortonormal de autovetores de L para o \mathbb{R}^n .

Seja $\beta = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ uma base ortonormal de autovetores de L . O vetor de escolha \vec{s} pode, portanto, ser escrito como uma combinação linear dos vetores de β , isto é

$$\vec{s} = \sum_{i=1}^n a_i \vec{v}_i, \quad (2.10)$$

em que $a_i = \vec{v}_i^T \vec{s}$.

Devido a ortonormalidade da base β temos ainda que

$$n = \vec{s}^T \vec{s} = \left(\sum_{i=1}^n a_i \vec{v}_i \right)^T \vec{s} = \sum_{i=1}^n a_i (\vec{v}_i^T \vec{s}) = \sum_{i=1}^n a_i^2. \quad (2.11)$$

Então partindo da Equação 2.9 e suprimindo o fator $\frac{1}{4}$, o qual não interfere na minimização, temos,

$$\begin{aligned} R &= \mathbf{s}^T L \mathbf{s} \\ &= \left(\sum_{i=1}^n a_i \vec{v}_i^T \right) L \left(\sum_{j=1}^n a_j \vec{v}_j \right) \\ &= \sum_{i,j} a_i a_j \lambda_j \delta_{i,j} \\ &= \sum_{i=1}^n a_i^2 \lambda_i, \end{aligned} \quad (2.12)$$

em que λ_i é o autovalor associado ao autovetor \vec{v}_i . Sem perda de generalidade, assumimos que os autovalores estão indexados em ordem crescente, isto é,

$$\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_n. \quad (2.13)$$

Observamos da Equação 2.12 que a tarefa de minimizar R é equivalente a escolher quantidades para a_i , tal que $\sum_{i=1}^n a_i^2 \lambda_i$ seja mínimo.

Como todos os autovalores são não negativos, $\lambda_1 = 0$ e $a_i^2 \geq 0$ temos que $\sum_{i=1}^n a_i^2 \lambda_i \geq 0$. Por outro lado, $\sum_{i=1}^n a_i^2 \lambda_i = 0$ se $a_i = 0$ para $i = 2, 3, 4, \dots, n$, mesmo que $a_1 \neq 0$. Ou seja, o corte mínimo é 0 e ocorre quando o vetor de escolha é múltiplo de \vec{v}_1 , isto é, $\vec{s} = a_1 \vec{v}_1$.

Quando interpretamos esta solução, observamos que ela é trivial, pois, conforme mostramos no Teorema 1.12, $\vec{v}_1 = \vec{1}$ e se $\vec{s} = \vec{v}_1$ então todos os vértices estariam na mesma comunidade.

Para evitar essa solução trivial, uma opção é fixar o tamanho dos dois grupos. Se um grupo tem tamanho n_1 e o outro n_2 , então a_1 será fixado, pois

$$a_1^2 = (\vec{v}_1^T \vec{s})^2 = \frac{(n_1 - n_2)^2}{2}. \quad (2.14)$$

Com essa restrição em a_1 e a restrição $s^T s = n$ para o vetor de escolha, a minimização de R ocorre se \vec{s} é proporcional a \vec{v}_2 , o autovetor associado ao segundo menor autovalor. O autovetor \vec{v}_2 é chamado de vetor de Fiedler.

A escolha desse vetor, coloca todo o peso da quantidade a ser minimizada R , no segundo menor autovalor λ_2 , conhecido como conectividade algébrica, definido por Miroslav Fiedler (1973), (ROCHA, 2015).

A restrição imposta sobre \vec{s} na Equação 2.2, significa que, na maioria dos casos, \vec{s} não será paralelo a \vec{v}_2 , o que dificulta a minimização.

Logo, a forma de escolher uma solução para que \vec{s} se torne o mais próximo possível de ser paralelo a \vec{v}_2 é a que maximiza a quantidade

$$|\vec{v}_2^T \vec{s}| = \left| \sum_i v_{2_i} s_i \right|, \quad (2.15)$$

onde v_{2_i} é o i -ésimo elemento de \vec{v}_2 .

Observe, que através da desigualdade triangular, temos

$$\left| \sum_{i=1}^n v_{2_i} s_i \right| \leq \sum_i |v_{2_i}|. \quad (2.16)$$

Assim o vetor de escolha mais próximo possível de ser paralelo a \vec{v}_2 é \vec{s} tal que

$$s_i = \begin{cases} +1 & \text{se } \vec{v}_{2_i} \geq 0 \\ -1 & \text{se } \vec{v}_{2_i} < 0, \end{cases} \quad (2.17)$$

pois nesse caso

$$\left| \sum_{i=1}^n v_{2_i} s_i \right| = \sum_i |v_{2_i}|. \quad (2.18)$$

Contudo, observamos que essa escolha não satisfaz a condição de que o número de $+1$ e -1 elementos de s sejam iguais aos tamanhos desejados de n_1 e n_2 dos dois grupos. Mas essa não é uma condição que desejamos impor neste trabalho.

O método de particionamento espectral é capaz de trazer resultados bastante satisfatórios para o caso em que λ_2 que está distante dos demais autovalores $\lambda_3, \lambda_4, \dots, \lambda_n$.

2.4 CORTE MÍNIMO EM GRAFOS PONDERADOS

Em (NEWMAN, 2004), Newman afirma que em uma rede o fluxo máximo depende da capacidade de fluxo de cada aresta. Assim, detectar uma comunidade tal que o fluxo dentro da mesma seja máximo exige que consideremos além da conectividade a conectividade ponderada, em que as arestas recebem pesos não necessariamente iguais.

Portanto, quando a relação entre os vértices de um grafo tem diferentes intensidades, essas diferenças devem ser consideradas na detecção de comunidades. Assim, faz-se necessário um método de corte mínimo ponderado.

O corte para um grafo ponderado, com matriz de adjacência ponderada A_p e matriz laplaciana ponderada L_p , será denotado por R_p , e definido como:

$$R_p = \frac{1}{2} \sum_{\substack{i \in G_1 \\ j \in G_2}} A_{p_{ij}}. \quad (2.19)$$

Como na sessão anterior, o fator $\frac{1}{2}$ compensa a contagem duplicada das arestas que compõe o corte.

O vetor de escolha é definido como no corte mínimo, tornando a Equação 2.19

$$R_p = \frac{1}{4} \sum_{i,j} (1 - s_i s_j) A_{p_{ij}}. \quad (2.20)$$

A soma do peso de arestas conectadas ao vértice i , ou seja o grau ponderado do vértice i , k_i , pode ser representado como

$$k_i = \sum_j A_{p_{ij}}, \quad (2.21)$$

e usando a Equação 2.21 em um somatório duplo, temos

$$\sum_{i,j} A_{p_{ij}} = \sum_i k_i = \sum_i s_i^2 k_i = \sum_{i,j} s_i s_j k_i \delta_{ij}, \quad (2.22)$$

na qual $s_i^2 = 1$, já que $s_i = \pm 1$ e

$$\delta_i = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{caso contrário.} \end{cases} \quad (2.23)$$

Portanto,

$$R = \frac{1}{4} \sum_{i,j} s_i s_j (k_i \delta_{ij} - A_{p_{ij}}). \quad (2.24)$$

Observando que, $(k_i \delta_{i,j} - A_{p_{i,j}}) = L_{p_{i,j}}$, podemos escrever o corte R_p da seguinte forma

$$R_p = \frac{1}{4} \mathbf{s}^T L_p \mathbf{s}. \quad (2.25)$$

Veja que o corte mínimo ponderado é análogo ao corte mínimo, apenas substituindo a matriz laplaciana pela laplaciana ponderada. Como a matriz laplaciana ponderada também é real, simétrica, não negativa definida, podemos aplicar exatamente a ela a mesma análise espectral do corte mínimo, quanto a minimização do corte ponderado.

Assim o vetor escolha, \vec{s} , para o corte mínimo ponderado deve ser o mais próximo possível de ser paralelo ao autovetor, \vec{v}_2 , associado ao segundo menor autovalor de L_p . Considerando as restrição da Equação 2.2 temos que as entradas de \vec{s} serão definidas como

$$s_i = \begin{cases} +1 & \text{se } \vec{v}_{2_i} \geq 0 \\ -1 & \text{se } \vec{v}_{2_i} < 0. \end{cases} \quad (2.26)$$

3 EXPERIMENTO COMPUTACIONAL

Neste capítulo vamos aplicar o método apresentado no Capítulo 2. Embora tenhamos adotado uma caracterização matemática para o conceito de comunidade, ainda é difícil afirmar que em um experimento as comunidades foram detectadas corretamente. Diante disso, vamos aplicar o método de corte mínimo a um problema para o qual outro método de detecção de comunidades já foi aplicado e as comunidades detectadas foram aceitas como corretas, tanto por cientistas que estudam esse assunto quanto por conhecedores do grafo em questão.

3.1 COMUNIDADES DO LIVRO *STORM OF SWORDS*

Em (BEVERIDGE; SHAN, 2016) foi apresentada uma análise baseada em grafos do livro *A Storm of Swords*, o terceiro livro da coleção *A Song of Ice and Fire* e na qual está baseada a série televisiva *Game of Thrones*. Dentre outros aspectos, como centralidade e proximidade, são detectadas comunidades de personagens do livro. O grafo extraído do livro *A Storm of Swords*, tem como vértices os personagens principais, 107 ao todo, e as 352 arestas entre os vértices/personagens foram criadas se existia alguma interação entre os vértices. Ainda, as arestas são ponderadas conforme o número de interações dos vértices. O arquivo contendo as arestas acima está disponível em:

<https://github.com/melaniewalsh/sample-social-network-datasets/blob/master/sample-datasets/game-of-thrones/got-edges.csv>

O resultado obtido por (BEVERIDGE; SHAN, 2016) e reproduzido neste trabalho conforme o código no anexo B, apresentou 7 comunidades, conforme a Figura 5.

Mais precisamente, as comunidades detectadas são:

Comunidade 1={ Bran, Hodor, Jojen, Luwin, Meera, Rickon, Nan, Theon },

Comunidade 2 ={ Belwas, Daario, Daenerys, Irri, Jorah, Missandei, Rhaegar, Viserys, Barristan, Illyrio, Drogo, Aegon, Kraznys, Rakharo, Worm },

Comunidade 3 ={ Aemon, Alliser, Craster, Eddison, Gilly, Janos, Jon, Mance, Rattleshirt, Samwell, Val, Ygritte, Grenn, Karl, Bowen, Dalla, Orell, Qhorin, Styr },

Comunidade 4 ={ Arya, Beric, Eddard, Gendry, Sandor, Anguy, Thoros },

Comunidade 5 ={ Davos, Melisandre, Shireen, Stannis, Cressen, Salladhor },

Comunidade 6 ={ Aerys, Amory, Balon, Brienne, Bronn, Cersei, Gregor, Jaime, Joffrey, Jon Arryn, Kevan, Loras, Lysa, Meryn, Myrcella, Oberyn, Podrick, Renly, Robert, Robert Arryn, Sansa, Shae, Tommen, Tyrion, Tywin, Varys, Walton, Petyr, Elia, Ilyn, Pycelle, Qyburn, Margaery, Lancel, Olenna, Marillion, Ellaria, Mace, Chataya, Doran },

em duas comunidades. Foram aplicadas 7 vezes o método do corte mínimo. Como resultado obtivemos 8 comunidades, apresentadas na Figura 6.

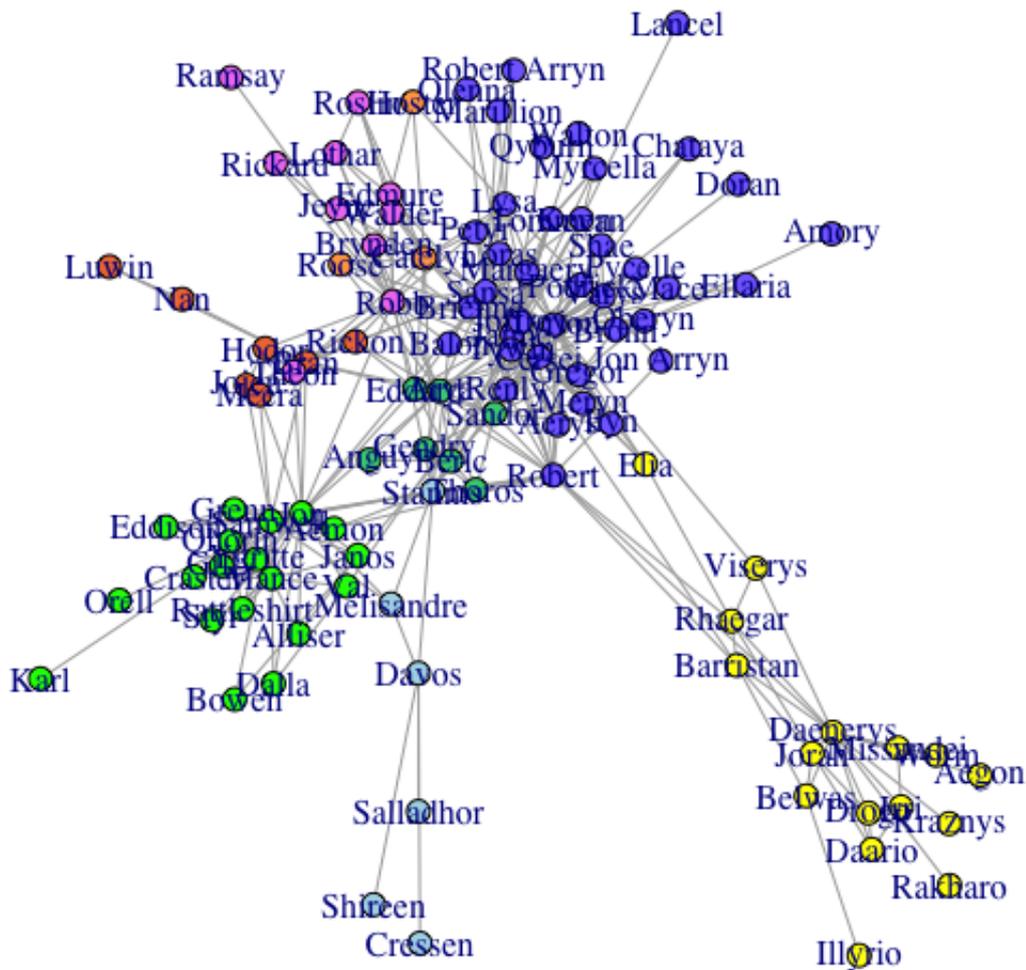


Figura 6 – Comunidades detectadas pelo corte mínimo

Mais precisamente, nossas comunidades são:

Comunidade A= { Bran, Hodor, Jojen, Luwin, Meera, Rickon, Nan },

Comunidade B= { Belwas, Daario, Daenerys, Irri, Jorah, Missandei, Rhaegar, Viserys, Barristan, Illyrio, Elia, Drogo, Aegon, Kraznys, Rakharo, Worm },

Comunidade C= { Aemon, Alliser, Craster, Eddison, Gilly, Janos, Jon, Mance, Rattleshirt, Samwell, Val, Ygritte, Grenn, Karl, Bowen, Dalla, Orell, Qhorin, Styr },

Comunidade D= { Arya, Beric, Eddard, Gendry, Sandor, Anguy, Thoros },

Comunidade E= { Davos, Melisandre, Shireen, Stannis, Cressen, Salladhor },

Comunidade F= { Aerys, Amory, Balon, Brienne, Bronn, Cersei, Gregor, Jaime, Joffrey, Jon Arryn, Kevan, Loras, Lysa, Meryn, Myrcella, Obery, Podrick, Renly, Robert, Robert Arryn, Sansa, Shae, Tommen, Tyrion, Tywin, Varys, Walton, Petyr, Ilyn, Pycelle, Qyburn, Margaery, Lancel, Olenna, Marillion, Ellaria, Mace, Chataya, Doran },

Comunidade G= { Brynden, Edmure, Lothar, Rickard, Robb, Walder, Theon, Jeyne, Roslin, Ramsay },

Comunidade H= { Catelyn, Hoster, Roose }.

Os resultados são bem parecidos, em particular as comunidades 3, 4 e 5 são exatamente iguais às comunidades C, D e E. Existem nas demais comunidades 3 diferenças principais, que são:

- Nos resultados de (BEVERIDGE; SHAN, 2016) o personagem Theon está na comunidade 1, cujo principal personagem é Bran, e nos resultados do corte mínimo Theon está na comunidade G, cujo principal personagem é Robb.
- Nos resultados de (BEVERIDGE; SHAN, 2016) a personagem Elia está na comunidade 6, cujo principal personagem é Tyrion, e nos resultados de corte mínimo Elia está na comunidade B, que é a comunidade cujo principal personagem é a Daenerys.
- A comunidade 7 dos resultados de (BEVERIDGE; SHAN, 2016) é, a menos do personagem Theon, a união das comunidades G e H dos resultados obtidos com o método do corte mínimo.

Essas diferenças, segundo nosso conhecimento sobre a estória, são plausíveis.

4 PROPOSTA DE APLICAÇÃO EM NÍVEL MÉDIO

Neste capítulo iremos subsidiar o professor de matemática, de Ensino Médio, ao trabalho com detecção de comunidades. Numa proposta que vai desde o estudo bibliográfico, por parte do professor, até a detecção de comunidades, orientada pelo professor e observada pelos alunos.

O objetivo dessa proposta é trazer uma ideia de trabalho com matrizes de forma diferenciada, atual e que desperte interesse, se não curiosidade, por parte dos estudantes. Logo, sugerimos que o estudo de matrizes venha acompanhado de conceitos novos, raramente abordados no Ensino Médio, mas que estão, no ramo da Álgebra Linear, intrinsecamente ligados aos conceitos matriciais e que são objetos de estudo desse conteúdo.

Outro forte argumento é trazer para dentro da sala de aula, um conteúdo recheado de aplicação reais, ou seja, conteúdos que promovam a resolução de situações problema do cotidiano e que tenham significado para o aluno. O que está de acordo com a nova BNCC (Base Nacional Comum Curricular), de 2018, (BNCC, 2018) que orienta ser necessário utilizar estratégias, conceitos e procedimentos matemáticos, em seus campos – Aritmética, Álgebra, Grandezas e Medidas, Geometria, Probabilidade e Estatística –, para interpretar, construir modelos e resolver problemas em diversos contextos, analisando a plausibilidade dos resultados e a adequação das soluções propostas, de modo a construir argumentação consistente.

Claramente conseguimos atingir essa competência, enunciada na BNCC, trazendo para o estudo de matrizes, a detecção de comunidades. O aluno não só será capaz de compreender uma aplicação real do estudo das matrizes, como estará apto a perceber que a matemática é ferramenta fundamental para a organização de redes, seja ela uma rede social, uma rede de negócios ou até mesmo uma situação cotidiana, como por exemplo, o mapa rodoviário de sua cidade, que está em rede.

Com o objetivo de dar suporte ao professor, desenvolvemos o roteiro que segue.

4.1 EMBASAMENTO TEÓRICO PARA O PROFESSOR

O primeiro passo para que o professor de matemática, se sinta seguro e preparado para inovar sua aula, é se aprofundar no assunto. Buscando referências bibliográficas que o auxiliem na compreensão geral do contexto, de detecção de comunidades, tanto no aspecto matemático como no histórico.

Portanto antes de iniciar o assunto em sala de aula, com seus alunos do Ensino Médio, sugerimos que o professor leia os capítulos anteriores dessa dissertação.

No Capítulo 1, podem ser feitos os estudos preliminares, para que os conceitos relacionados a Álgebra Linear, pertinentes a aplicação, sejam lembrados. Além disso, partimos do

pressuposto que nem todas as universidades, no curso de licenciatura em matemática, têm em seu currículo o estudo de Grafos.

Então, além dos tópicos de conceitos e teoremas descritos nessa dissertação o professor de matemática pode compreender o tema com maior profundidade consultando (LAWSON, 1997), um livro conceitual mais tradicional, que trás conceitos importantes. Com uma abordagem mais atual e aplicável, sugerimos (CASTRO A. M.; VIAMONTE, 2013), (MONTEIRO, 2010) e (ORTIZ; MOREIRA, 2016), o primeiro vem com um capítulo com uma breve introdução ao software MATLAB.

A parte histórica da detecção de comunidades irá auxiliar em sala de aula a sensibilizar os alunos para o tema. Fazer com que eles percebam que tópicos, como esse de redes, não surgiram apenas nos dias atuais. Isso pode favorecer no despertar da curiosidade e motivação dos estudantes.

Para ajudar o professor, nesse aspecto, envolvemos essa dissertação com fatos históricos que podem ser percebidos, principalmente no Capítulo 2. Para enriquecer a construção histórica do tema sugerimos a leitura de (NEWMAN, 2006), (RODRIGUES, 2009) e principalmente (EVES, 2007), esse é um livro de leitura agradável e que certamente servirá de base histórica não apenas para o tocante ao tema iniciado por Euler (1741), como também para outros temas pertinentes ao trabalho do professor em sala de aula.

Para entender como se construiu o conceito de rede, o Artigo (VERMELHO S. C.; BEERTONCELLO, 2015), aborda a construção desse conceito ao longo do tempo e é capaz de trazer exemplos de redes usuais.

Em uma das aulas propomos o uso do software RStudio, tutorial em (SANTANA, 2017), para conhecer seu funcionamento e aplicações, sugerimos a leitura do material disponível em <http://material.curso-r.com/rbase/>.

De domínio dos conceitos, agora o professor é capaz de desenvolver uma aula introdutória à detecção de comunidades.

4.2 PROPOSTA DE AULA SOBRE CONCEITOS FUNDAMENTAIS

É importante que o professor direcione o aluno à aprendizagem. Para que isso aconteça é fundamental introduzir definições e exemplos, que permitam que os estudantes compreendam os conceitos e sejam capazes de aplicá-los.

Para que o aluno chegue a compreensão dos objetivos de se detectar comunidades, deve partir dos conceitos chaves, da Álgebra Linear e da Teoria de Grafos. Claro que o nível de aprofundamento não deve ser universitário e sim, básico. Tão pouco é necessário fazer as demonstrações dos teoremas para os alunos.

Segue, uma proposta de roteiro para aulas introdutórias, sobre o detecção de comunidades em grafos. Salientamos que é importante que o aluno já tenha visto os conteúdos de matrizes e determinantes. A parte aqui sugerida é uma complementação desses conteúdos.

4.2.1 AULA SOBRE TEORIA DE GRAFOS

Conteúdo: Teoria de Grafos.

Objetivos: Conhecer a evolução histórica da Teoria de Grafos e compreender a aplicação dessa teoria em detecção de comunidades.

Ações: Inicie a aula entregando para cada aluno o xerox da Figura 7. Com base nessa figura, inicie contando o problema das sete pontes de Königsberg, veja Capítulo 2.

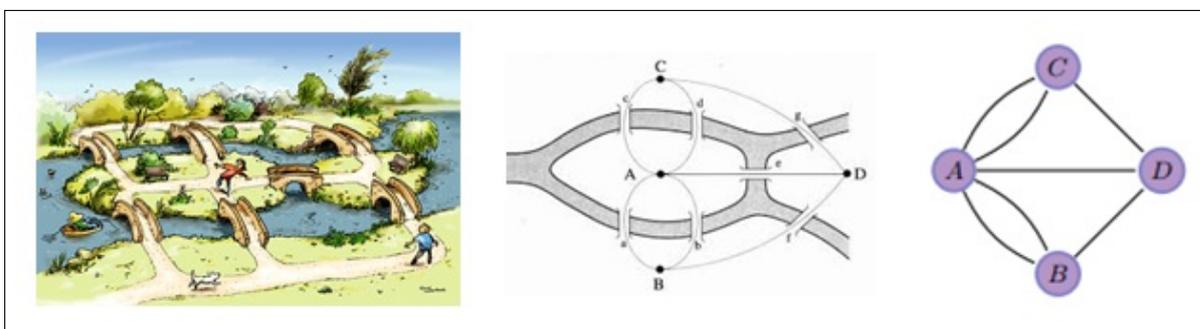


Figura 7 – Charge de Simon Kneebone sobre as pontes de Königsberg (à esquerda), representação gráfica do problema (centro) e grafo modelo gerado (à direita)

Em seguida, de preferência em projeção, defina grafo, aresta e vértice, utilizando, para isso, o modelo de grafo gerado pelo esquema das pontes de Königsberg.

Siga a aula comentando o experimento de Milgram, conhecido como "os seis passos de separação de Milgram", em seguida passe o vídeo "Seis graus de separação", do canal Nerdologia no Youtube, disponível em <<https://www.youtube.com/watch?v=YMI3CrChwSk>>. Use esse vídeo para reforçar a importância e utilidade do estudo de grafos.

Observação: sugerimos, para o professor, o documentário "Seis Graus de Separação (Documentário-2008)", também no Youtube, em

https://www.youtube.com/watch?v=BQ7UDWn_uws,

para que possa fazer observações pertinentes ao assunto com os alunos.

4.2.2 AULA SOBRE AUTOVALOR E AUTOVETOR

Conteúdo: Autovalor e autovetor.

Objetivos: Compreender os conceitos de autovetor e autovalor. Perceber que esses conceitos estão ligados aos conceitos matriciais.

Ações: Iniciar a aula, inserindo os conceitos de autovalor e autovetor, por meio de um exemplo, como o do no Capítulo 1, dessa dissertação.

Depois dessa introdução expositiva, os alunos reforçam os conceitos estudados, resolvendo uma lista de exercícios sobre o assunto.

Lista de exercícios sobre autovalor e autovetor	
1) Verificar, utilizando a definição, se os vetores dados são autovetores das correspondentes matrizes:	
a) $v = (-2, 1)^T$	b) $v = (1, 1, 2)^T$
$\begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 2 & 3 \end{pmatrix}$
2) Determinar os autovalores e os autovetores das seguintes matrizes:	
a) $\begin{pmatrix} 1 & 3 \\ -1 & 5 \end{pmatrix}$	b) $\begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}$
c) $\begin{pmatrix} 1 & -1 & 0 \\ 2 & 3 & 2 \\ 1 & 1 & 2 \end{pmatrix}$	d) $\begin{pmatrix} 3 & 3 & -2 \\ 0 & -1 & 0 \\ 8 & 6 & -5 \end{pmatrix}$

4.2.3 AULAS SOBRE ÁLGEBRA LINEAR APLICADA À TEORIA DE GRAFOS

Conteúdo: Álgebra linear associada à Teoria de Grafos (matriz adjacência, matriz diagonal de graus, matriz laplaciana e suas propriedades).

Objetivos: Apresentar para os alunos novos conceitos relacionados a Álgebra, no tocante matrizes, e que são aplicados na Teoria dos Grafos e conseqüentemente na detecção de comunidades.

Ações: Iniciar a aula expondo um exemplo de grafo usando os próprios alunos, como nos Exemplos 1 e 3, dessa dissertação. E a partir desses modelos introduza concomitantemente para grafos não direcionados e ponderados, os conceitos algébricos que estão abordados no Capítulo 1.

A medida que vai avançando na aplicação dos conceitos através de exemplos é importante que se faça o registro formal dos conceitos estudados. Pois, exercícios soltos, sem conceitos podem levar o aluno a resolver outros exercícios aplicando uma estratégia do tipo "siga o modelo", e dessa forma acaba não retendo os conteúdos matemáticos como deveria.

Assim que concluir essa parte, solicite aos alunos que se reúnam em equipes de 5 integrantes. Entregue para cada equipe o seguinte roteiro de atividade:

Atividade prática	
1º passo	Abram, individualmente, o WhatsApp e registrem, no caderno seus 10 primeiros contatos.
2º passo	Escolha outras duas redes sociais e faça o mesmo.
3º passo	Registre o nome de 10 colegas que estiveram com você durante o intervalo nas últimas semanas.
4º passo	Escreva o nome de 10 colegas que, em algum momento, realizaram uma tarefa, um exercício, ou trabalho, com você dentro de sala de aula.
5º passo	Escreva o nome de 10 colegas que estudaram com você em séries anteriores.
6º passo	Escreva o nome de 10 colegas que estiveram com você em alguma atividade fora da escola.
7º passo	Em equipe, mas cada um em seu caderno, monte um grafo, com o conjunto de vértices sendo $V =$ membro da equipe, e arestas $E =$ nome listado em algum dos passos de 1 a 6. Só deve contar como uma aresta válida se a ligação aparecer na lista dos dois vértices adjacentes.
8º passo	Repita o seu modelo de grafo e nele coloque a quantidade de vezes que o nome desses vértices aparecem ligados, consultando todas as listas da equipe. Formando um grafo ponderado.
9º passo	Compare os grafos de toda a equipe.
10º passo	Nos dois grafos obtidos, faça o estudo algébrico.

Observação: Essa descrição é para duas aulas, sugerimos que ainda na primeira aula, o professor tente iniciar a atividade em equipe, pelo menos até o 3º passo. Dessa forma poderá observar e intervir nas dúvidas que surgirem. E os alunos poderão ter algum tempo para pensarem sobre essas questões, assim aqueles que não conseguiram completar suas primeiras listas, podem fazê-lo. No final dessa atividade, o professor deve comentar com os alunos os resultados.

4.3 PROPOSTA DE ROTEIRO DE AULA SOBRE COMUNIDADES E REDES SOCIAIS

4.3.1 AULA SOBRE CONCEITO DE REDE E COMUNIDADE

Conteúdo: Conceitos e exemplos de redes sociais e detecção de comunidades

Objetivos: Explorar exemplos de redes e como foram detectados as comunidades em redes.

Ações: Inicie a aula instigando a curiosidade dos alunos com as imagens da Figura 8. Questione os alunos sobre o tipo de rede que eles percebem nas imagens. Se conhecem

outras redes como essas. Se são capazes de formular uma rede como as do exemplo. Nessa hora deseja-se que eles remetam as redes sociais propriamente ditas. E a partir disso, o professor, lança para eles a imagem da Figura 9 e com ela formaliza os conceitos de rede.

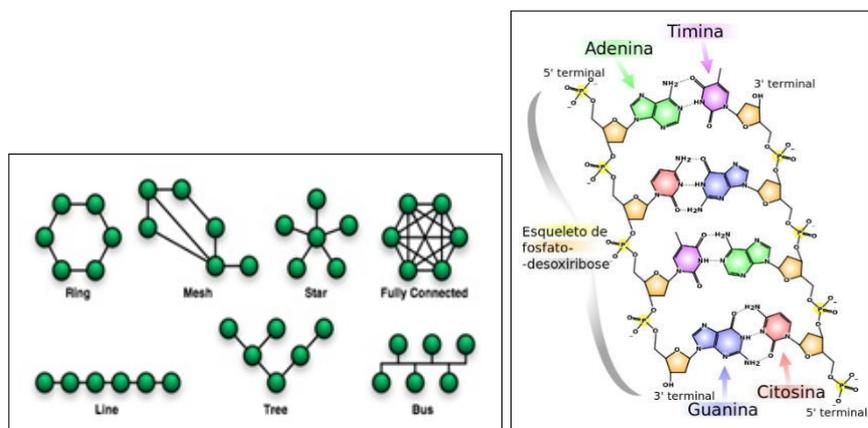


Figura 8 – Topologia de redes de comunicação (à esquerda) e bases nitrogenadas (à direita)

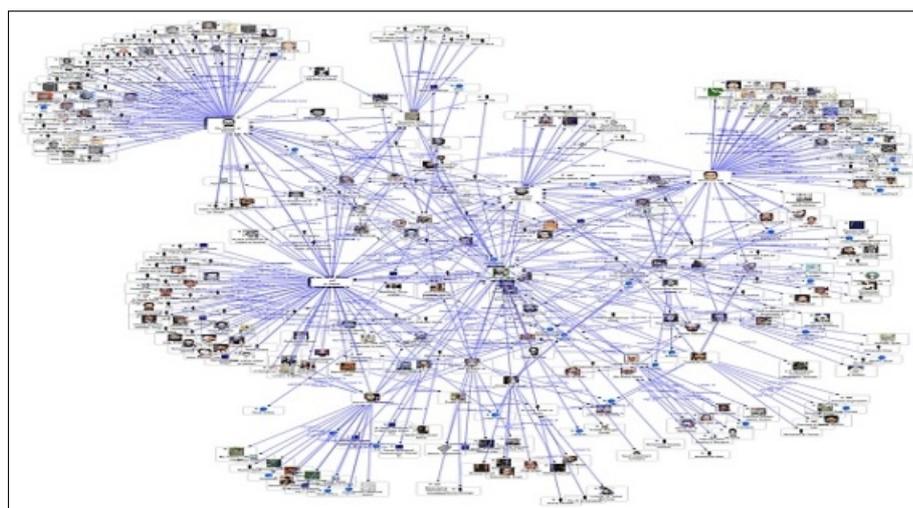


Figura 9 – Grafo de uma rede social

Em seguida, utilize o projetor ou cartelas com imagens impressas dos ícones de redes sociais, como nos exemplos da Figura 10, desafiando os alunos para ver qual reconhece mais redes pelas imagens.

Logo após essa atividade, o professor, deve fazer conexão, perguntando, como as comunidades se formam dentro de uma dessas redes. Como saber quais pessoas formam uma comunidade? Como estão interligadas essas pessoas? Que vantagens têm na detecção dessas comunidades? Com essas perguntas, e outras a critério do professor, deve ouvir as respostas dos alunos e ao mesmo tempo formalizar os conceitos de detecção de comunidades.

Na outra metade, dessa aula, o professor deve explorar o experimento de detecção de comunidades, exemplificado nessa dissertação, no Capítulo 3.

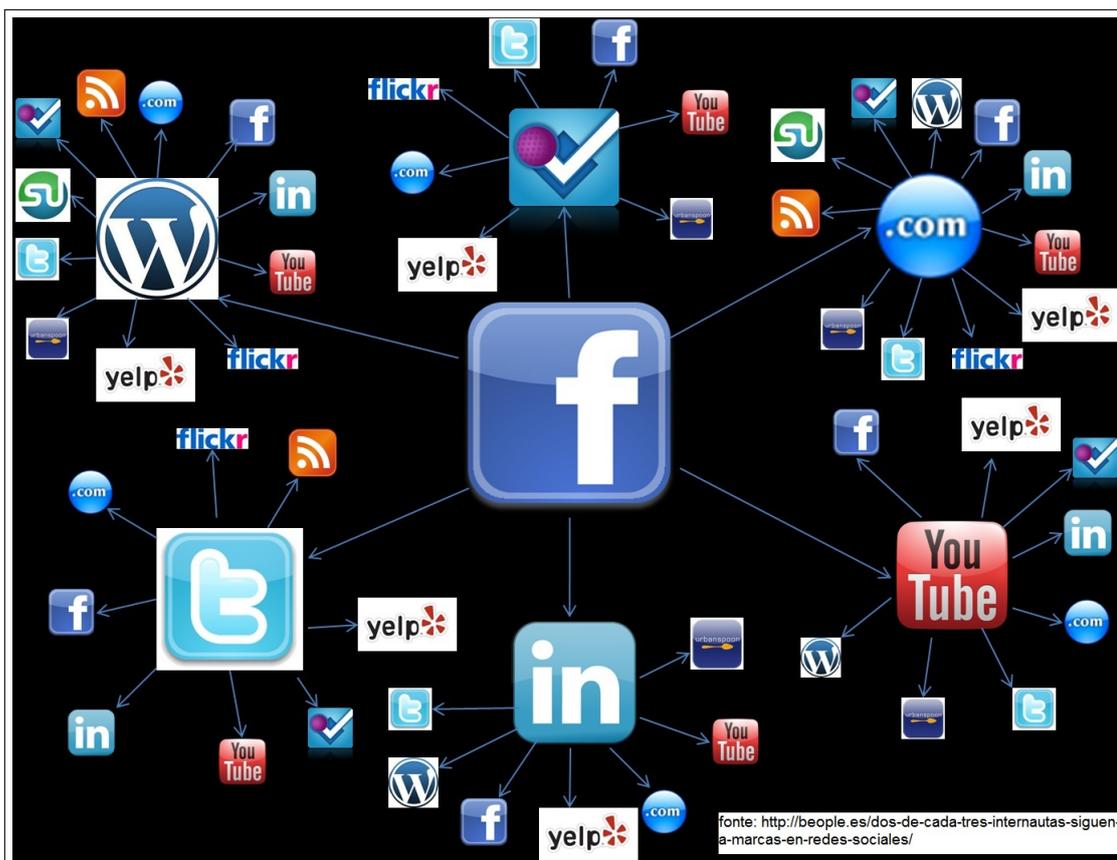


Figura 10 – Ícones de redes sociais

Observações: As fontes das imagens da Figura 8 são respectivamente, <https://pt.wikipedia.org/wiki/Topologia_de_rede> e <<https://mundoeducacao.bol.uol.com.br/biologia/a-formacao-acidos-nucleicos.htm>>. Já a fonte para a imagem da Figura 9, é <https://www.gta.ufrj.br/ensino/eel879/trabalhos_vf_2009_2/victor/Introduo.html>.

4.4 PROPOSTA DE ATIVIDADE PRÁTICA PARA A DETECÇÃO DE COMUNIDADES

4.4.1 AULA QUESTIONÁRIO PARA DETECÇÃO DE COMUNIDADES

Conteúdo: Análise de dados

Objetivos: realizar o levantamento de dados para a detecção de comunidades.

Ações: Antes dessa aula o professor deve montar um questionário num formulário online, com as perguntas:

1. Das disciplinas do Ensino Médio quais são as que você mais se identifica?
2. Quais atividades você costuma fazer durante suas horas vagas ou nos fins de semana?

3. Dos alunos de sua turma, quais os três que mais curtiram suas publicações nas redes sociais?
4. Dos alunos de sua turma, quais interagem mais com você dentro da escola?
5. Dos alunos de sua turma, quais interagem mais com você fora da escola?

Como alternativa para as questões sugerimos,

1. Colocar três caixas de seleção para as disciplinas curriculares, para ficar bem claro a questão de prioridade.
2. Permitir que só seja possível selecionar três opções, dentre elas, cinema, shopping, parques, esportes, estudar, video game, vídeos da internet, blogs, artesanato, filmes, séries, outros.
3. Colocar três caixas de seleção para o aluno selecionar por ordem os nomes dos alunos de sua turma. Inserir a opção nenhum.
4. Idem ao item 3.
5. Idem ao item 3.

O professor deve fornecer aos alunos o link para a realização do questionário. Precisa, baixar nos computadores da escola o software para geração de grafos, o qual, aqui, sugerimos o RStudio. Deve também deixar, em cada computador, salvo uma planilha com o nome dos alunos, escrito, cada um numa célula, da primeira linha e da primeira coluna, como demonstrado na Figura 11

	A	B	C	D	E	F	G	H
1		Amanda	Bruna	Carlos	Daniele	Eduardo	Fábio	Gabriela
2	Amanda	<input type="text"/>						
3	Bruna							
4	Carlos							
5	Daniele							
6	Eduardo							
7	Fábio							
8	Gabriela							

Figura 11 – Modelo para planilha de adjacência.

Na aula, em posse dos resultados fornecidos pelo formulário online, o professor deve encaminhar os alunos ao laboratório de informática e orientá-los para preencher a planilha de adjacência. Cada dupla receberá os dados de 6 alunos, referentes as questões 3, 4 e 5 do questionário. Outra dupla, receberá os mesmos nomes, a fim de que os erros de lançamento dos dados, nas planilhas, sejam minimizados.

Para preencher a planilha o aluno deve seguir seguir os critérios:

- (i) Se o colega foi mencionado em primeiro peso 3;
- (ii) Se o colega foi mencionado em segundo peso 2;
- (iii) Se o colega foi mencionado em terceiro peso 1 e;
- (iv) Se o colega não foi mencionado peso 0.

Outros dois grupos de duplas devem receber os dados referentes as questões 1 e 2 do questionário. Para o primeiro grupo considerar os critérios:

- (i) Se o colega marcou as mesmas disciplinas, na mesma ordem, o peso é 3;
- (ii) Se o colega marcou as duas primeiras opções iguais o peso é 2;
- (iii) Se o colega marcou a primeira opção igual o peso é 1 e;
- (iv) Se o colega, marcou mesmas disciplinas em ordem diferente ou não marcou disciplinas comuns o peso é 0.

O segundo grupo de duplas analisará a questão 2 do questionário, nos aspectos:

- (i) Se o colega selecionou três atividades em comum, peso 3;
- (ii) Se o colega selecionou duas atividades em comum, peso 2;
- (iii) Se o colega mencionou apenas uma atividade em comum, peso 1 e;
- (iv) Se o colega não mencionou nenhuma atividade comum, peso 0.

A fim de facilitar, o professor deve criar um email para essa atividade. Solicitar que os alunos enviem a planilha, para que ele possa compilar e agrupar os dados analisados pelos alunos em um único arquivo. Esse arquivo deve ser salvo com extensão *CSV* para ser manipulado no ambiente *R*.

4.4.2 AULA SOBRE CORTE MÍNIMO E DETECÇÃO DE COMUNIDADES NO R

Conteúdo: Corte mínimo e uso do RStudio

Objetivos: Utilizar o RStudio para detectar comunidades. Confrontar os resultados com o real. Conhecer o método do corte mínimo e a relação desse método com o autovalor e autovetor.

Ações: Antes de iniciar essa aula é importante que o professor já tenha alimentado o software R, em todas as máquinas do laboratório de informática, com os dados necessários para a detecção das comunidades na turma.

No início da aula levar a turma para o laboratório de informática e situar os alunos quanto ao ambiente R e como a implementação da detecção de comunidades acontece nele. É importante, que nesse momento, o professor explique aos alunos que o método de corte mínimo é apenas um algoritmo de detecção de comunidades e que existem outros, até mais eficientes que este, mas não tão simples.

Sugerimos que o professor explique que o corte mínimo se utiliza de um vetor de escolha para detectar comunidades e que ele está associado ao segundo menor autovalor e exatamente onde está esta informação nos códigos inseridos em R, bem como explique como isso acontece.

Para que o professor, consiga realizar a dinâmica de detectar comunidades usando o R, deve entender como foi implementado o experimento descrito no Capítulo 3. Compreender como ele se construiu, estudando os códigos de implementação que constam no Anexo ???. Cabe ao professor decidir se irá, ou não, realizar com alunos detecções sucessivas, porém, sugerimos que, o próprio professor, refine as comunidades para expor aos alunos e fazer, assim as devidas análises.

4.5 ESTUDO DE CASO

Nessa sessão vamos relatar a experiência de desenvolver essa proposta em sala de aula com os alunos do segundo ano do Ensino Médio.

A escola estadual onde a proposta foi aplicada está situada numa região de intenso comércio, tem grande procura pela comunidade e por alunos oriundos de escolas particulares. Dos trinta e seis alunos da turma sondada, apenas dois são novos, os demais estudam juntos desde o ano anterior. Todos os estudantes dessa turma possuem smartphones e em sua maioria com acesso a Internet, fato que tornou fácil o desenvolvimento das atividades aqui propostas.

A apresentação de grafos despertou curiosidade e interesse por parte dos alunos. Até mesmo os mais introspectivos contribuíram com ideias sobre o tema. Uma parte dos estudantes conseguiram vislumbrar aplicações futuras usando grafos, um deles sinalizou interesse em estudar assuntos relacionados a grafos pós conclusão do Ensino Médio.

A utilização das redes sociais como meio investigativo para construir grafos e o uso da tecnologia para implementá-los envolveu os estudantes, que se identificaram com as atividades propostas e foram capazes de perceber que a matemática vai além dos conteúdos curriculares expostos nos bancos escolares.

As atividades relacionadas a conceitos, exemplos e exercícios sobre teoria de grafos e álgebra linear foi de fácil assimilação, pois os alunos já tinham intimidade com o estudo de matrizes. Porém o tópico relacionado a detecção de comunidades usando o corte mínimo ficou confuso para compreensão dos alunos e assim, precisou de alguns ajustes.

Ao invés de abordar o método corte mínimo em seus detalhes, foi mais eficiente conceituá-

lo a partir da matriz laplaciana e da escolha do autovetor associado ao segundo menor autovalor, com um exemplo de como e porque isso acontece. Mostrar os comandos no software fez o aluno compreender a aplicação dos conceitos de autovetor e autovalor para a implementação do corte mínimo para detectar comunidades.

Os alunos não demonstraram reter as definições acerca de corte mínimo em si, porém sabem que esse método se utiliza dos conceitos relacionados a matrizes, autovalores e autovetores.

A quantidade de aulas destinadas a disciplina de matemática no Ensino Médio não contribui para a inclusão de atividades desse tipo em todos os conteúdos, mas escolas da rede estadual trabalham com atividades de feira do conhecimento ou projetos de iniciação científica que são momentos oportunos para desenvolver temas como o abordado nessa dissertação.

5 CONCLUSÃO

Os conceitos atrelados a àlgebra linear e a teoria de grafos apresentadas nesse trabalho foram de suma importância para a ideia de aplicação do tema de detecção de comunidades em grafos. Serviu como base para o cálculo do corte mínimo e além disso, foi capaz de transpor a sala de aula, com conteúdos não usuais e que podem favorecer o professor do Ensino Médio a integrar à suas aulas aplicações reais e úteis, despertando no estudante mais entusiasmo em estudar matemática.

Nesta dissertação abordamos o método do corte mínimo para a detecção de comunidades em grafos, ponderados ou não. Além disso implementamos o método do corte mínimo e o usamos para detectar comunidades em um problema padrão. Comparativamente com outro resultado da literatura, nossos resultados foram bastante satisfatórios. Isso mostra a viabilidade do método e motiva a aplicação desse tema em aulas de nível médio.

A implementação se deu sobre o terceiro livro da coleção *A Song of Ice and Fire*, pelo fato de existir um trabalho científico realizado por (BEVERIDGE; SHAN, 2016) e publicado. A implementação do método corte mínimo para os outros dois livros da coleção, ficará como desejo de trabalhos futuros.

Na parte final do trabalho, apresentamos uma proposta de uma sequência de aulas, sobre detecção de comunidades em grafos para o nível médio. Essa proposta está suportada nas demais partes da dissertação.

Diante do exposto nesse trabalho, podemos concluir que as disciplinas do PROFMAT na UTFPR serviram de incentivo e motivação ao tema. Além disso foram capazes de enriquecer o meu conhecimento matemático, não apenas para a parte física dessa dissertação, mas também em relação a minha prática quanto professora de matemática.

REFERÊNCIAS

- ASSIS, J. S. M. **Grafos Eurelianos no Ensino Médio**. Rio de Janeiro: IMPA, 2016. 10
- BEVERIDGE, A.; SHAN, J. Network of thrones. **Math Horizons**, v. 23, n. 4, p. 18–22, 2016. 8, 11, 28, 29, 31, 43
- BLONDEL V. D., G. J.-L. L. R.; LEFEBVRE, E. Fast unfolding of communities in large networks. **Fast unfolding of communities in large networks**, n. 10, 2008. 29
- BNCC. **Base Nacional Comum Curricular: Ensino Médio**. Brasília: MEC/Secretaria de Educação Básica, 2018. 32
- CASTRO A. M., S. A. V.; VIAMONTE, A. J. **Álgebra Matricial, conceitos, exercícios e aplicações**. Porto: Publindústria, 2013. 33
- CONTENT, M. D. R. **Social Media Trends 2018**. [S.l.]: <https://inteligencia.rockcontent.com/social-media-trends-2018/>, 2019. 22
- COSTA, R. V. **Grafos: Algumas aplicações a nível médio**. Dissertação (PROFMAT) — Universidade de Brasília, 2017. 10
- DIESTEL, R. **Graph Theory**. Hanburg: Springer, 2005. 12, 18, 22
- EULER, L. Solutio problematis ad geometriam situs pertinentis. **Math Horizons**, v. 8, n. 53, p. 128–140, 1741. 10
- EVES, H. **Introdução à História da Matemática**. São Paulo: Unicamp, 2007. 33
- FEOLILOFF P., K. Y.; WAKABAYASHI, Y. **Introdução à teoria de grafos**. São Paulo: IME-USP, 2012. 12
- LAWSON, T. **Álgebra Linear**. São Paulo: Blucher, 1997. 33
- LÉVI-STRAUSS, C. **Mito e significado**. Lisboa: Edições 70, 1987. 22
- MAIA L. F., N. H. B. O. C. F. Detecção de comunidades numa rede de relacionamento de alunos. **Revista Eletrônica Sistemas e Gestão**, n. 9, p. 480–487, 2014. 22
- MALTA, G. H. S. **Grafos no Ensino Médio: uma inserção possível**. Dissertação (PROFMAT) — Unversidade do Rio Grande do Sul, 2008. 10
- MILGRAN, S.; TRAVERS, J. An experimental study of the small world problem. **Sociometry**, v. 32, n. 4, p. 425–443, 1969. 21
- MONTEIRO, L. H. A. **Sistemas Dinâmicos Complexos**. São Paulo: Livraria da Física, 2010. 33
- NETTO, P. O. B.; JURKIEWICZ, S. **Grafos: Introdução e prática**. São Paulo: Blucher, 2009. 12, 13
- NEWMAN, M. Analysis of weighted networks. **arXiv:cond-mat/0407503v1**, v. 01, 2004. 26

- NEWMAN, M. Finding community structure in networks using eigenvectors of matrices. **arXiv:physics/0605087v3**, v. 03, 2006. 10, 20, 29, 33
- NOGUEIRA, D. K. **Introdução à Teoria de Grafos**: proposta para o ensino médio. Dissertação (PROFMAT) — Universidade de Brasília, 2015. 10
- ORTIZ, A. A. D.; MOREIRA, T. G. V. **Tópicos na Interseção entre a Teoria de Grafos e Álgebra**. Rio de Janeiro: SBM, 2016. 12, 33
- POOL, I. S.; KOCHEN, M. Contacts and influence. **Social Networks**, v. 1, n. 1, 1978. 21
- ROCHA, I. d. S. **Generalização da Teoria de Fiedler para a conectividade algébrica**. Dissertação (doutorado) — Universidade Federal do Rio Grande do Sul, 2015. 16, 18, 25
- RODRIGUES, D. M. **Detecção de Comunidades no sistema de correio eletrônico universitário**. Dissertação (Mestrado) — Instituto Superior da Ciências do Trabalho e da Empresa, 2009. 20, 21, 33
- SANTANA, V. **Tutorial R/RStudio**. FEA/USP, 2017. 11, 33
- VERMELHO S. C., V. A. P. M.; BEERTONCELLO, V. Sobre o conceito de redes sociais e seus pesquisadores. **Educ. Pesquisa**, v. 41, n. 4, p. 863–881, 2015. 33

ANEXOS

ANEXO A – CÓDIGO EM R PARA O MÉTODO DO CORTE MÍNIMO

```

library(igraph)# carrega o pacote igraph, que contem metodos uteis para grafos.
MyData<-read.csv(file="../../../stormofswords.csv", header = TRUE, sep=",")

from<-MyData[[1]]
to<-MyData[[2]]
weight<-MyData[[3]]
EdgeList <- data.frame(from, to)
Storm<- graph_from_data_frame( d= EdgeList, directed = FALSE)
E(Storm)$weight<-weight
V(Storm)$size<-5

n=gorder(Storm)
m=gsize(Storm)
L<-laplacian_matrix(Storm,weights = NULL)

ev<-eigen(L)
Evalues=ev[[1]]
Eectors<-ev[[2]]
s<-matrix(0,n,1)# vetor de escolha
#preenche o vetor de escolha conforme as entradas do
#autovetor associado ao segundo menor autovalor
for(i in 1:n){
  if(Eectors[i,n-1]>=0) s[i,1]=1 else s[i,1]=-1
}

# criar subgrafo para nova divisao
subX<-c()
Sub1<-c() #primeira comunidade
for(i in 1:n){
  if(s[i,1]==1) {
    b<-i
    subX<-c(subX,b)
  }
  else{
    b<-i
    Sub1<-c(Sub1,b)
  }
}

Com1<-V(Storm) [Sub1]
Storm2<-induced_subgraph(Storm,subX)
n=gorder(Storm2)

```

```

m=gsize(Storm2)

L<-laplacian_matrix(Storm2,weights = NULL)
ev<-eigen(L)
Evalues=ev[[1]]
Eectors<-ev[[2]]

s<-matrix(0,n,1)# vetor de escolha
for(i in 1:n){#preenche o vetor de escolha
  if(Eectors[i,n-1]>=0) s[i,1]=1 else s[i,1]=-1
}

# criar subgrafos para novas divisoes
subA<-c()
subB<-c()
for(i in 1:n){
  if(s[i,1]==-1) {
    b<-i
    subA<-c(subA,b)
  } else{
    b<-i
    subB<-c(subB,b)
  }
}

Storm3<-induced_subgraph(Storm2,subA)
n=gorder(Storm3)
m=gsize(Storm3)

L<-laplacian_matrix(Storm3,weights = NULL)
ev<-eigen(L)
Evalues=ev[[1]]
Eectors<-ev[[2]]
s<-matrix(0,n,1)# vetor de escolha
for(i in 1:n){#preenche o vetor de escolha
  if(Eectors[i,n-1]>=0) s[i,1]=1 else s[i,1]=-1
}

sub2<-c()
sub3<-c()
for(i in 1:n){
  if(s[i,1]==-1) {
    b<-i
    sub2<-c(sub2,b)
  } else{
    b<-i
    sub3<-c(sub3,b)
  }
}

```

```

Com2<-V(Storm3) [sub2]
Com3<-V(Storm3) [sub3]

# criar subgrafos para novas divisoes
Storm4<-induced_subgraph(Storm2,subB)
n=gorder(Storm4)
m=gsize(Storm4)

L<-laplacian_matrix(Storm4,weights = NULL)
ev<-eigen(L)
Evalues=ev[[1]]
Eectors<-ev[[2]]
s<-matrix(0,n,1)# vetor de escolha
for(i in 1:n){ #preenche o vetor de escolha
  if(Eectors[i,n-1]>=0) s[i,1]=1 else s[i,1]=-1
}

# cria subgrafo para nova divisao
subC<-c()
sub4<-c()
for(i in 1:n){
  if(s[i,1]==-1) {
    b<-i
    subC<-c(subC,b)
  }
  else{
    b<-i
    sub4<-c(sub4,b)
  }
}

Com4<-V(Storm4) [sub4]

Storm5<-induced_subgraph(Storm4,subC)
n=gorder(Storm5)
m=gsize(Storm5)

L<-laplacian_matrix(Storm5,weights = NULL)
ev<-eigen(L)
Evalues=ev[[1]]
Eectors<-ev[[2]]
s<-matrix(0,n,1)# vetor de escolha
for(i in 1:n){#preenche o vetor de escolha
  if(Eectors[i,n-1]>=0) s[i,1]=1 else s[i,1]=-1
}

# cria subgrafo para nova divisao

```

```

sub5<-c()
subD<-c()
for(i in 1:n){
  if(s[i,1]==-1) {
    b<-i
    sub5<-c(sub5,b)
  }
  else{
    b<-i
    subD<-c(subD,b)
  }
}

Com5<-V(Storm5)[sub5]
Storm6<-induced_subgraph(Storm5,subD)
n=gorder(Storm6)
m=gsize(Storm6)
L<-laplacian_matrix(Storm6,weights = NULL)
ev<-eigen(L)
Evalues=ev[[1]]
Eectors<-ev[[2]]
s<-matrix(0,n,1)# vetor de escolha
#preenche o vetor de escolha
for(i in 1:n){
  if(Eectors[i,n-1]>=0) s[i,1]=1 else s[i,1]=-1
}

# cria subgrafo para nova divisao
sub6<-c()
subE<-c()
for(i in 1:n){
  if(s[i,1]==1) {
    b<-i
    sub6<-c(sub6,b)
  }
  else{
    b<-i
    subE<-c(subE,b)
  }
}

Com6<-V(Storm6)[sub6]
Storm7<-induced_subgraph(Storm6,subE)
n=gorder(Storm7)
m=gsize(Storm7)
L<-laplacian_matrix(Storm7,weights = NULL)
ev<-eigen(L)

```

```

Evalues=ev[[1]]
Eectors<-ev[[2]]
s<-matrix(0,n,1)# vetor de escolha
for(i in 1:n){
  if(Eectors[i,n-1]>=0) s[i,1]=1 else s[i,1]=-1 #preenche o vetor de escolha.
}

sub7<-c()
sub8<-c()
for(i in 1:n){
  if(s[i,1]==1) {
    b<-i
    sub7<-c(sub7,b)
  }
  else{
    b<-i
    sub8<-c(sub8,b)
  }
}

Com7<-V(Storm7) [sub7]
Com8<-V(Storm7) [sub8]

#Atribuicao de cores as diferentes comunidades detectadas
for(i in 1:length(Com1))
{
  for(j in 1:107 )
  {
    if(Com1[i]$name==V(Storm) [j]$name)
    {
      V(Storm) [j]$color<-"yellow"
    }
  }
}

for(i in 1:length(Com2))
{
  for(j in 1:107 )
  {
    if(Com2[i]$name==V(Storm) [j]$name)
    {
      V(Storm) [j]$color<-"lightblue"
    }
  }
}
}

```

```
for(i in 1:length(Com3))
{
  for(j in 1:107 )
  {
    if(Com3[i]$name==V(Storm)[j]$name)
    {
      V(Storm)[j]$color<-"green"
    }
  }
}
```

```
for(i in 1:length(Com4))
{
  for(j in 1:107 )
  {
    if(Com4[i]$name==V(Storm)[j]$name)
    {
      V(Storm)[j]$color<-"violet"
    }
  }
}
```

```
for(i in 1:length(Com5))
{
  for(j in 1:107 )
  {
    if(Com5[i]$name==V(Storm)[j]$name)
    {
      V(Storm)[j]$color<-"slateblue1"
    }
  }
}
```

```
for(i in 1:length(Com6))
{
  for(j in 1:107 )
  {
    if(Com6[i]$name==V(Storm)[j]$name)
    {
      V(Storm)[j]$color<-"tan1"
      j=108
    }
  }
}
```

```
for(i in 1:length(Com7))
```

```
{
  for(j in 1:107 )
  {
    if(Com7[i]$name==V(Storm)[j]$name)
    {
      V(Storm)[j]$color<-"seagreen3"
      j=108
    }
  }
}

for(i in 1:length(Com8))
{
  for(j in 1:107 )
  {
    if(Com8[i]$name==V(Storm)[j]$name)
    {
      V(Storm)[j]$color<-"sienna2"
      j=108
    }
  }
}
plot(Storm, layout=coords)
```

ANEXO B – CÓDIGO EM R PARA DETECÇÃO DE COMUNIDADE USANDO A MODULARIDADE E O MÉTODO DE LOUVAIN

```
library(igraph)
MyData<-read.csv(file="../../../stormofswords.csv", header = TRUE, sep=",")

from<-MyData[[1]]
to<-MyData[[2]]
weight<-MyData[[3]]
EdgeList <- data.frame(from, to)
Storm<- graph_from_data_frame( d= EdgeList, directed = FALSE)
E(Storm)$weight<-weight
V(Storm)$size<-4
V(Storm)$label.cex = 0.8 #tamanho do nome de cada vertice

clusterlouvain <- cluster_louvain(Storm)

set.seed(17)
coords <-layout_fruchterman_reingold(Storm)
plot(Storm, vertex.color=rainbow(7,alpha=0.4)[clusterlouvain$membership],
layout=coords)
```