

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA APLICADA

Fundamentos Matemáticos de Estratégias Espectrais para Particionamento de Grafos

por

Guilherme Tadewald Varella

Dissertação submetida como requisito parcial
para a obtenção do grau de
Mestre em Matemática Aplicada

Prof. Dr. Carlos Hoppen
Orientador

Porto Alegre, fevereiro de 2020.

CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Varella, Guilherme Tadewald

Fundamentos Matemáticos de Estratégias Espectrais para Particionamento de Grafos / Guilherme Tadewald Varella.— Porto Alegre: PPGMAp da UFRGS, 2020.

106 p.: il.

Dissertação (mestrado) —Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Matemática Aplicada e Estatística, Porto Alegre, 2020.

Orientador: Hoppen, Carlos

Dissertação: Matemática Aplicada,
Teoria Espectral de Grafos

Fundamentos Matemáticos de Estratégias Espectrais para Particionamento de Grafos

por

Guilherme Tadewald Varella

Dissertação submetida ao Programa de Pós-Graduação em Matemática Aplicada do Instituto de Matemática e Estatística da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de

Mestre em Matemática Aplicada

Linha de Pesquisa: Teoria Espectral de Grafos

Orientador: Prof. Dr. Carlos Hoppen

Banca examinadora:

Prof. Dr. Claudio Rosito Jung
INF-UFRGS

Prof. Dr. Gabriel de Moraes Coutinho
DCC-UFMG

Prof. Dr. Luiz Emilio Allem
PPGMAp-UFRGS

Dissertação apresentada e aprovada em
6 de março de 2020.

Prof. Dr. Esequia Sauter
Coordenador

SUMÁRIO

LISTA DE FIGURAS	vi
LISTA DE TABELAS	vii
LISTA DE SÍMBOLOS	viii
RESUMO	x
ABSTRACT	xi
1 INTRODUÇÃO	1
1.1 Grafos de Similaridade	1
1.2 Funções de Particionamento	2
1.3 Problemas Relaxados	4
1.4 Desigualdade de Cheeger	6
1.5 Objetivos e Organização do Trabalho	7
2 PRELIMINARES	10
2.1 Teoria de Grafos	10
2.2 Álgebra Linear	14
2.3 Teoria Espectral dos Grafos	19
2.4 Resultados Técnicos	24
2.5 Modelo Probabilístico	26
3 PARTICIONAMENTO DE GRAFOS	32
3.1 Funções de Particionamento	32
3.2 Aproximação dos Problemas RatioCut e NCut	36
3.3 Métodos Espectrais	39

3.4	Particionamento de Árvores	45
4	GARANTIAS TEÓRICAS SOBRE O DESEMPENHO DE MÉTODOS ESPECTRAIS	52
4.1	Expansão e Condutância	52
4.2	Desigualdade de Cheeger de Ordem 2	53
4.2.1	Demonstração da Validade da Desigualdade de Cheeger	54
4.2.2	Algoritmo fornecido pela demonstração	60
4.3	Desigualdade de Cheeger de Ordem k	69
4.3.1	Demonstração para a Cota Inferior	70
4.3.2	Cota Superior - Funcionamento do Algoritmo	72
4.3.3	Demonstração para a Cota Superior	73
4.3.4	Detalhes sobre o Algoritmo	79
4.3.5	Demonstração dos Lemas Auxiliares	93
5	CONSIDERAÇÕES FINAIS	99
5.1	Trabalhos Futuros	100
	REFERÊNCIAS BIBLIOGRÁFICAS	102

LISTA DE FIGURAS

2.1	Exemplo de grafo	11
2.2	Exemplo de Grafo 3-regular	12
2.3	Exemplo de uma partição do conjunto de vértices do grafo . . .	13
2.4	Exemplo de Árvore	14
2.5	Exemplo de Grafo	20
3.1	Diferentes partições de V	34
3.2	Exemplo de grafo	42
3.3	Outras partições geradas pelo algoritmo k -médias	44
3.4	Exemplo de Árvore T^*	48
3.5	Remoção da aresta $\{u_i, v_i\}$, que está no caminho de V_{imp} a V_{par} , gerando os conjuntos de vértices A_i (em amarelo) e B_i (em verde). 49	49
4.1	Primeiro Exemplo de Grafo	62
4.2	Partições que podem ser a saída do algoritmo	64
4.3	Segundo Exemplo de Grafo	65
4.4	Diferentes Partições de P	67
4.5	Exemplo de grafo e sua respectiva partição gerada pelo algoritmo, que também gera a condutância de G	68
4.6	Exemplo de Grafo	83
4.7	Malha de Cubos	84
4.8	Conjuntos T_1, T_2, T_3, T_4 e T_5	85
4.9	Conjuntos X_1, X_2, Y_1 e Y_2	86
4.10	Segundo Exemplo	88
4.11	Malha do Segundo Exemplo	89
4.12	Partição que gera $\phi_G(3)$	91

LISTA DE TABELAS

4.1	Valores aproximados dos $\alpha_G(S_i)$	63
4.2	Autovetores de \mathcal{L}	66
4.3	Pesos das Arestas e Grau dos Vértices de G	68
4.4	Resultados dos $\alpha_G(S_i)$	69
4.5	Valor das Normas ao Quadrado dos $\mathbf{f}(v_i)$	85
4.6	Valor das Condutâncias	86
4.7	Condutância dos $S_{i,t}$	87
4.8	Projeções dos Vértices	88
4.9	Primeiro Teste das Condutâncias	92

LISTA DE SÍMBOLOS

$ S $	Número de vértices de um subconjunto S de vértices
d_i	Grau do vértice v_i
$\text{vol}(S)$	Volume de um subconjunto S de vértices
\bar{S}	Conjunto complementar de S
$W(A, B)$	Peso das arestas entre A e B
$\mathcal{P}_G(k)$	Conjunto das partições em k partes (ou coleção de k subconjuntos disjuntos de vértices) de G
$\phi_G(S_1, \dots, S_k)$	Condutância de S_1, \dots, S_k
$\phi_G(k)$	Condutância de G de ordem k
$G[S]$	Subgrafo induzido de S
$\langle \mathbf{u}, \mathbf{v} \rangle$	Produto interno entre os vetores
$\mathbf{u} \perp \mathbf{v}$	Vetores \mathbf{u} e \mathbf{v} perpendiculares
$\text{supp}(\mathbf{v})$	Suporte do vetor \mathbf{v}
$\text{Span}\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$	Espaço gerado pelos vetores $\mathbf{u}_1, \dots, \mathbf{u}_n$
$\text{tr}(M)$	Traço da matriz M
$(M)_{ij}$	Elemento da i -ésima linha e j -ésima coluna da matriz M
$\lambda_k^{(M)}$	k -ésimo menor autovalor da matriz M
$\det(M)$	Determinante da matriz M
$R_M(\mathbf{x})$	Quociente de Rayleigh da matriz M sobre o vetor \mathbf{x}
\mathcal{L}	Matriz laplaciana normalizada de um grafo
$P(A)$	Probabilidade do evento A
(Ω, P)	Espaço amostral Ω de probabilidade P
$p(x)$	Função densidade de probabilidade
$\mathbb{E}w$	Esperança da variável aleatória w

p_v	Peso do vértice v
$\psi_G(S)$	Expansão de S
$\psi_G(S_1, \dots, S_k)$	Expansão de S_1, \dots, S_k

RESUMO

O problema de particionamento consiste, basicamente, em agrupar dados semelhantes e separar aqueles que não se assemelham e pode ser modelado matematicamente a partir de grafos. Recentemente, foram publicados diversos resultados teóricos sobre o tema e alguns deles fornecem algoritmos que utilizam o espectro de uma matriz associada a um grafo para gerar uma partição que auxilia a demonstrar estes resultados. Tais resultados trazem argumentos teóricos que podem ser evidências de que este algoritmo teria um bom desempenho como método para encontrar a melhor partição do problema de particionamento. Neste trabalho, procuramos entender o funcionamento de alguns destes algoritmos. Especificamente, nosso objetivo é compreender os algoritmos obtidos pela demonstração da cota superior para condutâncias de ordem 2, apresentada por Chung [8], e ordem $k > 2$, demonstrada por Trevisan [33]. Com relação a este último, apesar da demonstração não exigir que alguns passos do algoritmo sejam implementados de uma única maneira, vamos especificá-los com base na análise dos exemplos em que o aplicamos. Nestes exemplos, observamos que os algoritmos definidos obtiveram um bom desempenho. Também estudamos o resultado apresentado por Wang et. al [35] sobre particionamento de árvores em duas classes e propomos uma possível estratégia para provar a versão estendida deste resultado para $k > 2$ classes. Parte desta estratégia envolve uma versão do resultado [35] estendido para grafos com peso nos vértices, cuja prova é uma contribuição original deste trabalho.

ABSTRACT

The clustering problem basically consists of grouping similar data and separating those that are not similar, and it can be using graphs. Recently, several theoretical results have been published on this subject, and some of them provide approximation algorithms that use the spectrum of a matrix associated with a graph to generate a partition to prove these results. In this work, we try to understand how some of these algorithms work. Specifically, our objective is to understand the algorithms obtained for proving an upper bound for partitions into two classes, presented by Chung [8], and into k -classes, presented by Trevisan [33]. Regarding the latter, since the proof does not specify how some steps of the algorithm must be implemented, we will define them based on the analysis of the examples in which we apply it. We observed that the algorithm performed well on these examples. We also studied the result presented by Wang et al. [35] about splitting trees into two classes and we propose a possible strategy to prove an extended version of this result for $k > 2$ classes. Part of this strategy involves an extension of a result in [35] extended to graphs with weights on vertices, whose proof is an original contribution of this work.

1 INTRODUÇÃO

Agrupar elementos semelhantes e separar aqueles que não se assemelham é um desafio nas mais diversas áreas, especialmente se há uma quantidade expressiva destes elementos. Uma das vantagens em se realizar este tipo de operação é poder tomar certas decisões para elementos de um mesmo grupo e decisões distintas para aqueles que estão em grupos diferentes. A problemas desta natureza damos o nome de *problemas de particionamento*.

1.1 Grafos de Similaridade

O problema de particionamento pode ser visto constantemente em áreas da computação como ciência de dados e mineração de dados, que foram popularizadas recentemente em razão dos avanços em inteligência artificial [2] e *machine learning* [38]. O estudo do problema de segmentação de imagem, que consiste em incluir numa mesma classe *pixels* cujas cores possuam tonalidades próximas e separar aquelas que são distantes, possui uma quantidade expressiva de resultados interessantes como os de Shi e Malik [32], Wang e Siskind [36], Liu, Zhao e Jiao [24] e Zeng et. al [39], entre muitos outros. Além destas, existem outras áreas onde o problema de particionamento é aplicado: no marketing poderão ser oferecidos os mesmos serviços para pessoas que possuem um histórico de compras semelhantes, na biologia animais e plantas são classificados conforme suas características e expressões genéticas, entre outros exemplos [14].

Saber qual é a melhor forma de classificar dados não é algo trivial e o problema pode ser formulado de diversas formas. Uma das maneiras de modelar esse tipo de problema, que será o foco deste trabalho, é utilizando um grafo de similaridade, onde cada vértice corresponde a um dado do problema original e as arestas

entre os vértices são ponderadas conforme o grau de semelhança entre os dados. O peso de cada aresta do grafo de similaridade pode ser obtido de diversas maneiras e há estudos que avaliam diferentes formas de definir estes pesos, dependendo algumas vezes do contexto em que está inserido. Conforme Shi e Malik [32], a utilização da distribuição Gaussiana como peso das arestas traz bons resultados para exemplos de segmentação de imagem. Bajusz [3] avalia o índice de Tanimoto, o índice de Dice, o coeficiente de Cosine e distância de Soergel como possíveis pesos para classificar impressões digitais moleculares. Já definidos os graus de similaridade, além de simplesmente construir o grafo onde o peso de cada aresta é o grau de similaridade entre os vértices, existem outros métodos de construção o grafo. Um exemplo é o método *k-nn* (*k*-nearest neighborhoods), proposto por Cover e Hart [9], que considera, para cada vértice, as arestas que ligam ele aos *k* vértices com maior grau de semelhança. Outro exemplo é o *r-neighborhood* que, dado um $r > 0$, desconsidera no grafo as arestas cujo grau de semelhança seja menor que *r*.

1.2 Funções de Particionamento

Transformando o problema para um grafo, o objetivo é encontrar a sua melhor partição, de tal forma que as arestas que incidem em vértices de classes diferentes possuam pesos mais baixos e as que incidem nos que estão na mesma classe tenham pesos elevados. Para modelar esse problema como um problema matemático bem posto, é preciso definir funções que medem a qualidade de uma partição para reescrevê-lo como um problema de otimização; a estas funções damos o nome de *funções de particionamento*. Seja um *grafo* $G = (V, E)$ ponderado com vértices v_1, \dots, v_n e arestas $\{v_i, v_j\}$ de peso $w_{ij} > 0$ cada e seja $S \subset V$ e $E(S) = \{\{v_i, v_j\} : v_i, v_j \in S\}$, as seguintes medidas que nos auxiliarão na definição de funções desse tipo: $|S|$ é o número de vértices de *S*; $d_i = \sum_{v_j \in V} w_{ij}$ é o *grau* do vértice v_i ; $W(G[S]) = \sum_{\{v_i, v_j\} \in E(S)} w_{ij}$ é o *peso das arestas internas de S*; $W(S, \bar{S}) =$

$\sum_{v_i \in S, v_j \in \bar{S}} w_{ij}$ é o *peso do corte induzido por S* ; e $\text{vol}(S) = \sum_{v_i \in S} d_i = 2 \cdot W(G[S]) + W(S, \bar{S})$ é o *volume* de S .

À procura de uma função de particionamento que possa ser considerada boa no contexto de segmentação de imagem, Wu e Leahy [37] definiram a função *corte*. Dado um grafo $G = (V, E)$ e uma partição S_1, \dots, S_k de V , a função corte é definida por $\text{Cut}(S_1, \dots, S_k) = \sum_{i=1}^k W(S_i, \bar{S}_i)$ para $k \geq 2$ fixo. O problema desta função é que ela favorece partições com $k - 1$ conjuntos pequenos e um conjunto que contenha todos os demais vértices. Por essa razão, Cox e Zhong [10] definiram uma função, chamada de *corte por razão*, que soma os cortes entre cada parte e seu complementar, normalizando-os pela quantidade de vértices que possuem, o que penaliza conjuntos pequenos. Esta função é dada por $\text{RatioCut}(S_1, \dots, S_k) = \sum_{i=1}^k \frac{\text{Cut}(S_i, \bar{S}_i)}{|S_i|}$. Porém, esta função não leva em consideração o peso das arestas internas de cada parte, que é uma característica que pode ser importante para definir a qualidade da partição. Para incorporar esse elemento, Shi e Malik [32] definiram o *corte normalizado* como $\text{RatioCut}(S_1, \dots, S_k) = \sum_{i=1}^k \frac{\text{Cut}(S_i, \bar{S}_i)}{\text{vol}(S_i)}$.

É possível notar que para cada uma dessas funções a quantidade de classes já é conhecida e, além disso, quanto menor o seu valor, melhor a qualidade da partição. Sabendo disso, podemos concluir que o problema de particionamento descrito inicialmente consiste, para grafos, em problemas de minimização, onde queremos encontrar a partição que retorna o menor valor possível para determinada função de particionamento. Em outras palavras, dado um grafo $G = (V, E)$, seja $F : \mathcal{P}_G(k) \rightarrow \mathbb{R}$ uma função de particionamento de G , onde $\mathcal{P}_G(k)$ é o conjunto de todas as partições de V em k partes. Queremos encontrar a partição $\{A_1, \dots, A_k\}$ que satisfaça

$$F(A_1, \dots, A_k) = \min_{S_1, \dots, S_k \in \mathcal{P}_G(k)} F(S_1, \dots, S_k).$$

Outra função de particionamento muito importante é a condutância. Dado um grafo $G = (V, E)$ e $k \geq 2$ subconjuntos $S_1, \dots, S_k \subset V$ disjuntos e não

vazios, a *condutância de* S_1, \dots, S_k é dada por

$$\phi_G(S_1, \dots, S_k) = \max_{i=1, \dots, k} \frac{\text{Cut}(S_i, \overline{S_i})}{\text{vol}(S_i)}.$$

Assim como os problemas Cut, RatioCut e NCut, o da condutância consiste em encontrar a solução do problema de minimização da função para um $k \geq 2$ fixo. A este valor mínimo de condutância damos o nome de *condutância de G de ordem k*, definida por

$$\phi_G(k) = \min_{S_1, \dots, S_k \text{ disjuntos}} \phi_G(S_1, \dots, S_k).$$

Conforme demonstrações apresentadas nos trabalhos de Wang e Siskind [36] e Shi e Malik [32], os problemas de otimização das funções RatioCut e NCut são NP-difíceis, ao contrário do problema Cut que é polinomial.

Vale ressaltar que, apesar de, neste trabalho, sempre tratarmos o número k de conjuntos na partição como um dado do problema, definir o número ideal de conjuntos é um problema fundamental de dados reais, que motiva ampla pesquisa.

1.3 Problemas Relaxados

Na Teoria Espectral de Grafos, os autovalores e autovetores de matrizes relacionados ao grafo, como por exemplo a de adjacências, a laplaciana e a laplaciana normalizada, poderão trazer resultados interessantes sobre um grafo. Alguns estudos se preocupam com a aplicação da Teoria Espectral em ciências específicas, como a química [17] e biologia [4]. Em outras situações, a preocupação com propriedades estruturais de um grafo é maior, como o resultado clássico de Teoria Espectral que diz que o número de componentes de um grafo é igual à multiplicidade do menor autovalor da matriz laplaciana do grafo, que é 0 [1]; o mesmo vale para a matriz laplaciana normalizada.

No caso de particionamento de grafos, a Teoria Espectral aparece em métodos que podem ser utilizados para tentar resolver um problema de particio-

namento. Os argumentos que medem a qualidade destes métodos são tipicamente empíricos e ainda não são perfeitamente compreendidos de um ponto vista rigoroso matemático.

Em alguns casos, descrever um problema matemático que se aproxime do problema de particionamento original pode sugerir um método para resolver este problema. Luxburg [34] apresenta demonstrações que provam que os problemas de minimização das funções RatioCut e NCut podem ser reescritos como problemas de otimização, onde a função objetivo é uma forma quadrática positiva semidefinida cujas variáveis satisfazem restrições lineares, mas possuem caráter discreto. Como estes também são problemas difíceis de serem resolvidos, uma estratégia muito utilizada para encontrar uma aproximação da solução do problema original é observar a resposta do mesmo problema de otimização, porém removendo o requisito de que as soluções estejam restritas a um conjunto discreto; dizemos que este é uma *relaxação* do problema original. Apesar da relaxação garantir cotas inferiores para o peso de uma partição ótima, sabe-se que a razão (*gap*) entre as soluções ótimas do problema original e do problema relaxado pode ser grande. Além disso, obter uma partição a partir da solução ótima do problema relaxado também é um desafio.

Aplicando o Teorema de Rayleigh-Ritz (que é assunto do Capítulo 2), é possível verificar que as relaxações dos problemas RatioCut e NCut têm como solução a matriz F cujas colunas são os k autovetores associados aos k menores autovalores das matrizes laplaciana, no problema RatioCut, e laplaciana normalizada, no problema NCut. Pela forma como foram definidas as restrições do problema original, é possível notar que há uma relação entre cada linha de F com um vértice do grafo G . Por essa razão, o que se faz é escolher algum método que decida como classificar as linhas de F em k conjuntos disjuntos não-vazios, o que implicaria numa classificação dos vértices de G em k classes de uma partição de V . Métodos como estes, que classificam os vértices de um grafo conforme informações obtidas a partir

do espectro de uma matriz relacionada a ele, serão chamados, neste trabalho, de *estratégias espectrais para particionamento de grafos*.

Existem estudos, como os de Ng, Jordan e Weiss [29], que propõem estratégias espectrais até então não conhecidas ou demonstram resultados com garantias sobre a qualidade do particionamento encontrado. Estudos como os de Dhillon, Guan e Kulis [13], Damle, Minden e Ying [6], Chakraborty e Das [11] e Fränti e Sieranoja [16], por exemplo, propõem alternativas de estratégias que utilizam um algoritmo de *clustering* popularmente utilizado e que será definido no Capítulo 3: o algoritmo das *k*-médias. É bom enfatizar que a avaliação das estratégias espectrais ocorre muitas vezes com argumentos de caráter empírico - nesse caso, podemos denominá-las de *heurísticas espectrais* -, mas também existem estratégias cujos argumentos que defendam o seu desempenho têm fundamentação teórica - nesse caso, serão chamadas de *algoritmos de aproximação espectrais*.

1.4 Desigualdade de Cheeger

Fiedler [15] apresenta resultados importantes em relação ao segundo menor autovalor da matriz laplaciana, chamado também de *conectividade algébrica*. Um exemplo destes resultados é que a conectividade algébrica é igual a zero se, e somente se, o grafo é desconexo. Além da conectividade algébrica, é possível verificar que este resultado vale também quando o segundo menor autovalor da matriz laplaciana normalizada também for igual a zero. Este resultado pode ser facilmente verificado observando a desigualdade clássica demonstrada por Cheeger [7] que relaciona a condutância de ordem 2 com o segundo menor autovalor da matriz laplaciana normalizada λ_2 . Para um grafo G , a *Desigualdade de Cheeger* de ordem 2 afirma que

$$\frac{\lambda_2}{2} \leq \phi_G(2) \leq \sqrt{2\lambda_2}.$$

Pela definição de condutância, podemos concluir com este resultado que, se λ_2 é pequeno o grafo necessariamente possui um corte esparso, enquanto que se λ_2 é grande o grafo não possui corte esparso.

Estudos recentes trouxeram versões desta desigualdade para ordem $k > 2$. De forma geral, temos que, para um grafo G , a Desigualdade de Cheeger de ordem $k > 2$ é dada por

$$\frac{\lambda_k}{2} \leq \phi_G(k) \leq C \cdot f(k) \cdot \sqrt{\lambda_k},$$

onde λ_k é o k -ésimo menor autovalor, $C > 0$ é uma constante absoluta e $f : \{2, \dots, n-1\} \rightarrow \mathbb{R}$ é uma função.

A função f mencionada na Desigualdade de Cheeger de ordem k aparece de várias formas na literatura. Lee, Gharan e Trevisan [22], por exemplo, mostram esse resultado para $f(k) = k^2$. Outros resultados são mais fracos, porém são menos complexos de provar, como é o caso das notas de aula de Trevisan [33], onde o resultado é demonstrado para $f(k) = k^{3.5}$. Normalmente, essas cotas superiores são demonstradas com provas construtivas, que fornecem um algoritmo espectral que retorna uma partição, nestes casos podendo então ser definido um algoritmo de aproximação espectral a partir da construção da prova.

Assim como a Desigualdade de Cheeger de ordem 2, a de ordem k pode ser utilizada para derivar resultados clássicos sobre a conectividade do grafo. Note, por exemplo, que se a multiplicidade de 0 como autovalor for pelo menos k , então o grafo possui pelo menos k componentes.

1.5 Objetivos e Organização do Trabalho

O objetivo deste trabalho é procurar entender alguns métodos utilizados para resolver o problema de particionamento para os quais existem argumentos teóricos que evidenciam o seu bom desempenho. Basicamente, iremos investigar dois

tipos de métodos que possuem estas características. O primeiro deles está ligado ao fato da partição ótima para uma determinada classe de grafos ter propriedades que permitem que haja um método que resolva o problema de particionamento em tempo polinomial. O segundo se refere a métodos que utilizam o espectro de uma matriz associado ao grafo do problema e que são obtidos a partir de uma demonstração de um resultado teórico rigoroso.

Sobre o primeiro tipo, estudamos o resultado apresentado por Wang et. al [35], que diz que a melhor partição em duas classes de uma árvore utilizando a função NCut garante que o grafo induzido de cada uma de suas classes é conexo. Eles afirmam que o resultado também vale para $k > 2$ classes, porém não demonstram esse fato e nem citam referências. Propomos neste trabalho uma possível estratégia de demonstração para este resultado utilizando indução. Para prosseguir com esta estratégia, é necessário utilizar uma versão do resultado de [35] estendido para árvores com pesos em seus vértices, cuja demonstração é uma contribuição original deste trabalho.

Sobre o segundo tipo, serão estudados os algoritmos espectrais de aproximação fornecidos pelas demonstrações das desigualdades de Cheeger de ordem 2 e $k > 2$. Como em alguns momentos a demonstração não é específica sobre os passos a serem realizados, propomos algumas decisões a serem tomadas no decorrer dos algoritmos e utilizamos alguns exemplos práticos para ilustrar que essas decisões podem ser boas.

O trabalho é dividido em quatro capítulos, além da introdução. No Capítulo 2, serão apresentadas definições e resultados preliminares das áreas de Álgebra Linear, Teoria de Grafos e Teoria Espectral de Grafos, além de outros resultados mais técnicos, que serão úteis para o restante do trabalho. No Capítulo 3, definimos as funções Cut, NCut e RatioCut e apresentamos alguns problemas relacionados a elas. No Capítulo 4, apresentamos as desigualdades de Cheeger de ordem 2 e $k > 2$ estudadas neste trabalho e os algoritmos obtidos através das demonstrações

destas desigualdades. No Capítulo 5, serão apresentadas as considerações finais e ideias para trabalhos futuros.

2 PRELIMINARES

Neste capítulo serão introduzidos alguns conceitos e resultados fundamentais e específicos das áreas de Teoria de Grafos, Álgebra Linear e Teoria Espectral que serão úteis para a compreensão e demonstração de resultados mais avançados apresentados nos capítulos seguintes. Por completude, incluímos todos os principais conceitos utilizados nesse trabalho, inclusive alguns que já foram definidos na introdução. Além disso, incluímos uma seção com resultados técnicos que serão úteis em demonstrações dos próximos capítulos. Por fim, definiremos um modelo probabilístico utilizado para justificar um passo de um algoritmo que será enunciado no Capítulo 4.

2.1 Teoria de Grafos

Nesta seção, veremos alguns conceitos básicos sobre grafos, encontrados em [5], que serão tratados no decorrer deste trabalho. Além disso, introduziremos algumas notações que serão utilizadas posteriormente.

Um *grafo* $G = (V, E)$ é um par de conjuntos finitos V e E , onde V é formado por elementos arbitrários chamados *vértices* e E é um subconjunto do conjunto de todos os possíveis pares não-ordenados de vértices distintos (se $\{u, v\} \in E$, então $u, v \in V$ e $u \neq v$). Aos pares $\{u, v\}$ damos o nome de *arestas*. Dizemos que um vértice u é *adjacente* a um vértice v se existe a aresta $\{u, v\} \in E$; além disso, a aresta $\{u, v\}$ é dita *incidente* em vértices u e v . Note que, conforme essa definição, neste trabalho serão desconsiderados os grafos que possuam arestas múltiplas, laços ou que sejam orientados.

Com o objetivo de obter uma melhor compreensão dessas estruturas, podemos visualizar os grafos como pontos, que representam os vértices, e ligações

entre pontos, que são as arestas. Quando for conveniente, podemos enumerar todos os vértices do grafo, como por exemplo definindo o conjunto de vértices como $V = \{v_1, \dots, v_n\}$. Vejamos um exemplo deste tipo de representação na Figura 2.1.

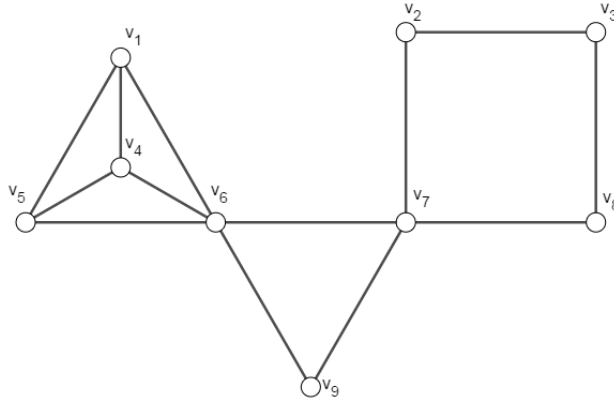


Figura 2.1: Exemplo de grafo

Um grafo $G = (V, E)$ pode ser visto como um *grafo ponderado* se existir uma função $w : E \rightarrow \mathbb{R}_{>0}$ que associa cada aresta a um valor não negativo. Denotaremos w_{ij} como o peso da aresta incidente nos vértices v_i e v_j e, por simplicidade, consideraremos $w_{ij} = 0$ caso v_i e v_j não sejam adjacentes. Quando quisermos tratar de um grafo não ponderado no contexto de grafos ponderados, consideraremos a função peso que atribui $w_{ij} = 1$, caso v_i e v_j sejam adjacentes, e $w_{ij} = 0$, caso contrário.

Para dois conjuntos, não necessariamente disjuntos, $A, B \subset V$, o conjunto de arestas entre os vértices de A e os vértices de B é o conjunto

$$E(A, B) = \{\{v_i, v_j\} : v_i \in A \text{ e } v_j \in B\}.$$

É possível definir o peso das arestas entre A e B como

$$W(A, B) = \sum_{\{v_i, v_j\} \in E(A, B)} w_{ij}.$$

Outro conceito importante é o do *grau* de um vértice. Dado um grafo G de vértices v_1, \dots, v_n , definimos o grau do i -ésimo vértice como

$$d_i = \sum_{j \neq i} w_{ij}.$$

Note que, no caso em que o grafo não é ponderado, o grau do vértice é simplesmente a quantidade de arestas incidentes a ele. Conhecendo o conceito do grau de um vértice, é então possível definir um tipo especial de grafo que será utilizado em alguns resultados deste trabalho. Dado um grafo G e um inteiro não-negativo d , dizemos que G é d -regular se todos os seus vértices têm grau d . Na Figura 2.2 vemos um exemplo de um grafo 3-regular.

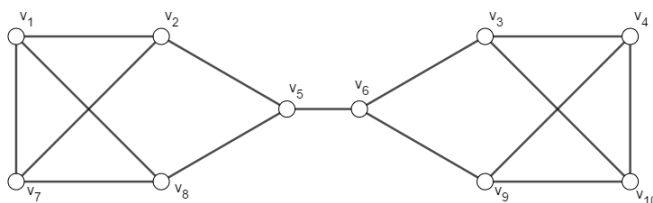


Figura 2.2: Exemplo de Grafo 3-regular

A seguir veremos duas definições que poderão ser aplicadas para o conjunto de vértices V de um grafo $G = (V, E)$ e possuem relação direta com os estudos realizados neste trabalho.

Definição 2.1. Dado um conjunto V , dizemos que $P = \{S_1, \dots, S_k\}$ é uma partição de V se, para quaisquer i e j tais que $i \neq j$, as seguintes afirmações são satisfeitas:

- (1) S_i é não vazio ($S_i \neq \emptyset$);
- (2) S_i e S_j são disjuntos entre si ($S_i \cap S_j = \emptyset$);
- (3) $\bigcup_{i=1}^k S_i = V$.

Neste caso, dizemos que S_1, \dots, S_k são as k partes (ou classes) de P .

Na Figura 2.3 temos um exemplo de partição do conjunto de vértices do grafo da Figura 2.1. Neste exemplo vemos uma partição em três partes $S_1 = \{v_1, v_3, v_5, v_8\}$ (vértices em verde), $S_2 = \{v_4, v_6, v_7\}$ (vértices em azul) e $S_3 = \{v_2, v_9\}$ (vértices em vermelho).

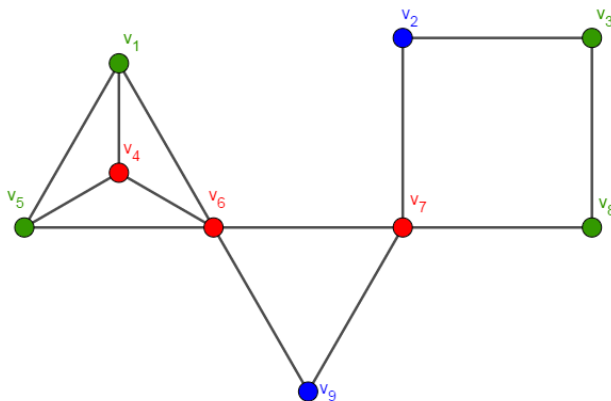


Figura 2.3: Exemplo de uma partição do conjunto de vértices do grafo

Na teoria de grafos, existem também definições que classificam certas estruturas de um grafo $G = (V, E)$. Um *passeio* de G é uma sequência de vértices intercalada com uma sequência de arestas, onde cada aresta é precedida e sucedida por seus vértices incidentes. Um *caminho* entre $u, v \in V$ é um passeio que não contem vértices e arestas repetidas e seu primeiro e último elemento são u e v . Um *ciclo* de G é um passeio sem vértices e arestas repetidas, exceto pelo primeiro e último vértice de sua sequência, que são exatamente os mesmos.

Um grafo $G = (V, E)$ é dito *conexo* se, para quaisquer $u, v \in V$ distintos, é possível obter um caminho entre u e v . Caso existam dois vértices que não satisfaçam essa condição, então G é dito *desconexo*. É possível verificar, por exemplo, que os grafos das figuras 2.1, 2.2 e 2.4 são conexos.

A partir de um grafo $G = (V, E)$ podemos gerar um *subgrafo* $G' = (V', E')$ tal que $V' \subset V$ e $E' \subset E$. Um subgrafo dito *subgrafo induzido* de G caso ele

contenha todas as arestas $\{u, v\} \in E$ tais que $u, v \in V'$. Denotaremos o subgrafo induzido G' por $G[V']$. Um subgrafo induzido $G' = (V', E')$ é dito *componente* de G se G' é conexo e se, para qualquer $v \in V \setminus V'$, o subgrafo induzido de $G'_v = (V'_v, E'_v)$, onde $V'_v = V' \cup \{v\}$, é desconexo.

Existem grafos que são classificados conforme alguma estrutura específica que ela possui. Uma *árvore*, por exemplo, é um grafo conexo que não possui nenhum ciclo e uma *floresta* é um grafo cujas componentes são todas árvores. Seja $G = (V, E)$ uma floresta, é possível verificar que a quantidade de arestas de G é igual a $|V| - k$, onde k é o número de componentes de G . Em particular, se G também é árvore, a quantidade de arestas é $|V| - 1$. Na Figura 2.4 vemos um exemplo de árvore.

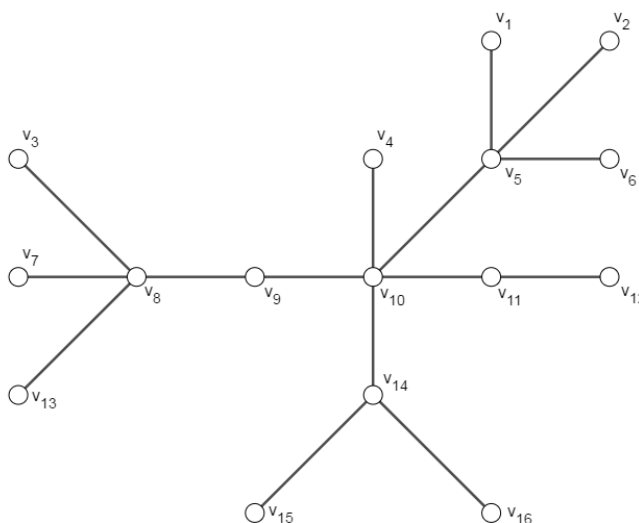


Figura 2.4: Exemplo de Árvore

2.2 Álgebra Linear

Nesta seção serão introduzidos alguns conceitos e resultados da área de Álgebra Linear. Aqueles de fácil compreensão e amplamente utilizados serão

enunciados no início desta seção e podem ser encontrados em livros como o de Lang [21]. No final desta seção, veremos alguns resultados mais avançados, aplicados para problemas mais específicos.

Antes de serem apresentados os conceitos e resultados, é importante apresentar algumas notações que serão utilizadas neste trabalho. A primeira delas é do produto interno usual entre dois vetores $\mathbf{u} = (u_1, \dots, u_n)$ e $\mathbf{v} = (v_1, \dots, v_n)$ pertencentes ao espaço vetorial \mathbb{R}^n , dado por $\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^n u_i v_i$. Também escrevemos $\mathbf{u} \perp \mathbf{v}$ quando os vetores são ortogonais entre si, ou seja, satisfazem $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. A notação para o *suporte* de um vetor \mathbf{v} , que é o conjunto dos índices i tais que $v_i \neq 0$, é dado por $\text{supp}(\mathbf{v})$. É importante notar que dois vetores $u, v \in \mathbb{R}^n$ de suporte disjunto são ortogonais entre si. O vetor $\mathbf{1} \in \mathbb{R}^n$ é aquele cujas coordenadas são todas iguais a 1. Também utilizamos a notação $\text{Span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ para o subespaço \mathbb{R}^n gerado pelos vetores $\mathbf{u}_1, \dots, \mathbf{u}_k \in \mathbb{R}^n$, ou seja, o conjunto de todos os vetores que podem ser escritos como uma combinação linear dos vetores $\mathbf{u}_1, \dots, \mathbf{u}_k$.

O conceito seguinte definirá o traço de uma matriz quadrada.

Definição 2.2. *Definimos o traço de uma matriz quadrada $M \in \mathbb{R}^{n \times n}$ como*

$$\text{tr}(M) = \sum_{i=1}^n (M)_{ii},$$

onde $(M)_{ii}$ é o i -ésimo elemento da diagonal principal de M .

Existem valores e vetores com propriedades especiais em relação a uma matriz quadrada. Tais valores e vetores serão definidos a seguir.

Definição 2.3. *Seja uma matriz quadrada $M \in \mathbb{R}^{n \times n}$, dizemos que $\lambda^{(M)} \in \mathbb{C}$ é autovalor de M se existe um vetor $\mathbf{v} \in \mathbb{C}^n$, $\mathbf{v} \neq \mathbf{0}$, tal que $M\mathbf{v} = \lambda^{(M)}\mathbf{v}$; neste caso, \mathbf{v} é dito autovetor associado a $\lambda^{(M)}$.*

Seja $M \in \mathbb{R}^{n \times n}$ e I a matriz identidade de ordem n . A partir do determinante $\det(x \cdot I - M)$, onde $x \in \mathbb{R}$ é indeterminado, é possível obter um

polinômio de grau n conhecido como *polinômio característico* da matriz M . O conjunto das raízes deste polinômio é exatamente o conjunto dos autovalores de M e a ele damos o nome de *espectro* de M . Neste trabalho, denotamos $\text{spect}(M)$ como o multiconjunto de autovalores de M , onde o número de vezes em que cada autovalor aparece é igual a sua multiplicidade como raiz do polinômio característico. O subespaço vetorial formado pelos autovetores de M associados ao autovalor $\lambda^{(M)}$ e o vetor nulo, denotado por $\mathbf{0}$, é chamado de autoespaço de M associado a $\lambda^{(M)}$.

Na seção de Teoria Espectral dos Grafos, veremos que um grafo pode ser associado a diversas matrizes e muitas delas têm a seguinte propriedade.

Definição 2.4. *Uma matriz quadrada $M \in \mathbb{R}^{n \times n}$ é simétrica se, para quaisquer i, j , temos $(M)_{ij} = (M)_{ji}$.*

O seguinte teorema traz resultados que caracterizam os autovalores e autovetores de uma matriz simétrica. Sua demonstração pode ser encontrada em [21, Theorem 3.2 pg. 214, Theorem 4.3 pg. 219].

Teorema 2.5. *Seja $M \in \mathbb{R}^{n \times n}$ simétrica. Então as seguintes afirmações são válidas:*

- (1) *M possui n autovalores reais $\lambda_1^{(M)}, \dots, \lambda_n^{(M)}$ (não necessariamente distintos);*
- (2) *Se \mathbf{u} e \mathbf{v} são autovetores associados a autovalores distintos de M , então $\langle \mathbf{u}, \mathbf{v} \rangle = 0$;*
- (3) *Existe um conjunto $\mathbf{u}_1, \dots, \mathbf{u}_n$, onde \mathbf{u}_i é autovetor correspondente ao autovalor $\lambda_i^{(M)}$ mencionado no item (1), que forma uma base ortonormal de \mathbb{R}^n .*

Com o item (1) do Teorema 2.5 conclui-se que é possível ordenar os autovalores de uma matriz simétrica M . Sendo assim, para M simétrica utilizaremos, no decorrer deste trabalho, $\mathbf{v}_1, \dots, \mathbf{v}_n$ para denotar os autovetores associados

aos autovalores $\lambda_1^{(M)} \leq \dots \leq \lambda_n^{(M)}$, respectivamente. Além disso, pelo item (3), é possível afirmar que a dimensão do autoespaço referente a um certo autovalor de M é exatamente igual à multiplicidade do respectivo autovalor.

A seguir serão enunciados alguns resultados mais avançados da área de Álgebra Linear. Para compreendê-los, antes definiremos a seguinte função que será constantemente utilizada nos resultados deste trabalho.

Definição 2.6. *O Quociente de Rayleigh associado a uma matriz $M \in \mathbb{R}^{n \times n}$ é a função $R_M : \mathbb{R}^n \setminus \{\mathbf{0}\} \rightarrow \mathbb{R}$ dada por*

$$R_M(\mathbf{y}) = \frac{\mathbf{y}^T M \mathbf{y}}{\mathbf{y}^T \mathbf{y}} = \frac{\langle \mathbf{y}, M \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle}.$$

É fácil verificar que, para qualquer $\alpha \in \mathbb{R}$ não nulo, vale $R_M(\alpha \cdot \mathbf{x}) = R_M(\mathbf{x})$.

Um dos resultados sobre o quociente de Rayleigh é o próximo teorema, que possui relação com o problema de otimização da função RatioCut, conforme será visto no capítulo 3. A prova de sua versão generalizada pode ser encontrada em [18, Theorem 4.2.2 pg. 234].

Teorema 2.7 (Teorema de Rayleigh). *Seja M uma matriz real e simétrica de ordem n com autovalores $\lambda_1^{(M)} \leq \dots \leq \lambda_n^{(M)}$ correspondentes aos autovetores $\mathbf{u}_1, \dots, \mathbf{u}_n$, respectivamente, que formam uma base ortogonal. Tomando $k \in \{1, \dots, n\}$ e definindo $\mathbf{W}_+ = \text{Span}\{\mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ e $\mathbf{W}_- = \text{Span}\{\mathbf{u}_1, \dots, \mathbf{u}_{k-1}\}$, temos as seguintes afirmações:*

(a)

$$\lambda_k^{(M)} = \max_{\mathbf{y} \neq \mathbf{0} \text{ e } \mathbf{y} \in \mathbf{W}_+^\perp} R_M(\mathbf{y}). \quad (2.1)$$

Além disso, se um vetor \mathbf{u} satisfaz a igualdade em (2.1) então \mathbf{u} é autovetor associado a $\lambda_k^{(M)}$;

(b)

$$\lambda_k^{(M)} = \min_{\mathbf{y} \neq 0 \text{ e } \mathbf{y} \in \mathbf{W}_\perp} R_M(\mathbf{y}). \quad (2.2)$$

Além disso, se um vetor \mathbf{u} satisfaz a igualdade em (2.2) então \mathbf{u} é autovetor associado a $\lambda_k^{(M)}$.

Será visto também no decorrer deste trabalho que o seguinte teorema possui relação com o problema de otimização do NCut. Sua demonstração pode ser encontrada em [28, Theorem 13 pg. 238].

Teorema 2.8 (Teorema de Rayleigh-Ritz). *Seja M uma matriz real e simétrica de ordem n com autovalores $\lambda_1^{(M)} \leq \dots \leq \lambda_n^{(M)}$ correspondentes a autovetores $\mathbf{u}_1, \dots, \mathbf{u}_n$ ortonormais, respectivamente. Para $k \leq n$, temos as seguintes afirmações:*

(a)

$$\min_{Y \in \mathbb{R}^{n \times k} \text{ e } Y^T Y = I} \text{tr}(Y^T M Y) = \lambda_1^{(M)} + \dots + \lambda_k^{(M)}.$$

A igualdade é atingida quando as colunas de Y geram $\text{Span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$;

(b)

$$\max_{Y \in \mathbb{R}^{n \times k} \text{ e } Y^T Y = I} \text{tr}(Y^T M Y) = \lambda_{n-k+1}^{(M)} + \dots + \lambda_n^{(M)}.$$

A igualdade é atingida quando as colunas de Y geram $\text{Span}\{\mathbf{u}_{n-k+1}, \dots, \mathbf{u}_n\}$;

Um teorema semelhante ao de Rayleigh é o de Courant-Fischer, que será utilizado para demonstrar um resultado importante mais adiante neste trabalho. Sua demonstração pode ser encontrada em [18, Theorem 4.2.6 pg. 236].

Teorema 2.9 (Courant-Fischer). *Seja M uma matriz real e simétrica de ordem n com autovalores $\lambda_1^{(M)} \leq \dots \leq \lambda_n^{(M)}$ correspondentes a autovetores $\mathbf{u}_1, \dots, \mathbf{u}_n$ ortonormais, respectivamente. Para $k \in \{1, \dots, n\}$, definimos \mathcal{W}_k o conjunto dos subespaços de \mathbb{R}^n de dimensão k . Então vale:*

(a)

$$\lambda_k = \min_{W \in \mathcal{W}_k} \max_{\mathbf{y} \in W \text{ e } \mathbf{y} \neq 0} R_M(\mathbf{y});$$

(b)

$$\lambda_k = \max_{W \in \mathcal{W}_{n-k+1}} \min_{\mathbf{y} \in W \text{ e } \mathbf{y} \neq 0} R_M(\mathbf{y}).$$

As duas igualdades são atingidas quando $\mathbf{y} = \mathbf{u}_k$.

2.3 Teoria Espectral dos Grafos

A Teoria Espectral de Grafos consiste em analisar propriedades estruturais dos grafos analisando o espectro de matrizes associadas a eles. Conforme [12], o estudo dessa área teve início em 1931 com o trabalho de Hückel em Química Quântica, porém ela começou a ser a vista de forma mais explícita como uma área da matemática em 1971 com a tese de doutorado Cvetković. Hoje em dia, a Teoria Espectral é muito estudada em diversas aplicações e o particionamento de grafos é uma delas. Nesta seção são definidas algumas matrizes associadas ao grafo, incluindo as matrizes laplaciana e laplaciana normalizada, que possuem ligação com o problema de particionamento de grafos e apresentadas algumas propriedades destas matrizes que são importantes para provarmos os resultados dos capítulos seguintes.

Primeiro, é importante conhecer outras duas matrizes relacionadas ao grafo para que possamos definir as matrizes laplaciana e laplaciana normalizada. Dado um grafo $G = (V, E)$ com vértices v_1, \dots, v_n e função peso $w : E \rightarrow \mathbb{R}_{>0}$, definimos sua *matriz de adjacências* A de tal forma que para cada i, j distintos temos $(A)_{ij} = w_{ij}$, onde w_{ij} é o peso da aresta $\{v_i, v_j\}$ caso v_i e v_j sejam adjacentes e $w_{ij} = 0$ caso contrário. Além disso, definimos sua *matriz de graus* D como uma matriz diagonal onde para cada i tem-se que $(D)_{ii}$ é igual ao grau do vértice v_i , denotado por d_i . Com a definição destas duas matrizes, será possível entender aquelas cujas definições virão a seguir.

Definição 2.10. Dado um grafo ponderado $G = (V, E)$ com vértices v_1, \dots, v_n e função peso $w : E \rightarrow \mathbb{R}_{>0}$, definimos sua matriz laplaciana $L = D - A$, ou seja, $(L)_{ij} = -w_{ij}$ se $i \neq j$ e $(L)_{ii} = d_i$.

A matriz laplaciana é simétrica e, por essa razão, se enquadra nos resultados enunciados na seção (2.2), como o fato de possuir somente autovalores reais e seus autovetores serem ortonormais.

Vejam os seguintes quais são as matrizes de adjacências, de graus e laplaciana do grafo representado na Figura 2.5.

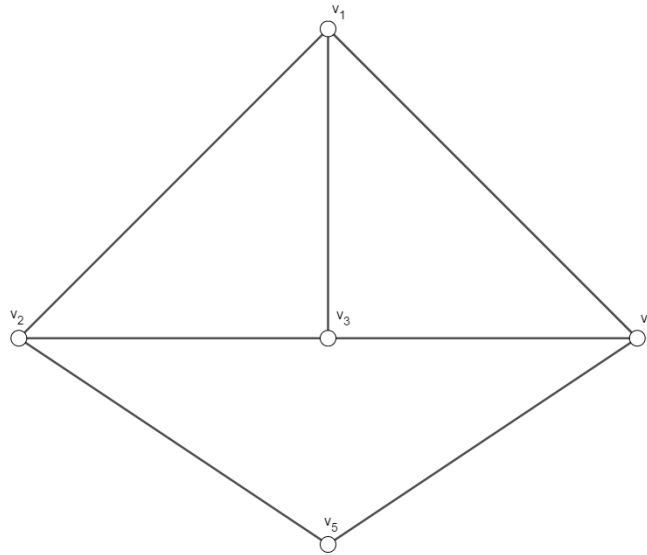


Figura 2.5: Exemplo de Grafo

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

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

A proposição seguinte traz uma propriedades da matriz laplaciana que será importante para um resultado obtido neste trabalho. Sua demonstração pode ser encontrada em [34, Proposition 1 pg. 4].

Proposição 2.11. *Seja a matriz laplaciana L de um grafo $G = (V, E)$ com vértices v_1, \dots, v_n e função peso $w : E \rightarrow \mathbb{R}_{>0}$. Para todo vetor $\mathbf{f} \in \mathbb{R}^n$ tem-se*

$$\mathbf{f}^T L \mathbf{f} = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2.$$

Apesar da matriz laplaciana estar envolvida em resultados importantes no particionamento de um grafo, a matriz relacionada ao grafo que será definida a seguir terá importância maior neste trabalho.

Definição 2.12. *Dado um grafo ponderado $G = (V, E)$ com vértices v_1, \dots, v_n não isolados e função peso $w : E \rightarrow \mathbb{R}_{>0}$, definimos sua matriz laplaciana normalizada de G como $\mathcal{L} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}$. Portanto, temos que os elementos de \mathcal{L} são*

$$(\mathcal{L})_{ij} = \begin{cases} 1 & \text{se } i = j \\ -\frac{1}{\sqrt{d_i d_j}} & \text{se } i \neq j \text{ e } \{v_i, v_j\} \in E \\ 0 & \text{se } i \neq j \text{ e } \{v_i, v_j\} \notin E \end{cases} .$$

Assim como a matriz laplaciana, a laplaciana normalizada também é simétrica e por isso se enquadra nos resultados enunciados na Seção 2.2. Note que a existência de $D^{-1/2}$ está condicionada ao grafo não possuir vértices isolados.

Na equação (2.3) temos a matriz laplaciana normalizada do grafo da Figura 2.5.

$$\mathcal{L} = \begin{bmatrix} \frac{\sqrt{3}}{3} & 0 & 0 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{3} & 0 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{3}}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{\sqrt{3}}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{\sqrt{2}}{2} \end{bmatrix} \cdot \begin{bmatrix} 3 & -1 & -1 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 \\ -1 & -1 & 3 & -1 & 0 \\ -1 & 0 & -1 & 3 & -1 \\ 0 & -1 & 0 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} \frac{\sqrt{3}}{3} & 0 & 0 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{3} & 0 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{3}}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{\sqrt{3}}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{\sqrt{2}}{2} \end{bmatrix}$$

$$\mathcal{L} = \begin{bmatrix} 1 & \frac{-1}{3} & \frac{-1}{3} & \frac{-1}{3} & 0 \\ \frac{-1}{3} & 1 & \frac{-1}{3} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-1}{3} & \frac{-1}{3} & 1 & \frac{-1}{3} & 0 \\ \frac{-1}{3} & 0 & \frac{-1}{3} & 1 & \frac{-1}{\sqrt{6}} \\ 0 & \frac{-1}{\sqrt{6}} & 0 & \frac{-1}{\sqrt{6}} & 1 \end{bmatrix}.$$

Um fato importante sobre as matrizes relacionadas ao grafo, como é o caso da laplaciana e da laplaciana normalizada, é que normalmente não há uma única maneira de representar um grafo com esses tipos de matrizes. No grafo da figura 2.5, por exemplo, poderíamos tomar a matriz laplaciana normalizada

$$\mathcal{L} = \begin{bmatrix} 1 & \frac{-1}{3} & 0 & \frac{-1}{3} & \frac{-1}{3} \\ \frac{-1}{3} & 1 & \frac{-1}{\sqrt{6}} & 0 & \frac{-1}{3} \\ 0 & \frac{-1}{\sqrt{6}} & 1 & \frac{-1}{\sqrt{6}} & 0 \\ \frac{-1}{3} & 0 & \frac{-1}{\sqrt{6}} & 1 & \frac{-1}{3} \\ \frac{-1}{3} & \frac{-1}{3} & 0 & \frac{-1}{3} & 1 \end{bmatrix}$$

se reordenássemos os vértices como $\{u_1, u_2, u_3, u_4, u_5\}$, onde $u_3 = v_5$, $u_5 = v_3$ e $u_i = v_i$ para os demais vértices. Como uma reordenação dos vértices produz uma matriz que pode ser obtida da matriz original através da permutação de linhas e das respectivas colunas, duas matrizes que diferem apenas pela ordenação dos vértices do grafo são matrizes similares e, por isso, possuem os mesmos autovalores. Lembre

que duas matrizes M e N de ordem n são similares se $M = PNP^{-1}$, onde P é uma matriz invertível de ordem n . No caso dos autovetores, um autovetor de uma das matrizes pode ser transformado em um autovetor da outra (associado ao mesmo autovalor) reordenando as coordenadas conforme a reordenação dos vértices. Isso mostra que é possível reordenar os vértices do grafo quando for conveniente, pois o espectro da matriz associada a ele não se altera.

Como nas matrizes associadas ao grafo vistas até este ponto as linhas e colunas são indexadas pelos vértices do grafo correspondente, é natural associar as coordenadas de um vetor de \mathbb{R}^n ao respectivo vértice no grafo. Outros vetores cujas coordenadas são indexadas conforme ordenação dos vértices serão apresentados na definição e proposição seguintes.

Definição 2.13. *Seja um grafo $G = (V, E)$ com vértices v_1, \dots, v_n e um subconjunto $S \subset V$. O vetor indicador $\mathbf{1}_S = (f_1, \dots, f_n)$ é tal que, para cada $i = 1, \dots, n$, temos*

$$f_i = \begin{cases} 1 & \text{se } v_i \in S \\ 0 & \text{se } v_i \notin S \end{cases} .$$

A proposição seguinte é um resultado que vale não só para matriz laplaciana normalizada, conforme enunciada, mas também para a laplaciana. Sua demonstração pode ser encontrada em [34, Proposition 4 pg. 6].

Proposição 2.14. *Seja um grafo $G = (V, E)$ e \mathcal{L} a sua matriz laplaciana normalizada. A multiplicidade k do autovalor 0 de \mathcal{L} é igual ao número de componentes S_1, \dots, S_k no grafo e, além disso, o autoespaço do autovalor 0 é gerado pelos vetores indicadores $D^{1/2}\mathbf{1}_{S_1}, \dots, D^{1/2}\mathbf{1}_{S_k}$ dessas componentes.*

Assim como a matriz laplaciana, a laplaciana normalizada também tem propriedades que serão importantes para demonstração de resultados no capítulos seguintes. O item (1) desta proposição é uma consequência da Proposição 2.14; o item (2) é uma consequência do resultado encontrado em [8, Equation 1.1 pg. 3] e; o item (3) é uma consequência do item (2) desta proposição e o Teorema 2.7.

Proposição 2.15. *Seja \mathcal{L} a matriz laplaciana normalizada de um grafo G com vértices v_1, \dots, v_n e $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(n)}$ os autovetores associados, respectivamente, aos seus autovalores $\lambda_1^{(\mathcal{L})} \leq \dots \leq \lambda_n^{(\mathcal{L})}$. Portanto as seguintes afirmações são verdadeiras:*

(1) $\mathbf{u}^{(1)} = D^{1/2}\mathbf{1}$ é autovetor associado ao autovalor $\lambda_1^{(\mathcal{L})} = 0$;

(2) *Seja o vetor $\mathbf{x} = (x_1, \dots, x_n)$ não nulo. Se G é d -regular, então*

$$R_{\mathcal{L}}(\mathbf{x}) = \frac{\sum_{\{v_i, v_j\} \in E} (x_i - x_j)^2}{d \sum_{j=1}^n x_j^2};$$

(3) *Para qualquer vetor $\mathbf{x} = (x_1, \dots, x_n)$ não nulo vale*

$$\lambda_2^{(\mathcal{L})} = \min_{\mathbf{x} \perp D\mathbf{1}} \frac{\sum_{\{v_i, v_j\} \in E} (x_i - x_j)^2}{\sum_{j=1}^n x_j^2 d_j}.$$

Além disso, a igualdade é atingida quando $\mathbf{x} = D^{-1/2}\mathbf{u}^{(2)}$.

2.4 Resultados Técnicos

Nesta seção veremos resultados mais técnicos que são utilizados em provas da Seção 2.5 e nos capítulos 3 e 4. Alguns deles são resultados bastante conhecidos e suas demonstrações podem ser encontradas em [21] e [23]. Outros são muito específicos para situações descritas neste trabalho que serão provados nesta seção.

Os três teoremas seguintes trazem três desigualdades bastante conhecidas: a de Bernoulli, a de Cauchy-Schwarz e a triangular.

Teorema 2.16 (Desigualdade de Bernoulli). *Sejam $d > -1$ e $k \geq 0$ inteiro, então vale*

$$(1 + d)^k \geq 1 + kd.$$

Teorema 2.17 (Desigualdade de Cauchy-Schwarz). *Para quaisquer vetores $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, temos que*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq |\mathbf{u}| \cdot |\mathbf{v}|.$$

De forma equivalente à desigualdade de Cauchy-Schwarz, também é possível afirmar que, dados os conjuntos $\{x_i\}_{i=1}^m$ e $\{y_i\}_{i=1}^m$, temos que

$$\left(\sum_{i=1}^m x_i y_i \right)^2 \leq \sum_{i=1}^m x_i^2 \sum_{i=1}^m y_i^2.$$

A desigualdade triangular é uma consequência da de Cauchy-Schwarz e será enunciada no corolário seguinte.

Corolário 2.18 (Desigualdade Triangular). *Para quaisquer vetores $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, temos que*

$$\left| \|\mathbf{u}\| - \|\mathbf{v}\| \right| \leq \|\mathbf{u} - \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

Outras duas desigualdades que serão importantes para provar alguns resultados deste trabalho, envolvendo médias, serão enunciados nos lemas seguintes.

Lema 2.19. *Seja o conjunto $\left\{ \frac{x_i}{y_i} \right\}_{i=1}^m$ com $x_i \geq 0, y_i > 0 \forall i \in \{1, \dots, m\}$, temos que*

$$\min \left\{ \frac{x_i}{y_i} \right\}_{i=1}^m \leq \frac{\sum_{i=1}^m x_i}{\sum_{i=1}^m y_i} \leq \max \left\{ \frac{x_i}{y_i} \right\}_{i=1}^m.$$

Lema 2.20. *Dadas funções $f, g : \mathbb{R} \rightarrow \mathbb{R}$, onde $g(x) > 0$ para qualquer $x \in \mathbb{R}$, e dado um número real $m > 0$, existe $0 < t \leq m$ tal que*

$$\frac{f(t)}{g(t)} \leq \frac{\int_0^m f(x) dx}{\int_0^m g(x) dx}.$$

Os seguintes lemas auxiliam na demonstrações de alguns resultados. Como são utilizados especificamente para desenvolver a prova dos resultados dos capítulos seguintes, iremos demonstrá-los nesta seção.

Lema 2.21. *Seja um grafo $G = (V, E)$ com vértices v_1, \dots, v_n e D sua matriz de graus. Seja um vetor $\mathbf{x} \perp D\mathbf{1}$, então*

$$\sum_{v_i} x_i^2 d_i = \min_{c \in \mathbb{R}} \sum_{v_i} (x_i - c)^2 d_i.$$

Demonstração. Se $\mathbf{x} \perp D\mathbf{1}$, então $\sum_{v_i} x_i d_i = 0$. Considere

$$\begin{aligned} \sum_{v_i} (x_i - c)^2 d_i &= \sum_{v_i} (x_i^2 - 2x_i c + c^2) d_i \\ &= \sum_{v_i} x_i^2 d_i + \sum_{v_i} c^2 d_i - 2c \sum_{v_i} x_i d_i = \sum_{v_i} x_i^2 d_i + \sum_{v_i} c^2 d_i. \end{aligned}$$

Portanto $\min_{c \in \mathbb{R}} \sum_{v_i} (x_i - c)^2 d_i$ é atingido quando $c = 0$. Logo concluímos que $\sum_{v_i} x_i^2 d_i = \min_{c \in \mathbb{R}} \sum_{v_i} (x_i - c)^2 d_i$. \square

Lema 2.22. *Seja $G = (V, E)$ um grafo cujas arestas $\{v_i, v_j\}$ tem peso w_{ij} e cada vértice v_i tem grau d_i . Dado um vetor $\mathbf{y} \in \mathbb{R}^n$ onde, para cada $i = 1, \dots, n$, a coordenada y_i está associada ao vértice v_i , podemos afirmar que*

$$\sum_{\{v_i, v_j\} \in E} w_{ij} (y_i + y_j)^2 \leq 2 \sum_{v_i \in V} d_i y_i^2.$$

Demonstração.

$$\begin{aligned} 2 \sum_{v_i \in V} d_i y_i^2 - \sum_{\{v_i, v_j\} \in E} w_{ij} (y_i + y_j)^2 &= 2 \sum_{v_i \in V} d_i y_i^2 - \sum_{\{v_i, v_j\} \in E} w_{ij} (y_i^2 + 2y_i y_j + y_j^2) \\ &= 2 \sum_{v_i \in V} d_i y_i^2 - \sum_{v_i \in V} d_i y_i^2 - \sum_{\{v_i, v_j\} \in E} w_{ij} (2y_i y_j) = \sum_{v_i \in V} d_i y_i^2 - \sum_{\{v_i, v_j\} \in E} w_{ij} (2y_i y_j) \\ &= \sum_{\{v_i, v_j\} \in E} w_{ij} (y_i^2 - 2y_i y_j + y_j^2) = \sum_{\{v_i, v_j\} \in E} w_{ij} (y_i - y_j)^2 \geq 0. \end{aligned}$$

Portanto, conclui-se que

$$\sum_{\{v_i, v_j\} \in E} w_{ij} (y_i + y_j)^2 \leq 2 \sum_{v_i \in V} d_i y_i^2.$$

\square

2.5 Modelo Probabilístico

Em uma das provas do capítulo 4 é utilizado um argumento probabilístico para demonstrar um resultado importante. Como isso é feito em um con-

texto específico, mas envolve um espaço de probabilidade contínuo com probabilidade uniforme, as definições necessárias serão apresentadas para uma classe particular de espaços, evitando menção a conceitos da Teoria da Medida. Tais definições serão dadas nesta seção e foram retiradas do livro de James [19], sofrendo alterações específicas para o nosso contexto. No final desta seção, descreveremos o espaço de probabilidade específico de que trataremos e os resultados que utilizaremos sobre ele.

Dado um experimento realizado sob certas condições fixas, chamamos de *espaço amostral*, denotado por Ω , o conjunto de todos os possíveis resultados para este experimento. O espaço amostral pode ser tanto de caráter discreto, quando Ω é finito ou enumerável, quanto contínuo, que neste modelo probabilístico consideraremos somente o caso em que Ω é uma região do espaço \mathbb{R}^k , para $k \geq 1$ inteiro. Um exemplo de espaço amostral discreto é o conjunto de todos os resultados possíveis de um arremesso de um dado comum ($\Omega = \{1, 2, 3, 4, 5, 6\}$); no caso do contínuo, a escolha de um número real entre 0 e 1 ($\Omega = (0, 1)$). Outro conceito importante é o de *evento* do espaço amostral, que se trata de um subconjunto $A \subset \Omega$.

Todo $A \subset \Omega$ está associado a um número real não negativo $P(A)$, chamado de *probabilidade* de A , que deve satisfazer as seguintes propriedades: (1) $P(\Omega) = 1$ e; (2) dados os eventos disjuntos $A_1, \dots, A_n \subset \Omega$, então $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$. Intuitivamente, a probabilidade associada a um evento indica com que frequência o evento A ocorre em relação ao espaço amostral. A essa altura podemos definir o modelo matemático (Ω, P) que representa um experimento de espaço amostral Ω e probabilidade P ; a este modelo damos o nome de *espaço de probabilidade*.

Uma *variável aleatória* de um espaço de probabilidade (Ω, P) é uma função $w : \Omega \rightarrow \mathbb{R}$ e, assim como o espaço amostral, também podemos classificá-la como contínua ou discreta. A variável aleatória *discreta* é aquela cujo contra-domínio \mathcal{W} é um conjunto finito ou enumerável, enquanto que na variável aleatória *contínua*

temos que \mathcal{W} é não-enumerável. Como exemplo, se $\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$, que é a região interna de um círculo unitário centrado na origem, podemos definir uma variável aleatória de diversas formas. Por exemplo, w é uma variável aleatória contínua se $w(x, y) = x + y$, pois $\mathcal{W} = [-\sqrt{2}, \sqrt{2}]$. Por outro lado, w definido por

$$w(x, y) = \begin{cases} \frac{x}{|x|} & \text{se } x \neq 0 \\ 0 & \text{se } x = 0 \end{cases}$$

é uma variável aleatória discreta, pois $\mathcal{W} = \{-1, 0, 1\}$. O modelo probabilístico que utilizaremos neste trabalho trata de variáveis aleatórias discretas definidas em um espaço amostral contínuo.

A função *densidade de probabilidade*, ou simplesmente *densidade*, definida em Ω é dada por $p(x) = P(\{x\})$, para qualquer $x \in \Omega$. Esta função satisfaz $\sum_{x \in \Omega} p(x) = P(\Omega) = 1$ quando Ω é discreto e $\int_{x \in \Omega} p(x) dx = 1$ quando é contínuo. Como trataremos de um espaço de probabilidade com distribuição uniforme, ou seja $p(x) = c \in \mathbb{R}$ para qualquer $x \in \Omega$, e de espaço amostral contínuo e compacto, podemos afirmar que $p(x) = \frac{1}{\text{volume de } \Omega}$ e que a probabilidade de um evento A é $P(A) = \frac{\text{volume de } A}{\text{volume de } \Omega}$. No caso onde $\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ e $A = \{(x, y) \in \Omega : |x| < \frac{1}{2}\}$, por exemplo, temos que

$$P(A) = \frac{\text{volume da região onde } |x| < 1/2}{\text{volume de } \Omega} = \frac{1}{3} + \frac{\sqrt{3}}{2\pi}.$$

Por simplicidade, denotamos, para um variável discreta w onde $\mathcal{W} = \{w_1, w_2, \dots\}$, a probabilidade $P(w_i) = P(\{x \in \Omega : w(x) = w_i\})$.

A seguir definiremos um valor que indica o valor médio de uma variável aleatória.

Definição 2.23. *Seja w a variável aleatória de um espaço de probabilidade (Ω, P) e função de densidade p em Ω , definimos a esperança de w como*

$$\mathbb{E}w = \int_{\Omega} w(x)p(x)dx \tag{2.3}$$

Em particular, se w é discreta de contra-domínio $\{w_1, w_2, \dots\}$, temos que (2.3) é igual a $\sum_i w_i P(w_i)$.

Uma propriedade importante da esperança é que ela é linear, ou seja, dadas duas variáveis aleatórias w, v do espaço de probabilidade (Ω, P) , então vale $\mathbb{E}(w + v) = \mathbb{E}w + \mathbb{E}v$. Por se tratar de um valor que indica o valor médio, alguns resultados conhecidos sobre o valor médio de um conjunto de valores também se aplica aqui. Um destes resultados é o teorema que virá a seguir.

Proposição 2.24. *Seja o espaço amostral de probabilidade Ω e w uma variável aleatória onde $\mathcal{W} = \{w_1, w_2, \dots\}$. Se para um certo $c \in \mathbb{R}$ vale $\mathbb{E}w \geq c$, então existe i tal que $w_i \geq c$.*

Agora iremos contextualizar a situação onde aplicaremos o modelo probabilístico definido. Dado um $k \geq 2$ inteiro, nos referimos a uma *malha de cubos* em \mathbb{R}^k a estrutura formada por infinitos cubos cujos lados têm mesmo tamanho e cada lado de um cubo também é lado do outro e é paralelo a um eixo de \mathbb{R}^k . Também definimos o núcleo de cada cubo da malha como cubos menores e de mesmos centros que os da malha. No contexto deste trabalho, os lados dos cubos são definidos por L , os núcleos possuem todos o mesmo lado e cada um tem distância igual a δ em relação a um núcleo que esteja em um cubo adjacente ao dele, o que implica que o lado dos núcleos são iguais a $L(1 - \delta/L)$.

Seja o espaço de probabilidade (Ω, P) de distribuição uniforme onde $\Omega = [0, L]^k$, associamos cada $(c_1, \dots, c_k) \in \Omega$, onde $c_i \in [0, L]$ para cada $i = 1, \dots, k$, à posição em que a malha se encontra no \mathbb{R}^k de tal forma que, dentre os vértices dos cubos das malhas cujas coordenadas são todas não-negativas, aquele que está mais próximo da origem é o ponto (c_1, \dots, c_k) . É possível verificar que todas as possíveis malhas de cubos de lado L estão sendo representadas por Ω e de forma única. Também definimos $C(x) = \bigcup_j C_j(x)$ a união das regiões dos núcleos da malha gerada pelo evento $x \in \Omega$.

Sejam os vetores $\mathbf{f}(v_1), \dots, \mathbf{f}(v_n)$ tais que $\sum_{i=1}^n \|\mathbf{f}(v_i)\|^2 = k$ e, para cada $i = 1, \dots, n$, $\bar{\mathbf{f}}(v_i) = \frac{\mathbf{f}(v_i)}{\|\mathbf{f}(v_i)\|}$ a projeção de $\mathbf{f}(v_i)$ na casca esférica unitária de \mathbb{R}^k . Usando o modelo probabilístico proposto, nosso objetivo é provar que existe uma posição da malha tal que a soma das normas ao quadrado de todos os vetores que estejam em algum núcleo da malha seja pelo menos $(1 - \frac{\delta k}{L}) \cdot k$. Em outras palavras, buscamos concluir que existe $x \in \Omega$ tal que $\sum_{\bar{\mathbf{f}}(v_i) \in C(x)} \|\mathbf{f}(v_i)\|^2 \geq (1 - \frac{\delta k}{L}) \cdot k$.

Para cumprir com esse objetivo, para cada $i = 1, \dots, n$ definimos as variáveis aleatórias m_i discretas de (Ω, P) como

$$m_i(x) = \begin{cases} \|\mathbf{f}(v_i)\|^2 & \text{se } \bar{\mathbf{f}}(v_i) \in C(x) \\ 0 & \text{se } \bar{\mathbf{f}}(v_i) \notin C(x) \end{cases}$$

e $M(x) = \sum_{i=1}^n m_i(x)$. Como (Ω, P) é de distribuição uniforme, a probabilidade de um ponto qualquer estar em $C(x)$ é igual a $\frac{\text{volume do núcleo}}{\text{volume do cubo}} = \frac{L^k(1-\delta/L)^k}{L^k} = (1 - \frac{\delta}{L})^k$.

Pela linearidade da esperança, temos que

$$\begin{aligned} \mathbb{E}M &= \mathbb{E}\left(\sum_{i=1}^n m_i(x)\right) \\ &= \sum_{i=1}^n (\mathbb{E}m_i(x)) \\ &= \sum_{i=1}^n (\|\mathbf{f}(v_i)\|^2 \cdot P(\bar{\mathbf{f}}(v_i) \in C(x)) + 0 \cdot P(\bar{\mathbf{f}}(v_i) \notin C(x))) \\ &= \sum_{i=1}^n \|\mathbf{f}(v_i)\|^2 \cdot \left(1 - \frac{\delta}{L}\right)^k = \left(1 - \frac{\delta}{L}\right)^k \cdot \sum_{i=1}^n \|\mathbf{f}(v_i)\|^2 \\ &= \left(1 - \frac{\delta}{L}\right)^k \cdot k \\ &\geq \left(1 - \frac{\delta k}{L}\right) \cdot k \end{aligned} \tag{2.4}$$

onde em (2.4) foi utilizada a Desigualdade de Bernoulli (Teorema 2.16). Portanto, pela Proposição 2.24 conclui-se que existe $x \in \Omega$ tal que $\sum_{\bar{\mathbf{f}}(v_i) \in C(x)} \|\mathbf{f}(v_i)\|^2 \geq (1 - \frac{\delta k}{L}) \cdot k$.

Uma possível estratégia para produzir um vetor (c_1, \dots, c_k) com essa propriedade é escolher vetores x_1, x_2, \dots sequencialmente até que o vetor escolhido

satisfaça a propriedade. A escolha deste vetor será útil, no Capítulo 4, para que o algoritmo possa iniciar a construção de uma partição a partir de um conjunto de vetores com massa significativa.

O fato de Ω ser de um espaço de probabilidade uniforme e a variável aleatória ser uma propriedade geométrica desse espaço garante que a probabilidade de que haja passos a serem realizados nesse processo seja pequena. Na linguagem de probabilidade, isso está relacionado ao fato da variável aleatória estar concentrada perto do seu valor esperado.

3 PARTICIONAMENTO DE GRAFOS

Em geral, o problema de particionamento procura classificar um conjunto de elementos em classes tais que elementos de uma mesma classe são similares e elementos de classes distintas são diferentes. No caso de grafos, isso consiste em encontrar um particionamento no grafo de tal forma que as arestas entre vértices de partes diferentes tenham pesos baixos e as arestas entre vértices dentro de uma mesma parte tenham pesos elevados. Isso exige que, na otimização, tenhamos que considerar dois objetivos diferentes: maximizar o peso total entre vértices de mesma parte e minimizar o peso total entre vértices de partes distintas.

Na primeira seção deste capítulo, iremos conhecer algumas funções que poderão ser relevantes para o seu particionamento. Na segunda seção, definimos, reformulamos os problemas de minimização RatioCut e NCut para problemas matemático bem definidos. Na terceira seção serão dados alguns exemplos de heurísticas espectrais que classificam os vértices observando estes autovetores. Na quarta e última seção, é apresentado um resultado que envolve o particionamento de árvores em duas partes utilizando NCut e sua demonstração.

3.1 Funções de Particionamento

Às funções mencionadas na introdução e que serão apresentadas nesta seção, damos o nome de *funções de particionamento*. Cada uma delas possui características singulares e a escolha de qual é a melhor opção entre elas vai depender da aplicação ou do tipo de partição desejada. Além disso, elas são definidas em termos de um grafo $G = (V, E)$ e de um valor inteiro $1 < k \leq |V|$. O domínio das funções definidas nesta seção é o conjunto de todas as partições de V em k partes e será denotado por $\mathcal{P}_G(k)$.

Antes de definir funções específicas, é importante definir o seguinte.

Definição 3.1. *Sejam G um grafo, $k \geq 2$ um número inteiro e F uma função de particionamento de G com domínio $\mathcal{P}_G(k)$. Definimos o problema F de G como aquele que tem como objetivo encontrar $P^* \in \mathcal{P}_G(k)$ tal que*

$$F(P^*) = \min_{P \in \mathcal{P}_G(k)} F(P).$$

Chamamos $F(P^)$ de valor ótimo e P^* de solução ótima do problema F .*

As medidas de volume ($\text{vol}(S)$) e peso do corte induzido ($W(A, B)$) definidas na introdução deste trabalho serão úteis para definir as funções de particionamento definidas nesta seção. A primeira função que veremos foi apresentada por Wu e Leahy [37] em segmentação de imagem.

Definição 3.2. *Seja um grafo ponderado $G = (V, E)$ de função peso $w : E \rightarrow \mathbb{R}_{>0}$, $1 < k \leq |V|$ inteiro e $\mathcal{P}_G(k)$ o conjunto das possíveis partições de V em k classes. Definimos a função $\text{Cut}_{G,k} : \mathcal{P}_G(k) \rightarrow \mathbb{R}$ como:*

$$\text{Cut}(S_1, \dots, S_k) = \frac{1}{2} \sum_{i=1}^k W(S_i, \bar{S}_i).$$

Dizemos que $\text{Cut}(S_1, \dots, S_k)$ é o valor do corte associado à partição $\{S_1, \dots, S_k\}$.

Para poder compreender o funcionamento desta e de outras funções de particionamento, vejamos um exemplo. Seja o grafo $G = (V, E)$ não ponderado de vértices v_1, \dots, v_{11} e suas partições $C = \{C_1, C_2\}$, $R = \{R_1, R_2\}$ e $N = \{N_1, N_2\}$, onde as partes de índice 1 são os vértices azuis e 2 são os vermelhos, conforme as figuras 3.1b, 3.1c e 3.1d.

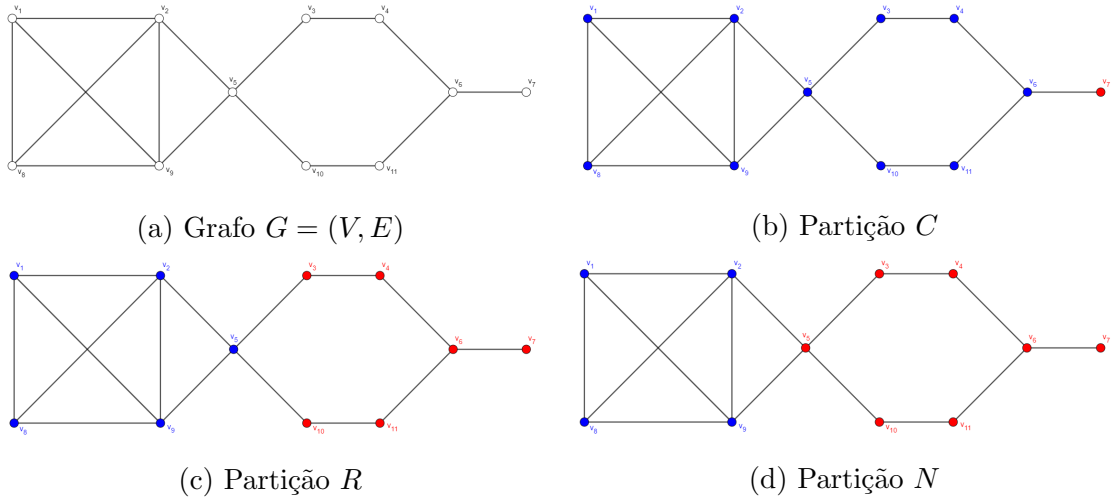


Figura 3.1: Diferentes partições de V

É possível verificar que $\text{Cut}(C) = 1$, $\text{Cut}(R) = 2$ e $\text{Cut}(N) = 2$, o que implica que C é a melhor das três partições quando utilizada a função de corte. Esse exemplo ilustra que a função Cut pode favorecer partições que tenham uma das classes com número pequeno de vértices, o que indica que a solução do seu problema de otimização poderá ter um certo desequilíbrio entre o número de vértices. Para contornar isso, Cox e Zhong [10] propuseram uma função que normaliza cada parcela pela cardinalidade da parte correspondente.

Definição 3.3. *Seja um grafo ponderado $G = (V, E)$ de função peso $w : E \rightarrow \mathbb{R}_{>0}$, $1 < k \leq |V|$ inteiro e $\mathcal{P}_G(k)$ o conjunto das possíveis partições de V em k classes. Definimos a função $\text{RatioCut}_{G,k} : \mathcal{P}_G(k) \rightarrow \mathbb{R}$ como:*

$$\text{RatioCut}(S_1, \dots, S_k) = \sum_{i=1}^k \frac{\text{Cut}(S_i, \bar{S}_i)}{|S_i|}.$$

A esta função damos o nome de corte por razão.

É possível verificar que $\text{RatioCut}(C) \approx 1.1$, $\text{RatioCut}(R) \approx 0.73$ e $\text{RatioCut}(N) \approx 0.78$, o que implica que R é a melhor das três partições quando utilizada a função de corte por razão. Observamos que as classes de R possuem um número equilibrado de vértices entre si, o que não ocorre em C .

Apesar de ser uma característica relevante para o problema de particionamento, a quantidade de arestas internas de uma mesma parte não interfere na função de corte por razão conforme a definição. Neste sentido, ao invés de normalizar a função de particionamento pelo tamanho das partes, podemos normalizar pelos seus respectivos volumes, conforme função apresentada por Shi e Malik [32].

Definição 3.4. *Seja um grafo ponderado $G = (V, E)$ de função peso $w : E \rightarrow \mathbb{R}_{>0}$, $1 < k \leq |V|$ inteiro e $\mathcal{P}_G(k)$ o conjunto das possíveis partições de V em k classes. Definimos a função $\text{NCut}_{G,k} : \mathcal{P}_G(k) \rightarrow \mathbb{R}$ como:*

$$\text{NCut}(S_1, \dots, S_k) = \sum_{i=1}^k \frac{\text{Cut}(S_i, \overline{S}_i)}{\text{vol}(S_i)}.$$

A esta função damos o nome de corte normalizado.

Pela forma como foi definida a função NCut , é possível observar que classes cujas arestas internas têm pesos elevados são favorecidas, pois, para cada $i = 1, \dots, k$, temos $\text{vol}(S_i) = 2 \cdot W([S_i]) + W(S_i, \overline{S}_i)$, conforme definição de volume na introdução deste trabalho.

Sobre as partições do grafo G da figura 3.1a, é possível verificar que $\text{RatioCut}(C) \approx 1.03$, $\text{RatioCut}(R) \approx 0.28$ e $\text{RatioCut}(N) \approx 0.27$, o que implica que N é a melhor das três partições quando utilizada a função normalizada. Notamos que as melhores partições das funções RatioCut e NCut se diferenciam apenas pelo vértice v_5 , que está na parte azul em R e vermelha em N . Isto se deve ao fato de v_1, v_2, v_8 e v_9 estarem todos conectados entre si e v_5 só estar conectado a dois destes quatro. Portanto, a função NCut favoreceu a partição N por esta conter uma classe com quantidade máxima de arestas internas, enquanto RatioCut se preocupou em favorecer partições cujas classes tenham uma quantidade equilibrada de vértices entre si.

3.2 Aproximação dos Problemas RatioCut e NCut

Das três funções de particionamento definidas na seção anterior, as duas que trazem mais informações em sua estrutura que podem ser relevantes para o problema de particionamento são as funções RatioCut e NCut. Sendo assim, queremos encontrar as partições que resolvam os problemas de minimização

$$\min_{\{S_1, \dots, S_k\} \in \mathcal{P}_G(k)} \text{RatioCut}(S_1, \dots, S_k) \quad (3.1)$$

e

$$\min_{\{S_1, \dots, S_k\} \in \mathcal{P}_G(k)} \text{NCut}(S_1, \dots, S_k). \quad (3.2)$$

Wang e Siskind [36] e Shi e Malik [32] apresentam demonstrações que provam que os problemas RatioCut e NCut, respectivamente, são NP-difíceis. O que se busca então são problemas cujas soluções tenham uma boa aproximação das soluções dos problemas originais e que possam ser resolvidos em tempo polinomial. Nesta seção reformularemos os problemas de minimização das funções RatioCut e NCut na forma de problemas lineares com restrições discretas e em seguida definiremos uma aproximação deste problema, conhecida como problema relaxado, onde consideramos o mesmo problema, só que com restrições contínuas. A reformulação dos problemas foi apresentada por Luxburg [34] em seu tutorial.

Primeiro, queremos encontrar uma reformulação do problema de minimização RatioCut, dado em (3.1). Para isso, seja $P = \{S_1, \dots, S_k\}$ uma partição de V em k classes. A matriz $F_P \in \mathbb{R}^{n \times k}$ tem como colunas os k vetores $\mathbf{f}^{(i)} = (f_1^{(i)}, \dots, f_n^{(i)})$ cujas coordenadas são

$$f_j^{(i)} = \begin{cases} \frac{1}{\sqrt{|S_i|}} & \text{se } v_j \in S_i \\ 0 & \text{se } v_j \notin S_i \end{cases}$$

para cada $j \in \{1, \dots, n\}$.

Utilizando a Proposição 2.11, pode-se observar que

$$\begin{aligned}
(\mathbf{f}^{(i)})^T L \mathbf{f}^{(i)} &= \frac{1}{2} \sum_{\ell, j=1}^n w_{\ell j} (f_{\ell}^{(i)} - f_j^{(i)})^2 \\
&= \frac{1}{2} \left(\sum_{v_{\ell} \in S_i, v_j \in \bar{S}_i} w_{\ell j} \frac{1}{|S_i|} + \sum_{v_{\ell} \in \bar{S}_i, v_j \in S_i} w_{\ell j} \frac{1}{|S_i|} \right) \\
&= \frac{1}{2} \cdot \frac{1}{|S_i|} \left(\sum_{v_{\ell} \in S_i, v_j \in \bar{S}_i} w_{\ell j} + \sum_{v_{\ell} \in \bar{S}_i, v_j \in S_i} w_{\ell j} \right) \\
&= \frac{1}{2} \cdot \frac{1}{|S_i|} (W(S_i, \bar{S}_i) + W(\bar{S}_i, S_i)) \\
&= \frac{\text{Cut}(S_i, \bar{S}_i)}{|S_i|}.
\end{aligned}$$

Como $(\mathbf{f}^{(i)})^T L \mathbf{f}^{(i)} = (F_P^T L F_P)_{ii}$, então

$$\begin{aligned}
\text{RatioCut}(S_1, \dots, S_k) &= \sum_{i=1}^k \frac{\text{Cut}(S_i, \bar{S}_i)}{|S_i|} = \sum_{i=1}^k (\mathbf{f}^{(i)})^T L \mathbf{f}^{(i)} = \sum_{i=1}^k (F_P^T L F_P)_{ii} \\
&= \text{tr}(F_P^T L F_P).
\end{aligned}$$

Portanto o problema de minimização do RatioCut, definido em (3.1), pode ser reescrito como

$$\min_{P \in \mathcal{P}_G(k)} \text{tr}(F_P^T L F_P).$$

Agora iremos reformular o problema NCut, dado em (3.2). Para isso, seja $P = \{S_1, \dots, S_k\}$ uma partição de V em k classes. A matriz $X_P \in \mathbb{R}^{n \times k}$ tem como colunas os k vetores $\mathbf{x}^{(i)} = (x_1^{(i)}, \dots, x_n^{(i)})$ cujas coordenadas são

$$x_j^{(i)} = \begin{cases} \frac{1}{\sqrt{\text{vol}(S_i)}} & \text{se } v_j \in S_i \\ 0 & \text{se } v_j \notin S_i \end{cases}$$

para cada $j \in \{1, \dots, n\}$.

Utilizando a Proposição 2.11, pode-se observar que

$$(\mathbf{x}^{(i)})^T L \mathbf{x}^{(i)} = \frac{1}{2} \sum_{\ell, j=1}^n w_{\ell j} (x_{\ell}^{(i)} - x_j^{(i)})^2$$

$$\begin{aligned}
&= \frac{1}{2} \left(\sum_{v_\ell \in S_i, v_j \in \bar{S}_i} w_{\ell j} \frac{1}{\text{vol}(S_i)} + \sum_{v_\ell \in \bar{S}_i, v_j \in S_i} w_{\ell j} \frac{1}{\text{vol}(S_i)} \right) \\
&= \frac{1}{2} \cdot \frac{1}{\text{vol}(S_i)} \left(\sum_{v_\ell \in S_i, v_j \in \bar{S}_i} w_{\ell j} + \sum_{v_\ell \in \bar{S}_i, v_j \in S_i} w_{\ell j} \right) \\
&= \frac{1}{2} \cdot \frac{1}{\text{vol}(S_i)} (W(S_i, \bar{S}_i) + W(\bar{S}_i, S_i)) \\
&= \frac{\text{Cut}(S_i, \bar{S}_i)}{\text{vol}(S_i)}.
\end{aligned}$$

Como $(\mathbf{x}^{(i)})^T L \mathbf{x}^{(i)} = (X_P^T L X_P)_{ii}$, temos que

$$\text{NCut}(S_1, \dots, S_k) = \sum_{i=1}^k (\mathbf{x}^{(i)})^T L \mathbf{x}^{(i)} = \sum_{i=1}^k (X_P^T L X_P)_{ii} = \text{tr}(X_P^T L X_P).$$

Substituindo $Y_P = D^{-1/2} X_P$, tem-se que $X_P^T L X_P = Y_P^T D^{-1/2} L D^{-1/2} Y_P = Y_P^T \mathcal{L} Y_P$. Portanto, é possível reescrever o problema de minimização NCut, definido em (3.2), como

$$\min_{P \in \mathcal{P}_G(k)} \text{tr}(Y_P^T \mathcal{L} Y_P). \quad (3.3)$$

É possível verificar que $Y_P^T Y_P = I$. Portanto, o problema de minimização dado em (3.2) pode ser reescrito como um problema cuja solução é uma matriz que satisfaz $Y_P^T Y_P = I$ e que pertençam a uma coleção finita de matrizes, já que os Y_P estão associados a partições de k classes. Caso este problema valesse para qualquer matriz $Y \in \mathbb{R}^{n \times k}$ que satisfizesse $Y^T Y = I$, ele seria reescrito como

$$\min_{Y \in \mathbb{R}^{n \times k} \text{ e } Y^T Y = I} \text{tr}(Y^T \mathcal{L} Y). \quad (3.4)$$

Sejam $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(k)}$ os autovetores ortonormais associados aos k menores autovalores $\lambda_1^{(\mathcal{L})} \leq \dots \leq \lambda_k^{(\mathcal{L})}$, respectivamente, pelo Teorema 2.8 (Rayleigh-Ritz), sabemos que a expressão em (3.4) é igual a $\lambda_1^{(\mathcal{L})} + \dots + \lambda_k^{(\mathcal{L})}$ e, além disso, Y satisfaz essa igualdade se suas colunas são vetores que geram $\text{Span} = \{\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(k)}\}$. Esta transformação do problema com restrições discretas para o de restrições contínuas, neste caso do problema em (3.3) para (3.4), é o que chamamos de *relaxação do problema*.

Desenvolvendo de forma semelhante ao caso do problema NCut, é possível também obter o problema relaxado do RatioCut, que é dado por

$$\min_{F \in \mathbb{R}^{n \times k} \text{ e } F^T F = I} \text{tr}(F^T L F).$$

Uma possível matriz Y que satisfaz a solução em (3.4) é aquela cujas colunas são os vetores $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(k)}$. É importante notar que esta solução não tem relação com uma partição de V , ao contrário daquela encontrada em (3.3). Por outro lado, a solução do problema relaxado é mais fácil de ser encontrada em muitos casos e, por essa razão, se espera que ela traga informações sobre os vértices do grafo que poderão ser úteis para tentar encontrar a partição do problema original. Possíveis estratégias de como encontrar esta solução serão discutidas na seção seguinte.

3.3 Métodos Espectrais

Na seção anterior, vimos que é possível obter um problema relaxado a partir dos problemas de minimização das funções RatioCut e NCut e que sua solução é obtida utilizando os autovetores das matrizes laplaciana e laplaciana normalizada associadas ao grafo de entrada, respectivamente. Em geral, esses autovetores não satisfazem as restrições discretas definidas no problema obtido antes da relaxação. Por outro lado, como já comentamos, é natural associar as coordenadas de cada vetor aos vértices correspondentes, produzindo um vetor para cada vértice. Se aplicássemos um método de classificação para estes vetores, intuitivamente estaríamos obtendo uma classificação de vértices do grafo que poderia ser uma partição com boa qualidade. A métodos que utilizam propriedades do espectro para classificação damos o nome de *algoritmos espectrais* ou *heurísticas espectrais*. Neste trabalho chamaremos de heurísticas espectrais os procedimentos para os quais ainda não se têm alguma garantia teórica sobre a qualidade da solução apresentada.

No Capítulo 4 serão apresentados dois algoritmos espectrais que obtêm os vetores-linha de forma distinta da que foi apresentada no parágrafo anterior. Além disso, os métodos de classificação de vetores-linha se diferem um do outro. De forma geral, as estratégias espectrais que estudaremos seguem do seguinte algoritmo.

Algoritmo Espectral:

Input: Grafo $G = (V, E)$ com vértices v_1, \dots, v_n e inteiro positivo $k \geq 2$

Output: Partição (ou coleção de k subconjuntos disjuntos não vazios) $\{S_1, \dots, S_k\}$ de V

- 1 Escolher uma matriz M associada a G ;
 - 2 A partir dos autovetores de M , definir um conjunto de vetores $\mathbf{f}(v_1), \dots, \mathbf{f}(v_n) \in \mathbb{R}^\ell$ associados aos vértices de V ;
 - 3 Utilizar algum procedimento para particionar os vetores $\mathbf{f}(v_j)$ em k conjuntos F_1, \dots, F_k ;
 - 4 $S_i = \{v_j : \mathbf{f}(v_j) \in F_i\}$.
-

Neste trabalho, iremos nos referir aos passos (1) e (2), juntos, como o *passo spectral* do algoritmo, enquanto o (3) e o (4) formam o *passo de clustering*. Tipicamente, a matriz M do passo (1) é simétrica, de forma que produza um conjunto ortonormal de autovetores associados a autovalores reais, utilizados no passo (2) para definir os vetores associados aos vértices do grafo.

Com relação ao passo de *clustering*, dentre todos aqueles conhecidos na literatura, o mais popular é o algoritmo das k -médias. Apresentado por MacQueen [27] e aprimorado por Lloyd [25], o algoritmo funciona da seguinte forma. Sejam os vetores $\mathbf{f}(v_1), \dots, \mathbf{f}(v_n) \in \mathbb{R}^\ell$, o algoritmo inicia tomando, segundo um critério pré-estabelecido (em muitos casos, escolhendo de forma aleatória), k vetores $\mathbf{c}_1, \dots, \mathbf{c}_k \in \mathbb{R}^\ell$ distintos chamados de *centroides*. Em seguida verifica, para cada $\mathbf{f}(v_i)$, qual é a centroide \mathbf{c}_j mais próximo a ele e o aloca a um conjunto F_j associado a essa centroide. Depois de realizar esta operação para cada $i = 1, \dots, n$, cada centroide \mathbf{c}_j será recalculada através da média dos vetores $\mathbf{f}(v_i)$ pertencentes a F_j . Estes passos

se repetem até que as centroides não mudem mais e, com isso, obtemos os conjuntos de vértices S_1, \dots, S_k tais que $S_j = \{v_i : \mathbf{f}(v_i) \in F_j\}$.

É possível mostrar que esse procedimento sempre chega em uma partição final após um número finito de passos. Para isso, definimos a função objetivo $H : \{1, 2, \dots\} \rightarrow \mathbb{R}$ dada por

$$H(x) = \sum_{i=1}^k \sum_{\mathbf{f}(v_j) \in S_i(x)} \|\mathbf{f}(v_j) - \mathbf{c}_i(x)\|^2,$$

onde x representa a iteração do algoritmo, $S_i(x)$ é o i -ésimo conjunto produzido pelo algoritmo após a x -ésima iteração e

$$\mathbf{c}_i = \frac{1}{|S_i(x)|} \sum_{\mathbf{f}(v_j) \in S_i} \mathbf{f}(v_j).$$

Selim e Ismail [31] apresentam uma demonstração que prova que a função H converge e decresce cada vez que o algoritmo altera a partição. Por essa razão, o algoritmo das k -médias termina.

Vejamos o que ocorre se quiséssemos particionar o grafo $G = (V, E)$, representado pela Figura 3.2a, em $k = 3$ classes utilizando a matriz laplaciana normalizada (conforme o problema relaxado do NCut) e o algoritmo de classificação das k -médias. Para efeitos de comparação, observamos primeiro que a partição que tem menor NCut para $k = 3$ é aquela apresentada na Figura 3.2b.

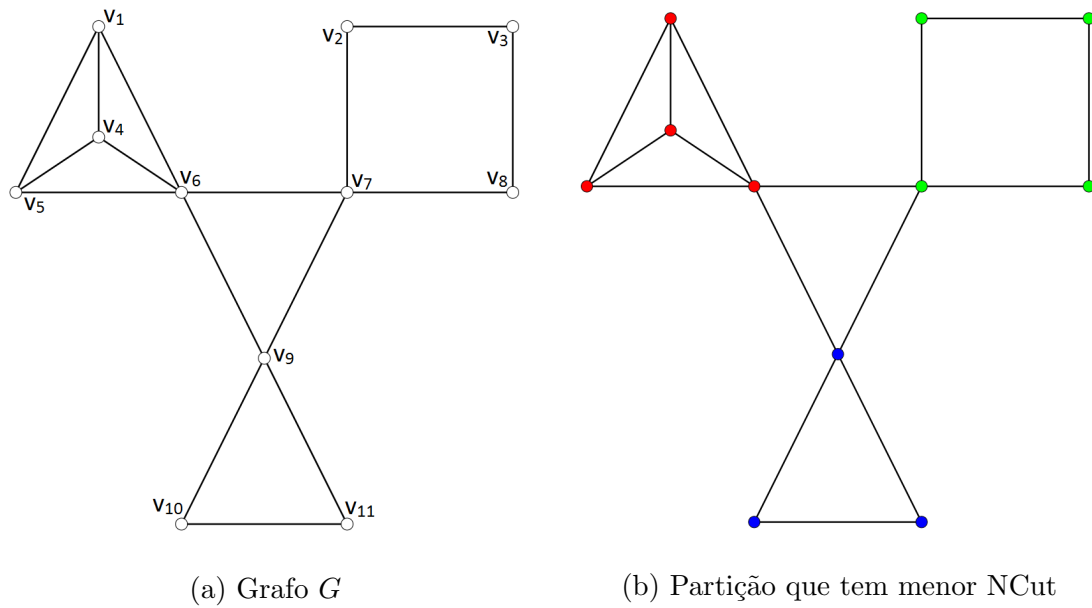


Figura 3.2: Exemplo de grafo

Ao aplicar o passo espectral do algoritmo, começamos calculando a matriz

$$X = \begin{bmatrix} 0.31 & -0.37 & 0.17 \\ 0.25 & 0.36 & 0.22 \\ 0.25 & 0.43 & 0.29 \\ 0.31 & -0.37 & 0.17 \\ 0.31 & -0.37 & 0.17 \\ 0.40 & -0.25 & 0.08 \\ 0.35 & 0.26 & 0.09 \\ 0.25 & 0.36 & 0.22 \\ 0.35 & 0.07 & -0.42 \\ 0.25 & 0.07 & -0.52 \\ 0.25 & 0.07 & -0.52 \end{bmatrix}$$

cujas colunas são os autovetores associados aos $k = 3$ menores autovalores da matriz laplaciana normalizada de G . A cada vértice de G , associamos o vetor dado pela

linha da matriz X . Por exemplo, o vértice v_6 está associado ao vetor $\mathbf{f}(v_6) = (0.40, -0.25, 0.08)$.

Para distribuir os vértices em classes, vamos aplicar o algoritmo das k -médias a esses vetores. Caso iniciássemos o algoritmo com as centroides $\mathbf{c}_1 = (0.32, -0.08, -0.01)$, $\mathbf{c}_2 = (0.25, 0.22, -0.15)$ e $\mathbf{c}_3 = (0.25, 0.43, 0.29)$, encontraríamos como resultado a partição $\{v_1, v_4, v_5, v_6\}$, $\{v_2, v_3, v_7, v_8\}$ e $\{v_9, v_{10}, v_{11}\}$, que coincide com a dada pela Figura 3.2b.

As heurísticas espectrais de forma geral não garantem que a solução ótima do problema original seja encontrada. Por exemplo, um particionamento do exemplo acima foi obtido por um procedimento para os quais não se tem uma garantia teórica. Porém, existem estudos que apontam que essa heurística apresenta um bom desempenho em muitas situações. Muitos desses estudos são baseados em experimentos computacionais ou estudos empíricos. Em muitos casos, o algoritmo das k -médias é utilizado para classificar os vetores obtidos pelo passo espectral da heurística.

Por outro lado, também existem muitos trabalhos que demonstram fatos sobre o desempenho desses algoritmos, veremos exemplo disso no próximo capítulo. Porém, é raro que esses algoritmos envolvam o passo das k -médias, pois esse é um algoritmo difícil de analisar. Uma das dificuldades na análise do desempenho das k -médias é sua inicialização, pois poderão haver diferentes resultados finais do algoritmo dependendo da inicialização realizada.

Voltando para aplicação do k -médias para o grafo 3.2a, caso as centroides escolhidas inicialmente fossem $\mathbf{c}_1 = (0.28, 0.03, 0.23)$, $\mathbf{c}_2 = (0.35, 0.07, -0.42)$ e $\mathbf{c}_3 = (0.25, 0.07, -0.52)$ obteríamos a partição da Figura 3.3a como resultado. Com as mesmas centroides \mathbf{c}_1 e \mathbf{c}_3 e tomando $\mathbf{c}_2 = (0.35, 0.07, -0.42)$ a partição resultante seria a da Figura 3.3b. Em ambos os casos a partição final é diferente daquela representada na Figura 3.2b.

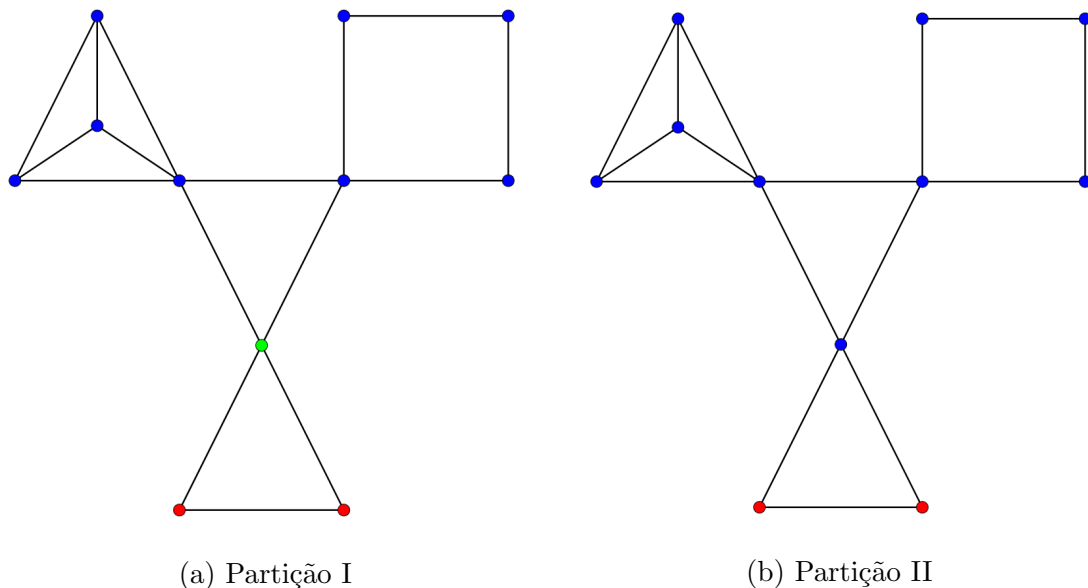


Figura 3.3: Outras partições geradas pelo algoritmo k -médias

Sobre a partição da Figura 3.3a, como os vetores $\mathbf{f}(v_9)$ e $\mathbf{f}(v_{10}) = \mathbf{f}(v_{11})$ estão muito próximos um do outro e as centroides $\mathbf{c}_2 = (0.35, 0.07, -0.42)$ e $\mathbf{c}_3 = (0.25, 0.07, -0.52)$ estão muito próximas a $\mathbf{f}(v_9)$ e $\mathbf{f}(v_{10}) = \mathbf{f}(v_{11})$, respectivamente, o vértice v_9 acabou ficando em classe diferente de v_{10} e v_{11} na partição final. Quanto ao caso em que a partição obtida foi a da Figura 3.3b, a centroide $\mathbf{c}_2 = (0.35, 0.07, -0.42)$ não está posicionada de forma que seja a centroide mais próxima de algum vértice, o que ocasiona numa situação em que um dos conjuntos se torna vazio após a primeira iteração e, portanto, não resultando numa partição em $k = 3$ classes como resultado final. Trabalhos como os de Fränti e Sieranoja [16] se concentram em estudar qual poderia ser uma inicialização de centroides apropriada para o k -médias.

Um passo de *clustering* mais simples e que funcionaria para o grafo 3.2a seria classificar as linhas de X conforme os sinais das coordenadas. Convenientemente, há três tipos diferentes de sinais para cada linha de X : $(+)(-)(+)$, $(+)(+)(+)$ e $(+)(+)(-)$. Classificando os vértices conforme a classificação dos sinais

de seus vetores-linha correspondentes, obtemos a partição da Figura 3.2b. Se nosso objetivo fosse particionar o grafo em $k > 3$ classes, talvez a análise de sinais não seria a melhor opção.

3.4 Particionamento de Árvores

Como o nosso objetivo neste trabalho é investigar a qualidade das estratégias espectrais, é necessário que possamos, pelo menos em alguns casos, conhecer a partição ótima para que o resultado obtido pela estratégia possa ser comparado a ela. Em certas situações, conhecer propriedades da partição ótima permite, inclusive, resolver problemas em tempo polinomial.

As árvores são tipos importantes de grafos, afinal todo grafo conexo possui um subgrafo gerador que seja uma árvore, chamada de *árvore geradora*. No trabalho realizado por Wang et. al [35], foi apresentado um teorema que prova que a partição que possui menor valor NCut em um árvore é tal que os subgrafos induzidos de suas classes são conexos. Além disso, eles afirmam que os subgrafos induzidos também são conexos caso eles pertencessem à partição de menor NCut para $k > 2$, porém não apresentam a prova e não citam a referência. Com este último resultado, seria possível encontrar a melhor partição NCut para k fixo, pois cada partição seria obtida removendo exatamente as $k - 1$ arestas que ligam as k classes duas a duas. Podemos verificar o valor de todas as possíveis partições desse tipo em tempo polinomial.

Sabendo que o resultado vale para $k = 2$, uma ideia de demonstração para $k > 2$ é utilizando a indução. Suponha que o resultado vale para $k - 1 \geq 2$ e que tenhamos a garantia de que, na melhor partição $P = \{S_1, \dots, S_k\}$ em k classes, existe S_i que possui subgrafo induzido conexo. Seja $G' = (V', E')$ o subgrafo de G obtido pela remoção de S_i de suas arestas incidentes, seria possível afirmar que a partição $\{S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_k\}$ é aquela que possui menor NCut entre as partições

de V' em $k - 1$ classes; caso contrário, P não seria a melhor partição de V . Com isso, concluiríamos que todas as classes de P possuem subgrafos induzidos conexos.

Um indício de que existe este conjunto S_i é um resultado que será enunciado e demonstrado a seguir (Teorema 3.7), que diz que, na partição ótima do problema NCut, um vértice pendente (que é adjacente a somente um vértice) pertence sozinho a uma classe dessa partição ou o vértice adjacente pertence à mesma classe que ele. Caso contraíssemos as arestas da classe contendo o vértice pendente, obteríamos um grafo em que as informações sobre essa classe não apareceria. Para tentar contornar isso, gostaríamos que o Teorema 3.7 fosse estendido para grafos com vértices ponderados. Dessa forma, será possível armazenar o peso das arestas do conjunto contraído neste vértice.

Por esse motivo, queremos estender o resultado de Wang et. al para grafos com vértices ponderados. Antes de enunciá-lo, definimos um grafo $G = (V, E)$ cujos vértices e arestas estão associados às funções peso $p : V \rightarrow \mathbb{R}_{\geq 0}$ e $w : E \rightarrow \mathbb{R}_{> 0}$, respectivamente. Neste tipo de grafo, definimos $W(S) = \sum_{v \in S} 2p_v + 2w(G[S]) + w(S, \bar{S})$, onde $w(G[S]) = \sum_{v_i, v_j \in S} w_{ij}$ e $w(S, \bar{S}) = \sum_{v_i \in S, v_j \in \bar{S}} w_{ij}$. A quantidade $W(S)$ faz o papel do volume no caso de grafos com pesos nos vértices. Por essa razão, dada a partição $\{S_1, \dots, S_k\}$ de V , definimos para este tipo de grafo o corte normalizado como

$$\text{NCut}(S_1, \dots, S_k) = \sum_{i=1}^k \frac{w(S_i, \bar{S}_i)}{W(S_i)}.$$

Teorema 3.5. *Seja $T = (V, E)$ uma árvore cujos vértices e arestas estão associados às funções peso $p : V \rightarrow \mathbb{R}_{\geq 0}$ e $w : E \rightarrow \mathbb{R}_{> 0}$, respectivamente. Sejam os conjuntos de vértices $A, B \in V$ tais que*

$$\text{NCut}(A, B) = \min_{X \cap Y = \emptyset, X \cup Y = V} \text{NCut}(X, Y),$$

então os subgrafos induzidos $T[A]$ e $T[B]$ são conexos.

Demonstração. Por simplicidade seja, para um subconjunto $S \in V$, $\alpha_S = W(S) - w(S, \bar{S})$. Vamos supor por absurdo que pelo menos um dos subgrafos $T[A]$ e $T[B]$ é desconexo. Logo, existe um conjunto de arestas $\{(u_i, v_i)\}_{i=1}^m$ com $m > 1$, $u_i \in A$ e $v_i \in B$. Removendo cada uma dessas arestas, obtemos um conjunto de $m + 1$ subárvores conexas $\{V_j\}_{j=0}^m$. Com essas informações, podemos concluir que

$$\begin{aligned}
\text{NCut}(A, B) &= \frac{w(A, B)}{W(A)} + \frac{w(A, B)}{W(B)} = \frac{\sum_{i=1}^m w(u_i, v_i)}{\alpha_A + \sum_{i=1}^m w(u_i, v_i)} + \frac{\sum_{i=1}^m w(u_i, v_i)}{\alpha_B + \sum_{i=1}^m w(u_i, v_i)} \\
&= \frac{\left(\alpha_A + \alpha_B + 2 \sum_{i=1}^m w(u_i, v_i)\right) \sum_{i=1}^m w(u_i, v_i)}{\left(\alpha_A + \sum_{i=1}^m w(u_i, v_i)\right) \left(\alpha_B + \sum_{i=1}^m w(u_i, v_i)\right)} \\
&= \frac{\alpha_T \sum_{i=1}^m w(u_i, v_i)}{\left(\alpha_A + \sum_{i=1}^m w(u_i, v_i)\right) \left(\alpha_B + \sum_{i=1}^m w(u_i, v_i)\right)}.
\end{aligned}$$

Agora sejam $\{A_i, B_i\}_{i=1}^m$ o conjunto das m possíveis partições de T , onde $\{A_i, B_i\}$ é aquela que é obtida quando a aresta $\{u_i, v_i\}$ é removida. Dessa forma temos que

$$\begin{aligned}
\text{NCut}(A_i, B_i) &= \frac{w(A_i, B_i)}{W(A_i)} + \frac{w(A_i, B_i)}{W(B_i)} = \frac{w(u_i, v_i)}{\alpha_{A_i} + w(u_i, v_i)} + \frac{w(u_i, v_i)}{\alpha_{B_i} + w(u_i, v_i)} \\
&= \frac{[\alpha_{A_i} + \alpha_{B_i} + 2w(u_i, v_i)] \sum_{i=1}^m w(u_i, v_i)}{[\alpha_{A_i} + w(u_i, v_i)][\alpha_{B_i} + w(u_i, v_i)]} \\
&= \frac{\alpha_T \cdot w(u_i, v_i)}{[\alpha_{A_i} + w(u_i, v_i)][\alpha_{B_i} + w(u_i, v_i)]}.
\end{aligned}$$

Definindo $\text{NCut}_i^* = \frac{\text{NCut}(A_i, B_i)}{\alpha_T}$ e utilizando o Lema 2.19 e um lema técnico (utilizado somente nesta prova e que será provado após o término desta), temos que

$$\min\{\text{NCut}_i^*\}_{i=1}^m \leq \frac{\sum_{i=1}^m w(u_i, v_i)}{\sum_{i=1}^m [\alpha_{A_i} + w(u_i, v_i)][\alpha_{B_i} + w(u_i, v_i)]} \quad (\text{Lema 2.19})$$

$$< \frac{\sum_{i=1}^m w(u_i, v_i)}{\left(\alpha_A + \sum_{i=1}^m w(u_i, v_i)\right) \left(\alpha_B + \sum_{i=1}^m w(u_i, v_i)\right)} \quad (\text{Lema 3.6}).$$

Portanto, existe uma partição (A_i, B_i) cujo corte normalizado é menor que o de (A, B) , o que contradiz a hipótese. Logo, $T[A]$ e $T[B]$ são conexos. \square

Lema 3.6. *Suponha que as hipóteses do Teorema 3.5 sejam válidas, então*

$$\sum_{i=1}^m [\alpha_{A_i} + w(u_i, v_i)][\alpha_{B_i} + w(u_i, v_i)] > \left[\alpha_A + \sum_{i=1}^m w(u_i, v_i) \right] \left[\alpha_B + \sum_{i=1}^m w(u_i, v_i) \right]. \quad (3.5)$$

Demonstração. Seja T^* uma árvore cujos vértices são representados pelas subárvores $\{V_j\}_{j=0}^m$ e cujas arestas são os $\{u_i, v_i\}$.

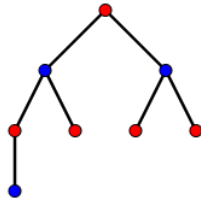


Figura 3.4: Exemplo de Árvore T^*

Tomando um dos vértices de T^* como raiz, note que

- 1) os vértices que pertencem a um nível ímpar de T^* (em vermelho na Figura 3.4) estão em A e a um nível par estão em B (em azul na Figura 3.4);
- 2) em T^* , a partição $\{A_i, B_i\}$ ocorre com a remoção da aresta $\{u_i, v_i\}$.

Com essas duas afirmações, podemos concluir que

$$\alpha_A \alpha_B = \left(\sum_{V_{imp} \in A} \alpha_{V_{imp}} \right) \left(\sum_{V_{par} \in B} \alpha_{V_{par}} \right) < \sum_{i=1}^m \left(\sum_{V_j \in A_i} \alpha_{A_i} \right) \left(\sum_{V_j \in B_i} \alpha_{B_i} \right). \quad (3.6)$$

A desigualdade (3.6) é atingida, pois todos os termos $\alpha_{V_{imp}} \cdot \alpha_{V_{par}}$ da expressão à esquerda da desigualdade aparecem pelo menos uma vez na da direita. Isto pode ser verificado se removermos uma aresta $\{u_i, v_i\}$ que esteja no caminho entre V_{imp} e V_{par} , fazendo com que pertençam a A_i e B_i , respectivamente.

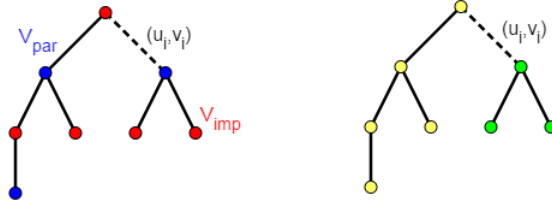


Figura 3.5: Remoção da aresta $\{u_i, v_i\}$, que está no caminho de V_{imp} a V_{par} , gerando os conjuntos de vértices A_i (em amarelo) e B_i (em verde).

Expandindo o lado esquerdo da desigualdade em (3.5), temos

$$\begin{aligned} & \sum_{i=1}^m [\alpha_{A_i} + w(u_i, v_i)] [\alpha_{B_i} + \omega(u_i, v_i)] \\ &= \sum_{i=1}^m \left[\sum_{V_j \in A_i} \alpha_{V_j} + \sum_{(u_j, v_j) \in A_i} 2w(u_j, v_j) + w(u_i, v_i) \right] \\ & \quad \cdot \left[\sum_{V_j \in B_i} \alpha_{V_j} + \sum_{(u_j, v_j) \in B_i} 2w(u_j, v_j) + w(u_i, v_i) \right] \\ &> \sum_{i=1}^m \left[\sum_{V_j \in A_i} \alpha_{V_j} \sum_{V_j \in B_i} \alpha_{V_j} + \left(\sum_{V_j \in A_i} \alpha_{V_j} + \sum_{V_j \in B_i} \alpha_{V_j} \right) w(u_i, v_i) \right. \\ & \quad \left. + \left(2 \sum_{(u_j, v_j) \in A_i} w(u_j, v_j) + w(u_i, v_i) \right) \left(2 \sum_{(u_j, v_j) \in B_i} w(u_j, v_j) + w(u_i, v_i) \right) \right] \end{aligned}$$

$$\begin{aligned}
&> \alpha_A \alpha_B + \sum_{j=0}^m \alpha_{V_j} \sum_{i=1}^m w(u_i, v_i) + \sum_{i=1}^m \left[(w(u_i, v_i))^2 + w(u_i, v_i) \sum_{i \neq j} w(u_j, v_j) \right] \\
&= \alpha_A \alpha_B + (\alpha_A + \alpha_B) \sum_{i=1}^m w(u_i, v_i) + \sum_{i=1}^m w(u_i, v_i) \left[w(u_i, v_i) + \sum_{i \neq j} w(u_j, v_j) \right] \\
&= \alpha_A \alpha_B + (\alpha_A + \alpha_B) \sum_{i=1}^m w(u_i, v_i) + \left[\sum_{i=1}^m w(u_j, v_j) \right]^2 \\
&= \left[\alpha_A + \sum_{i=1}^m w(u_i, v_i) \right] \left[\alpha_B + \sum_{i=1}^m w(u_i, v_i) \right]
\end{aligned}$$

□

O seguinte teorema é aplicado para grafos que não possuem pesos nos vértices. Porém, caso este resultado seja estendido para grafos cujos vértices estão associados a pesos, poderia auxiliar a proceder com a estratégia de demonstração do Teorema 3.5 para $k > 2$ mencionada anteriormente.

Teorema 3.7. *Seja um grafo G com arestas ponderadas onde o peso é dado por $w : E \rightarrow \mathbb{R}$ e seja $\{S_1, \dots, S_k\}$ a partição de V que é solução do problema NCut para $k \geq 2$ fixo. Se $v \in V$ é um vértice pendente cujo vértice adjacente é $u \in V$ e $v \in S_j$, então uma das duas afirmações é verdadeira:*

(1) $S_j = \{v\}$;

(2) Se existe $w \neq v$ tal que $w \in S_j$, então $u \in S_j$.

Demonstração. Seja $\{S_1, \dots, S_k\}$ a partição de V que é solução do problema NCut para $k \geq 2$ fixo e $v \in S_j$ adjacente a somente um vértice $u \in V$. Para provar este resultado, basta tomar $u \in S_i$, onde $i \neq j$, e $w \in S_j$ diferente de v e encontrar uma outra partição que gera NCut menor que a gerada por $\{S_1, \dots, S_k\}$, chegando a uma contradição.

Tomamos a partição $\{S_1^*, \dots, S_k^*\}$ onde $S_i^* = S_i \cup \{v\}$, $S_j^* = S_j \setminus \{v\}$ e $S_\ell^* = S_\ell$ para ℓ diferente de i e j . Queremos provar que $\text{NCut}(S_1^*, \dots, S_k^*) >$

$\text{NCut}(S_1^*, \dots, S_k^*)$. Seja $N = \text{NCut}(S_1, \dots, S_k) - \text{NCut}(S_1^*, \dots, S_k^*)$, note que

$$\begin{aligned}
N &= \sum_{\ell=1}^k \frac{\text{Cut}(S_\ell, \overline{S}_\ell)}{\text{vol}(S_\ell)} - \sum_{\ell=1}^k \frac{\text{Cut}(S_\ell^*, \overline{S}_\ell^*)}{\text{vol}(S_\ell^*)} \\
&= \frac{\text{Cut}(S_i, \overline{S}_i)}{\text{vol}(S_i)} + \frac{\text{Cut}(S_j, \overline{S}_j)}{\text{vol}(S_j)} - \frac{\text{Cut}(S_i^*, \overline{S}_i^*)}{\text{vol}(S_i^*)} - \frac{\text{Cut}(S_j^*, \overline{S}_j^*)}{\text{vol}(S_j^*)} \\
&= \frac{\text{Cut}(S_i, \overline{S}_i)}{\text{vol}(S_i)} + \frac{\text{Cut}(S_j, \overline{S}_j)}{\text{vol}(S_j)} - \frac{\text{Cut}(S_i, \overline{S}_i) - w(u, v)}{\text{vol}(S_i) + w(u, v)} - \frac{\text{Cut}(S_j, \overline{S}_j) - w(u, v)}{\text{vol}(S_j) - w(u, v)}.
\end{aligned}$$

Como $\text{vol}(S_j) \geq \text{Cut}(S_j, \overline{S}_j)$, então $N > 0$, chegando a uma contradição. \square

4 GARANTIAS TEÓRICAS SOBRE O DESEMPENHO DE MÉTODOS ESPECTRAIS

Neste capítulo, apresentaremos resultados teóricos rigorosos que evidenciam o bom desempenho de algoritmos espectrais. O parâmetro para o qual os resultados serão provados é chamado de condutância, que é uma variação da função NCut, conforme definição que será dada na primeira subseção deste capítulo. Estes algoritmos serão apresentadas e discutidas passo a passo e ilustradas por aplicações em exemplos.

4.1 Expansão e Condutância

Dada uma sequência de k valores não negativos, duas maneiras tradicionais de definir uma norma são a norma 1, que é o valor da soma de todos os elementos da sequência, e norma infinito, que toma o maior valor dentre os elementos dela. Seja uma partição $\{S_1, \dots, S_k\}$ do conjunto de vértices de um grafo, a norma 1 das sequências $\left\{ \frac{\text{Cut}(S_i, \bar{S}_i)}{|S_i|} \right\}_{i=1}^k$ e $\left\{ \frac{\text{Cut}(S_i, \bar{S}_i)}{\text{vol}(S_i)} \right\}_{i=1}^k$ são as funções RatioCut e NCut, respectivamente. Neste capítulo, consideraremos funções de particionamento, definidas por Chung [8], que são as normas infinito destas duas sequências mencionadas: expansão e condutância.

Definição 4.1. *Seja $G = (V, E)$ e $S \subseteq V$. Definimos a expansão de S como*

$$\psi_G(S) = \frac{\text{Cut}(S, \bar{S})}{|S|}.$$

Agora, dada uma coleção de conjuntos disjuntos $P = \{S_1, \dots, S_k\}$, definimos a expansão de P como

$$\psi_G(S_1, \dots, S_k) = \max_{i=1, \dots, k} \psi_G(S_i).$$

Definição 4.2. *Seja $G = (V, E)$ e $S \subseteq V$. Definimos a condutância de S como*

$$\phi_G(S) = \frac{\text{Cut}(S, \bar{S})}{\text{vol}(S)}.$$

Agora, dada uma coleção de conjuntos disjuntos $P = \{S_1, \dots, S_k\}$, definimos a condutância de P como

$$\phi_G(S_1, \dots, S_k) = \max_{i=1, \dots, k} \phi_G(S_i).$$

Conforme as definições dadas, é fácil de verificar que, para uma partição $\{S_1, \dots, S_k\}$, valem as desigualdades

$$\begin{aligned} \psi_G(S_1, \dots, S_k) &\leq \text{RatioCut}(S_1, \dots, S_k) \leq k \cdot \psi_G(S_1, \dots, S_k) \text{ e} \\ \phi_G(S_1, \dots, S_k) &\leq \text{NCut}(S_1, \dots, S_k) \leq k \cdot \phi_G(S_1, \dots, S_k). \end{aligned}$$

Além disso, conforme definições do valor de corte e da medida de volume, temos que, para um subconjunto $S \subseteq V$, vale $\phi_G(S) \leq 1$. Esta desigualdade implica que valores máximos e mínimos relacionados à condutância, como $\phi_G(S_1, \dots, S_k)$, são menores ou iguais a 1.

Ao contrário das funções Cut , RatioCut e NCut , aqui definimos o domínio de $\psi_G(S_1, \dots, S_k)$ e $\phi_G(S_1, \dots, S_k)$ como todas as possíveis coleções de k conjuntos de vértices não vazios e disjuntos, e não apenas todas as possíveis partições em k classes do grafo. Optamos por este domínio, pois ele permite que demonstremos os resultados das seções seguintes. Por simplicidade iremos nos referir às coleções de k conjuntos utilizadas nestas funções como partições e também denotaremos esta coleção por $\mathcal{P}_G(k)$.

4.2 Desigualdade de Cheeger de Ordem 2

Assim como os problemas de minimização do RatioCut e NCut , o de condutância também é NP-difícil. O que se sabe é que há garantias sobre o valor ótimo do problema, que possui cotas inferiores e superiores dependentes de um

autovalor da matriz laplaciana normalizada do grafo em questão. Para podermos demonstrar algumas destas provas, vamos primeiro definir o valor ótimo deste problema da seguinte forma.

Definição 4.3. *Seja $G = (V, E)$, $k \geq 2$ inteiro e $\mathcal{P}_G(k)$ o conjunto de todas as possíveis coleções de k subconjuntos disjuntos e não vazios de V . Definimos a condutância de G de ordem k como:*

$$\phi_G(k) = \min_{\{S_1, \dots, S_k\} \in \mathcal{P}_G(k)} \phi_G(S_1, \dots, S_k) = \min_{\{S_1, \dots, S_k\} \in \mathcal{P}_G(k)} \max_{i=1, \dots, k} \phi_G(S_i).$$

Em particular, a condutância de G de ordem 2 pode ser denominada simplesmente por *condutância de G* e é possível verificar que

$$\phi_G(2) = \min_{\text{vol}(S) \leq \frac{\text{vol}(V)}{2}} \phi_G(S) = \min_{S \neq \emptyset \subset V} \frac{\text{Cut}(S, \bar{S})}{\min\{\text{vol}(S), \text{vol}(\bar{S})\}}.$$

Tais versões da condutância de G indicam que sempre haverá uma partição $\{S, \bar{S}\}$ tal que $\phi_G(S, \bar{S}) = \phi_G(2)$, enquanto que para ordem $k > 2$ bastaria, em alguns casos, que todos os k conjuntos fossem disjuntos, mas que não formassem necessariamente uma partição.

4.2.1 Demonstração da Validade da Desigualdade de Cheeger

Nesta seção, demonstraremos a desigualdade de Cheeger de ordem 2. A prova da cota superior desta desigualdade é construtiva e fornece, a partir do espectro da matriz laplaciana normalizada, uma partição $\{S, \bar{S}\}$ tal que $\phi_G(2) \leq \phi_G(S, \bar{S}) \leq 2\sqrt{\phi_G(2)}$, trazendo evidências de que a partição encontrada tenha condutância próxima do valor ótimo $\phi_G(2)$.

Teorema 4.4 (Desigualdade de Cheeger de Ordem 2). *Seja $G = (V, E)$ e \mathcal{L} sua matriz laplaciana normalizada, cujo segundo menor autovalor é $\lambda_2^{(\mathcal{L})}$. Então,*

$$\frac{\lambda_2^{(\mathcal{L})}}{2} \leq \phi_G(2) \leq \sqrt{2\lambda_2^{(\mathcal{L})}}.$$

Demonstração. Inicialmente vamos mostrar que $\frac{\lambda_2^{(\mathcal{L})}}{2} \leq \phi_G(2)$.

Seja n o número de vértices em G e S um conjunto de vértices tal que $\phi_G(2) = \phi_G(S)$. Defina $\mathbf{y} \in \mathbb{R}^n$ por:

$$y_i = \begin{cases} \frac{1}{\text{vol}(S)} & \text{se } v_i \in S \\ -\frac{1}{\text{vol}(\bar{S})} & \text{se } v_i \in \bar{S} \end{cases}.$$

É possível notar que $\langle \mathbf{y}, D\mathbf{1} \rangle = 0$. Logo, pelo item (3) da Proposição 2.15, temos que

$$\lambda_2^{(\mathcal{L})} = \min_{\mathbf{x} \perp D\mathbf{1}} \frac{\sum_{\{v_i, v_j\} \in E} w_{ij} (x_i - x_j)^2}{\sum_{j=1}^n x_j^2 d_j} \leq \frac{\sum_{\{v_i, v_j\} \in E} w_{ij} (y_i - y_j)^2}{\sum_{v_i} y_i^2 d_i}. \quad (4.1)$$

Como $\sum_{\{v_i, v_j\} \in E} (y_i - y_j)^2$ se refere à soma sobre todo o conjunto de arestas em E , temos

$$\begin{aligned} \sum_{\{v_i, v_j\} \in E} w_{ij} (y_i - y_j)^2 &= \sum_{\{v_i, v_j\} \in E(S, \bar{S})} w_{ij} (y_i - y_j)^2 \\ &= \text{Cut}(S, \bar{S}) \left(\frac{1}{\text{vol}(S)} + \frac{1}{\text{vol}(\bar{S})} \right)^2, \end{aligned} \quad (4.2)$$

onde a igualdade em (4.2) é atingida, pois $(y_i - y_j)^2 = 0$ quando v_i e v_j pertencem ambos a S ou a \bar{S} .

Já o denominador $\sum_{v_i} y_i^2 d_i$ será

$$\sum_{v_i} y_i^2 d_i = \sum_{v_i \in S} \frac{1}{\text{vol}(S)^2} d_i + \sum_{v_i \in \bar{S}} \frac{1}{\text{vol}(\bar{S})^2} d_i$$

$$\begin{aligned}
&= \frac{\text{vol}(S)}{\text{vol}(S)^2} + \frac{\text{vol}(\bar{S})}{\text{vol}(\bar{S})^2} \\
&= \frac{1}{\text{vol}(S)} + \frac{1}{\text{vol}(\bar{S})}.
\end{aligned}$$

Desse modo,

$$\lambda_2^{(\mathcal{L})} \leq \frac{\text{Cut}(S, \bar{S}) \left(\frac{1}{\text{vol}(S)} + \frac{1}{\text{vol}(\bar{S})} \right)^2}{\frac{1}{\text{vol}(S)} + \frac{1}{\text{vol}(\bar{S})}} = \text{Cut}(S, \bar{S}) \left(\frac{1}{\text{vol}(S)} + \frac{1}{\text{vol}(\bar{S})} \right).$$

Sabendo que $\left(\frac{1}{\text{vol}(S)} + \frac{1}{\text{vol}(\bar{S})} \right) \leq \frac{2}{\min \{ \text{vol}(S), \text{vol}(\bar{S}) \}}$, concluímos que

$$\lambda_2^{(\mathcal{L})} \leq 2 \cdot \frac{\text{Cut}(S, \bar{S})}{\min \{ \text{vol}(S), \text{vol}(\bar{S}) \}} = 2\phi_G(2) \iff \frac{\lambda_2^{(\mathcal{L})}}{2} \leq \phi_G(2).$$

Agora provaremos que $\lambda_2^{(\mathcal{L})} \geq \frac{\phi_G(2)^2}{2}$.

Seja \mathbf{x} o vetor que satisfaz a igualdade em (4.1). Ordenamos os vértices de G de acordo com a ordenação decrescente das coordenadas de \mathbf{x} , isto é, renomeamos os vértices de tal forma que $x_i \geq x_{i+1}$ para $1 \leq i \leq n-1$. Além disso, definiremos $S_i = \{v_1, \dots, v_i\}$, $S_0 = \emptyset$ e

$$\alpha_G = \min_{i=1, \dots, n-1} \frac{\text{Cut}(S_i, \bar{S}_i)}{\min \{ \text{vol}(S_i), \text{vol}(\bar{S}_i) \}}.$$

Seja ℓ o maior inteiro tal que $\text{vol}(S_\ell) = \min \{ \text{vol}(S_\ell), \text{vol}(\bar{S}_\ell) \}$ ¹. Visto que $\mathbf{x} \perp D\mathbf{1}$, pelo Lema 2.21, temos

$$\sum_{v_i} x_i^2 d_i = \min_{c \in \mathbb{R}} \sum_{v_i} (x_i - c)^2 d_i \leq \sum_{v_i} (x_i - x_\ell)^2 d_i. \quad (4.3)$$

¹Note que $\ell \neq 0$, pois $\text{vol}(S_1) = \text{vol}(\{v_1\}) = d_1 \leq \text{vol}(\bar{S}_1)$.

Definimos a parte positiva e a parte negativa da diferença entre x_i e x_ℓ , denotadas por \mathbf{g} e \mathbf{h} respectivamente, como segue:

$$g_i = \begin{cases} x_i - x_\ell & \text{se } v_i \in S_\ell, \\ 0 & \text{se } v_i \in \bar{S}_\ell \end{cases}$$

$$h_i = \begin{cases} x_\ell - x_i & \text{se } v_i \in \bar{S}_\ell, \\ 0 & \text{se } v_i \in S_\ell \end{cases}.$$

Da equação (4.3), como $w_{ij} > 0$ para todo $\{v_i, v_j\} \in E$, temos

$$\lambda_2^{(\mathcal{L})} = \frac{\sum_{\{v_i, v_j\} \in E} w_{ij} (x_i - x_j)^2}{\sum_{v_i} x_i^2 d_i} \geq \frac{\sum_{\{v_i, v_j\} \in E} w_{ij} (x_i - x_\ell)^2}{\sum_{v_i} (x_i - x_\ell)^2 d_i}.$$

No numerador, é possível verificar que

$$\begin{aligned} \sum_{\{v_i, v_j\} \in E} w_{ij} (x_i - x_j)^2 &= \sum_{\{v_i, v_j\} \in E} w_{ij} (x_i - x_\ell - x_j + x_\ell)^2 \\ &\geq \sum_{\{v_i, v_j\} \in E} w_{ij} [(g_i - g_j)^2 + (h_j - h_i)]^2. \end{aligned}$$

Enquanto no denominador, temos

$$\sum_{v_i} (x_i - x_\ell)^2 d_i = \sum_{v_i} ((x_i - x_\ell)^2 + 0) d_i = \sum_{v_i} (g_i^2 + h_i^2) d_i.$$

Portanto

$$\begin{aligned} \lambda_2^{(\mathcal{L})} &\geq \frac{\sum_{\{v_i, v_j\} \in E} (w_{ij} (g_i - g_j)^2 + w_{ij} (h_j - h_i)^2)}{\sum_{v_i} (g_i^2 + h_i^2) d_i} \\ &= \frac{\sum_{\{v_i, v_j\} \in E} w_{ij} (g_i - g_j)^2 + \sum_{\{v_i, v_j\} \in E} w_{ij} (h_j - h_i)^2}{\sum_{v_i} g_i^2 d_i + \sum_{v_i} h_i^2 d_i}. \end{aligned}$$

Sejam

$$\begin{aligned} \mathcal{R}(\mathbf{g}) &= \frac{\sum_{\{v_i, v_j\} \in E} w_{ij} (g_i - g_j)^2}{\sum_{v_i} g_i^2 d_i} \\ \mathcal{R}(\mathbf{h}) &= \frac{\sum_{\{v_i, v_j\} \in E} w_{ij} (h_j - h_i)^2}{\sum_{v_i} h_i^2 d_i}. \end{aligned}$$

Sem perda de generalidade, consideramos $\lambda_2^{(\mathcal{L})} \geq \mathcal{R}(\mathbf{g})$ quando $0 < \mathcal{R}(\mathbf{g}) \leq \mathcal{R}(\mathbf{h})$ (pelo Lema 2.19) ou quando $h_i = 0$, para todo $i = 1, \dots, n$.

Multiplicando numerador e denominador de $\mathcal{R}(\mathbf{g})$ por $\sum_{\{v_i, v_j\} \in E} w_{ij}(g_i + g_j)^2$, temos

$$\begin{aligned} \lambda_2^{(\mathcal{L})} &\geq \frac{\sum_{\{v_i, v_j\} \in E} w_{ij}(g_i - g_j)^2}{\sum_{v_i} g_i^2 d_i} \\ &= \frac{\left(\sum_{\{v_i, v_j\} \in E} w_{ij}(g_i - g_j)^2\right) \left(\sum_{\{v_i, v_j\} \in E} w_{ij}(g_i + g_j)^2\right)}{\sum_{v_i} g_i^2 d_i \sum_{\{v_i, v_j\} \in E} w_{ij}(g_i + g_j)^2}. \end{aligned} \quad (4.4)$$

Pela desigualdade de Cauchy-Schwarz (Teorema 2.17), temos que o numerador em (4.4) é

$$\begin{aligned} \sum_{\{v_i, v_j\} \in E} w_{ij}(g_i - g_j)^2 \sum_{\{v_i, v_j\} \in E} w_{ij}(g_i + g_j)^2 &\geq \left(\sum_{\{v_i, v_j\} \in E} w_{ij}(g_i - g_j)(g_i + g_j) \right)^2 \\ &= \left(\sum_{\{v_i, v_j\} \in E} w_{ij}(g_i^2 - g_j^2) \right)^2 \end{aligned}$$

Quanto ao denominador, utilizamos o Teorema 2.22 para concluir que

$$2 \left(\sum_{v_i} g_i^2 d_i \right) \geq \sum_{\{v_i, v_j\} \in E} w_{ij}(g_i + g_j)^2.$$

Portanto

$$\lambda_2^{(\mathcal{L})} \geq \frac{\left(\sum_{\{v_i, v_j\} \in E} w_{ij}(g_i^2 - g_j^2)\right)^2}{2 \left(\sum_{v_i} g_i^2 d_i\right)^2}.$$

Podemos reescrever o numerador como

$$\left(\sum_{\{v_i, v_j\} \in E} w_{ij} \underbrace{(g_i^2 - g_{i+1}^2 + g_{i+1}^2 - g_{i+2}^2 + \dots + g_{j-1}^2 - g_j^2)}_{g_r^2 - g_{r+1}^2} \right)^2.$$

Dado $r \in \{1, \dots, n-1\}$, observe que no segundo fator da soma acima o termo $g_r^2 - g_{r+1}^2$ aparece para toda a aresta $\{v_i, v_j\}$ tal que $i \leq r$ e $r+1 \leq j$, isto é, $v_i \in S_r$ e $v_j \in \bar{S}_r$.

Por outro lado, os pesos que aparecem multiplicando $g_r^2 - g_{r+1}^2$ são todos aqueles que estão no corte entre S_r e \bar{S}_r . Logo, podemos afirmar que

$$\left(\sum_{\{v_i, v_j\} \in E} w_{ij} (g_i^2 - g_j^2) \right)^2 = \left(\sum_{r=1}^{n-1} (g_r^2 - g_{r+1}^2) \cdot \text{Cut}(S_r, \bar{S}_r) \right)^2.$$

$$\text{Assim, } \lambda_2^{(\mathcal{L})} \geq \frac{\left(\sum_{r=1}^{n-1} (g_r^2 - g_{r+1}^2) \cdot \text{Cut}(S_r, \bar{S}_r) \right)^2}{2 \left(\sum_{v_i} g_i^2 d_i \right)^2}.$$

Denotamos $\text{vol}_m(S) = \min \{ \text{vol}(S), \text{vol}(\bar{S}) \}$, de modo que $\text{Cut}(S_i, \bar{S}_i) \geq \alpha_G \cdot \text{vol}_m(S_i)$, lembrando que $\alpha_G = \min_{i=1, \dots, n-1} \frac{\text{Cut}(S_i, \bar{S}_i)}{\text{vol}_m(S_i)}$. Com essa notação, multiplicando e dividindo o numerador por $\text{vol}_m(S_r)$ obtemos

$$\begin{aligned} \left(\sum_{r=1}^{n-1} (g_r^2 - g_{r+1}^2) \cdot \text{Cut}(S_r, \bar{S}_r) \right)^2 &= \left(\sum_{r=1}^{n-1} (g_r^2 - g_{r+1}^2) \cdot \frac{\text{vol}_m(S_r)}{\text{vol}_m(S_r)} \cdot \text{Cut}(S_r, \bar{S}_r) \right)^2 \\ &\geq \left(\sum_{r=1}^{n-1} (g_r^2 - g_{r+1}^2) \cdot \text{vol}_m(S_r) \cdot \alpha_G \right)^2. \end{aligned} \quad (4.5)$$

É possível verificar que $g_r^2 - g_{r+1}^2 \geq 0$ para qualquer $r = 1, \dots, n-1$ e, por isso, a desigualdade em (4.5) é atingida. Considerando $S_0 = \emptyset$ e $S_n = V$, temos

$$\begin{aligned} \lambda_2^{(\mathcal{L})} &\geq \frac{\alpha_G^2}{2} \cdot \frac{\left(\sum_{r=1}^{n-1} (g_r^2 - g_{r+1}^2) \cdot \text{vol}_m(S_r) \right)^2}{\left(\sum_{v_i} g_i^2 d_i \right)^2} \\ &= \frac{\alpha_G^2}{2} \cdot \frac{\left(\sum_{r=1}^{n-1} (g_r^2 \cdot \text{vol}_m(S_r) - g_{r+1}^2 \cdot \text{vol}_m(S_{r+1})) \right)^2}{\left(\sum_{v_i} g_i^2 d_i \right)^2} \\ &= \frac{\alpha_G^2}{2} \cdot \frac{\left(\sum_{r=1}^{n-1} g_r^2 \cdot (\text{vol}_m(S_r) - \text{vol}_m(S_{r-1})) \right)^2}{\left(\sum_{v_i} g_i^2 d_i \right)^2}. \end{aligned} \quad (4.6)$$

O numerador da segunda fração em (4.6) também pode ser escrito como

$$\left(\sum_{r=1}^{\ell} g_r^2 \cdot (\text{vol}_m(S_r) - \text{vol}_m(S_{r-1})) + \sum_{r=\ell+1}^{n-1} g_r^2 \cdot (\text{vol}_m(S_r) - \text{vol}_m(S_{r-1})) \right)^2.$$

Note que a soma da direita é igual a zero, pois $g_r = 0$ para todo $r \geq \ell+1$. Por outro lado, pela forma como foi definido ℓ , podemos afirmar que $\text{vol}_m(S_r) = \text{vol}(S_r)$ para

todo $r \leq \ell$. Por essa razão, não iremos alterar o valor em (4.6) se considerarmos os próprios volumes dos S_r , mesmo se $r \geq \ell$, e, além disso, podemos estender o somatório até n , pois $g_n = 0$. Prosseguindo com a desigualdade, temos

$$\begin{aligned} \lambda_2^{(\mathcal{L})} &\geq \frac{\alpha_G^2}{2} \cdot \frac{\left(\sum_{r=1}^n g_r^2 \cdot (\text{vol}(S_r) - \text{vol}(S_{r-1})) \right)^2}{\left(\sum_{v_i} g_i^2 d_i \right)^2} \\ &= \frac{\alpha_G^2}{2} \cdot \frac{\left(\sum_{r=1}^n g_r^2 d_r \right)^2}{\left(\sum_{i=1}^n g_i^2 d_i \right)^2} = \frac{\alpha_G^2}{2} \\ &\geq \frac{\phi_G(2)^2}{2}. \end{aligned}$$

Este resultado também vale se $\lambda_2^{(\mathcal{L})} \geq \mathcal{R}(\mathbf{h})$ quando $0 < \mathcal{R}(\mathbf{h}) < \mathcal{R}(\mathbf{g})$ ou quando, para qualquer $i = 1, \dots, n$, $g_i = 0$. A prova possui algumas pequenas alterações em relação a esta que foi demonstrada, como por exemplo o numerador da segunda fração em (4.6) ser $\left(\sum_{r=1}^n n_r^2 \cdot (\text{vol}(\overline{S}_{r-1}) - \text{vol}(\overline{S}_r)) \right)^2$. Note que não há como $g_i = h_i = 0$ para todo i , pois isto implicaria que \mathbf{x} fosse um vetor constante, o que contradiz o fato de que $\mathbf{x} \perp D\mathbf{1}$. \square

4.2.2 Algoritmo fornecido pela demonstração

Podemos observar que a prova de que vale a desigualdade $\phi_G(2) \leq \sqrt{2\lambda_2^{(\mathcal{L})}}$ gera uma partição $\{S_i, \overline{S}_i\}$ que satisfaz $\frac{\text{Cut}(S_i, \overline{S}_i)}{\min\{\text{vol}(S_i), \text{vol}(\overline{S}_i)\}} \leq \sqrt{2\lambda_2^{(\mathcal{L})}}$. Além disso, pelo item (3) da Proposição 2.15, o vetor \mathbf{x} que satisfaz (4.1) é autovetor associado a $\lambda_2^{(\mathcal{L})}$. A garantia trazida pela cota superior e o fato de ser utilizado na demonstração um autovetor para gerar uma partição cuja condutância respeite essa cota sugere que seja possível definir um algoritmo espectral a partir da demonstração que tenha bons resultados para resolver o problema

$$\phi_G(2) = \min_{S \neq \emptyset \subset V} \frac{\text{Cut}(S, \overline{S})}{\min\{\text{vol}(S), \text{vol}(\overline{S})\}}.$$

A seguir vejamos passo a passo o funcionamento deste algoritmo.

Algoritmo 1: Algoritmo para Condutância de G

Input: Vértices v_1, \dots, v_n e matrizes D e \mathcal{L} de G

Output: Conjuntos S e \bar{S}

- 1 Calcular o autovetor $\mathbf{u}^{(2)}$ associado a $\lambda_2^{(\mathcal{L})}$;
 - 2 $\mathbf{x} = D^{-1/2}\mathbf{u}^{(2)}$;
 - 3 Para cada $i = 1, \dots, n$, definir a função de permutação
 $\sigma(i) = \{j : x_j \text{ é a } i\text{-ésima maior coordenada de } \mathbf{x}\}$;
 - 4 $S_0 = \emptyset$;
 - 5 **for** $i < n$ **do**
 - 6 $S_i = S_{i-1} \cup \{v_{\sigma(i)}\}$;
 - 7 $\alpha_G(S_i) = \frac{\text{Cut}(S_i, \bar{S}_i)}{\min\{\text{vol}(S_i), \text{vol}(\bar{S}_i)\}}$;
 - 8 **end**
 - 9 $S = \{S_i : \alpha_G(S_i) = \min_{j=1, \dots, n-1} \alpha_G(S_j)\}$.
-

Vejamos os resultados obtidos com a aplicação deste algoritmo nos exemplos de grafos a seguir.

Seja $G = (V, E)$ o grafo da Figura 4.1. Um possível autovetor associado ao segundo menor autovalor da matriz laplaciana normalizada \mathcal{L} de G é

$$\mathbf{u}^{(2)} = (0.032, 0.030, -0.248, -0.306, 0.025, 0.011, -0.167, -0.321, -0.321, 0.032, \\ 0.030, -0.248, -0.306, 0.174, 0.247, 0.247, 0.273, 0.273, 0.273, 0.273).$$

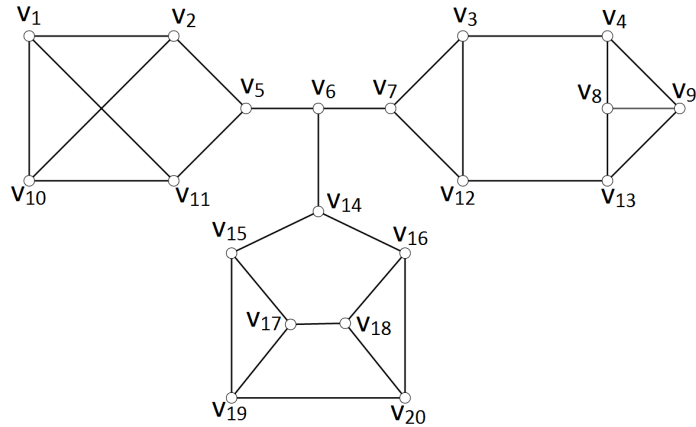


Figura 4.1: Primeiro Exemplo de Grafo

Como G é um grafo 3-regular, temos que $\mathbf{x} = D^{-1/2}\mathbf{u}^{(2)} = \frac{1}{\sqrt{3}} \cdot \mathbf{u}^{(2)}$. Portanto, neste exemplo podemos simplesmente observar o ordenamento das coordenadas de $\mathbf{u}^{(2)}$, ao invés das de \mathbf{x} . Ordenando as coordenadas de \mathbf{u} temos

$$\begin{aligned}
 u_{20}^{(2)} &\geq u_{19}^{(2)} \geq u_{18}^{(2)} \geq u_{17}^{(2)} \geq u_{16}^{(2)} \geq u_{15}^{(2)} \geq u_{14}^{(2)} \geq u_{10}^{(2)} \geq u_1^{(2)} \geq u_{11}^{(2)} \geq u_2^{(2)} \geq u_5^{(2)} \\
 &\geq u_6^{(2)} \geq u_7^{(2)} \geq u_{12}^{(2)} \geq u_3^{(2)} \geq u_{13}^{(2)} \geq u_4^{(2)} \geq u_9^{(2)} \geq u_8^{(2)}.
 \end{aligned}$$

Mostramos na Tabela 4.1 os conjuntos S_i definidos pelo algoritmo e seus respectivos valores de $\alpha_G(S_i)$.

Tabela 4.1: Valores aproximados dos $\alpha_G(S_i)$

Conjunto S_i	$\alpha_G(S_i)$	Conjunto S_i	$\alpha_G(S_i)$
$S_1 = \{v_{20}\}$	1,000	$S_{11} = S_{10} \cup \{v_2\}$	0,111
$S_2 = S_1 \cup \{v_{19}\}$	0,667	$S_{12} = S_{11} \cup \{v_5\}$	0,083
$S_3 = S_2 \cup \{v_{18}\}$	0,555	$S_{13} = S_{12} \cup \{v_6\}$	0,048
$S_4 = S_3 \cup \{v_{17}\}$	0,333	$S_{14} = S_{13} \cup \{v_7\}$	0,111
$S_5 = S_4 \cup \{v_{16}\}$	0,200	$S_{15} = S_{14} \cup \{v_{12}\}$	0,200
$S_6 = S_5 \cup \{v_{15}\}$	0,111	$S_{16} = S_{15} \cup \{v_3\}$	0,167
$S_7 = S_6 \cup \{v_{14}\}$	0,048	$S_{17} = S_{16} \cup \{v_{13}\}$	0,333
$S_8 = S_7 \cup \{v_{10}\}$	0,167	$S_{18} = S_{17} \cup \{v_4\}$	0,667
$S_9 = S_8 \cup \{v_1\}$	0,185	$S_{19} = S_{18} \cup \{v_9\}$	1,000
$S_{10} = S_9 \cup \{v_{11}\}$	0,133		

Podemos observar que os S_i que obtiveram melhores resultados foram

$$S_7 = \{v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}\}$$

e

$$S_{13} = \{v_1, v_2, v_5, v_6, v_{10}, v_{11}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}\}.$$

Vejamos como estão representadas as partições $\{S_7, \overline{S_7}\}$ e $\{S_{13}, \overline{S_{13}}\}$, que são os dois possíveis resultados deste algoritmo, nas figuras 4.2a e 4.2b, respectivamente.

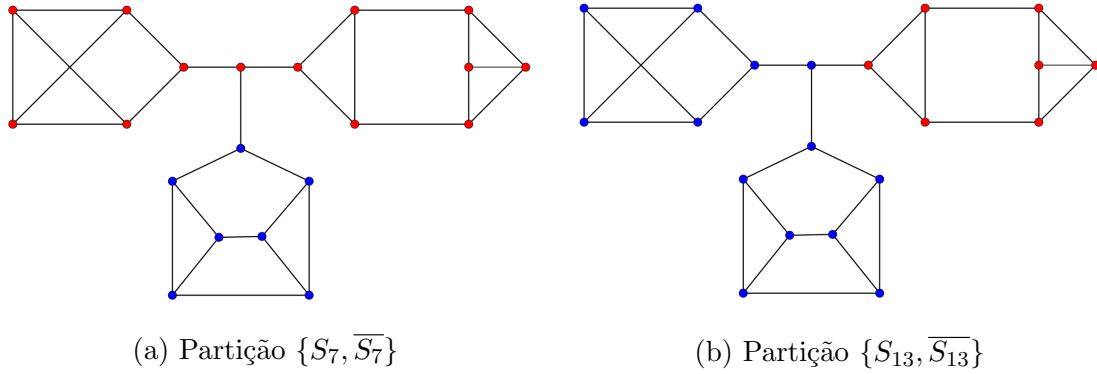


Figura 4.2: Partições que podem ser a saída do algoritmo

Sabendo que, como G é regular, $\phi_G(2) = \min_{|S| \leq 10} \frac{\text{Cut}(S, \overline{S})}{3 \cdot |S|}$, para provar que as partições das figuras 4.2a e 4.2b de fato geram $\phi_G(2)$ mostraremos que não há como obter S diferente de S_7 e S_{13} , com $|S| \leq 10$, tal que $\frac{\text{Cut}(S, \overline{S})}{3 \cdot |S|} < \frac{\text{Cut}(S_7, \overline{S_7})}{3 \cdot |S_7|} = \frac{\text{Cut}(S_{14}, \overline{S_{14}})}{3 \cdot |S_{14}|} = \frac{1}{18}$. Inicialmente, suponha que $\text{Cut}(S, \overline{S}) = 1$ e $\{u, v\}$ é a aresta de corte de S com seu complementar. É possível verificar que $\{u, v\}$ não está em nenhum um ciclo de G , caso contrário $\text{Cut}(S, \overline{S}) > 1$. Observando a Figura 4.1, notamos que as únicas arestas que satisfazem essa condição são $\{v_5, v_6\}$, $\{v_6, v_7\}$ e $\{v_6, v_{14}\}$, estas duas últimas são cortes que separam as classes das partições $\{S_7, \overline{S_7}\}$ e $\{S_{14}, \overline{S_{14}}\}$, respectivamente. No caso da aresta $\{v_5, v_6\}$, ela separa duas classes cuja partição tem condutância igual a $\frac{1}{15} > \frac{1}{18}$. Para S tal que $\text{Cut}(S, \overline{S}) > 1$, seria necessário que $|S| > 13$ para que $\frac{\text{Cut}(S, \overline{S})}{3 \cdot |S|} < \frac{1}{18}$, o que contradiz o fato de $|S| \leq 10$. Logo, as partições das figuras 4.2a e 4.2b são as que geram $\phi_G(2)$.

Antes de observar o que ocorre quando aplicamos o algoritmo no grafo de Petersen, definido por $P = (V, E)$ e representado na Figura 4.3a, iremos verificar que uma possível partição que gera $\phi_P(2) = \frac{1}{3}$ é a que está representada na Figura 4.3b.

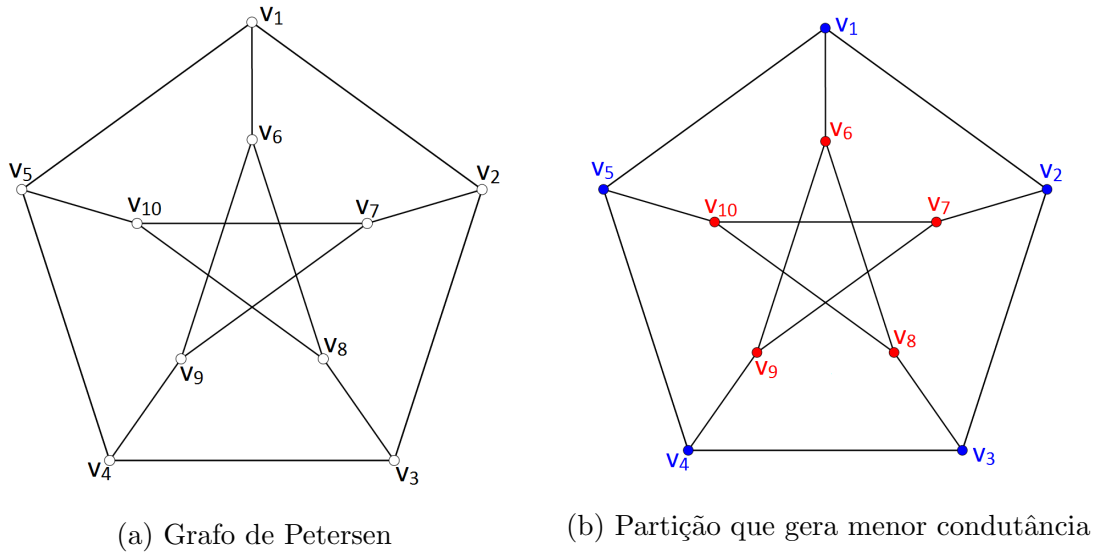


Figura 4.3: Segundo Exemplo de Grafo

Como P é regular, temos que $\phi_P(2) = \min_{|S| \leq 5} \phi_P(S)$. Nosso objetivo, portanto, é concluir que para qualquer S tal que $1 \leq |S| \leq 5$ tem-se que $\phi_P(S) \geq \frac{1}{3}$.

Sabemos que o grafo de Petersen não contém ciclo de comprimento menor que 5. Logo temos que, para S tal que $1 \leq |S| < 5$, $G[S]$ é uma floresta. Isto implica que a quantidade máxima que S pode possuir de arestas internas é $2(|S| - 1)$. Por essa razão, podemos afirmar que $\text{Cut}(S, \bar{S}) \geq 3|S| - 2(|S| - 1) = |S| + 2$ e, portanto,

$$\phi_G(S) = \frac{\text{Cut}(S, \bar{S})}{\text{vol}(S)} \geq \frac{|S| + 2}{3|S|} \geq \frac{6}{12} > \frac{1}{3}.$$

Caso $|S| = 5$ e $G[S]$ acíclico, temos que $\phi_G(S) \geq \frac{7}{15} > \frac{1}{3}$. Se $G[S]$ é um ciclo, cada vértice se liga a dois vértices de S e, portanto, se liga a um vértice que não esteja em S . Logo, nesta situação temos

$$\phi_G(S) = \frac{|S|}{3|S|} = \frac{1}{3},$$

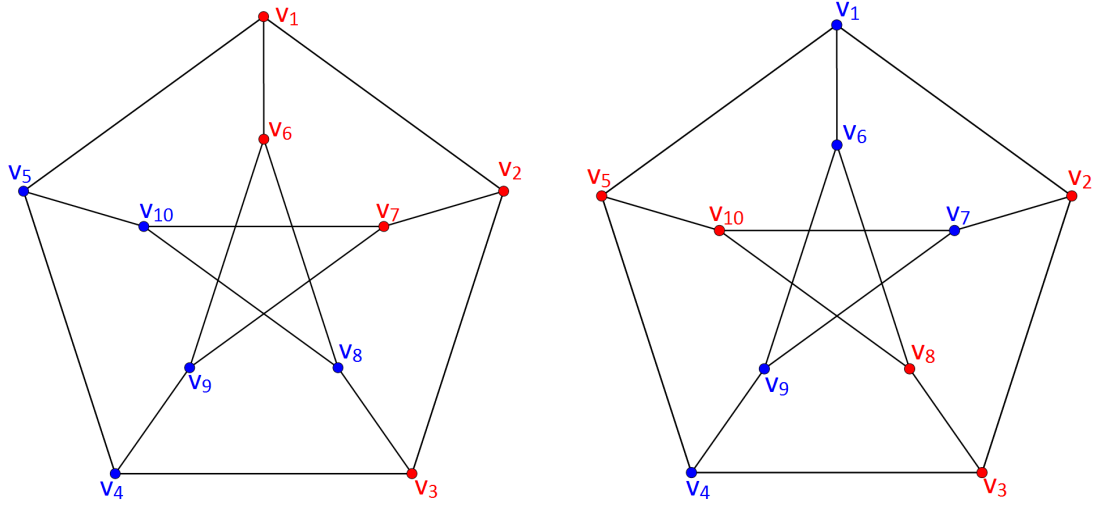
confirmando que a partição da Figura 4.3b gera $\phi_P(2)$.

Agora aplicaremos o Algoritmo para Condutância de P . Assim como o grafo do exemplo anterior, P também é regular e, por isso, podemos simplesmente analisar o ordenamento das coordenadas do autovetor associado a $\lambda_2^{(\mathcal{L})}$ para proceder com o algoritmo. Sabendo que $\text{spect}_{\mathcal{L}}(P) = \{0, 0.67^{(5)}, 1.67^{(4)}\}$, é possível obter os autovetores da associados aos autovalores $\lambda_2^{(\mathcal{L})} = \lambda_3^{(\mathcal{L})} = \lambda_4^{(\mathcal{L})} = \lambda_5^{(\mathcal{L})} = \lambda_6^{(\mathcal{L})} = 0.67$ conforme Tabela 4.2.

Tabela 4.2: Autovetores de \mathcal{L}

Vetor	Coordenadas									
	1	2	3	4	5	6	7	8	9	10
$\mathbf{u}^{(2)}$	0.71	0.24	-0.24	-0.24	0.24	0.24	-0.24	-0.24	-0.24	-0.24
$\mathbf{u}^{(3)}$	0.15	0.61	0.37	-0.25	-0.48	0.03	0.08	0.01	-0.13	-0.39
$\mathbf{u}^{(4)}$	0.20	0.53	0.43	0.19	-0.17	-0.16	-0.10	-0.29	-0.07	-0.56
$\mathbf{u}^{(5)}$	-0.05	0.12	0.61	0.17	-0.28	0.12	-0.45	0.33	-0.17	-0.40
$\mathbf{u}^{(6)}$	-0.04	0.15	0.37	-0.04	0.23	-0.43	-0.17	0.25	-0.64	0.32

Como o grafo de Petersen possui muitas simetrias, o autovalor $\lambda_2^{(\mathcal{L})} = 0.67$ tem multiplicidade maior que um, o que implica que a dimensão do seu autoespaço associado também é maior que um. Por essa razão, é possível que tenhamos diferentes partições como respostas do algoritmo, dependendo de qual autovetor do autoespaço associado a $\lambda_2^{(\mathcal{L})}$ utilizamos. Nas Figuras 4.4a e 4.4b, por exemplo, estão as partições que seriam encontradas pelo algoritmo se considerássemos $\mathbf{u}^{(3)}$ e $\mathbf{u}^{(6)}$, respectivamente, como autovetores associados a $\lambda_2^{(\mathcal{L})}$; em ambas as situações, a partição encontrada tem valor de condutância igual a $\frac{7}{15} > \phi_P(2)$.



(a) Partição encontrada ao considerar $\mathbf{u}^{(3)}$ (b) Partição encontrada ao considerar $\mathbf{u}^{(6)}$

Figura 4.4: Diferentes Partições de P

É possível notar que os outros autovetores da Tabela 4.2, além de $\mathbf{u}^{(3)}$ e $\mathbf{u}^{(6)}$, também não geram uma partição cuja condutância seja igual a $\phi_G(2)$. Porém sabemos que todo $\mathbf{y} \in \text{Span}\{\mathbf{u}^{(2)}, \mathbf{u}^{(3)}, \mathbf{u}^{(4)}, \mathbf{u}^{(5)}, \mathbf{u}^{(6)}\}$ não nulo também é autovetor associado a $\lambda_2^{(\mathcal{L})}$, o que nos leva a questionar se há vetores no autoespaço que, utilizados no algoritmo, acabam gerando a partição cuja condutância seja igual a $\phi_G(2)$. De fato, o vetor

$$\begin{aligned} \mathbf{y} &= 1.5\mathbf{u}^{(2)} + 0.25\mathbf{u}^{(3)} + 1.5\mathbf{u}^{(4)} + \mathbf{u}^{(5)} + \mathbf{u}^{(6)} \\ &= (1.30, 1.56, 1.36, -0.00, -0.07, -0.19, -1.11, -0.20, -1.29, -1.37) \end{aligned}$$

além de estar no autoespaço também gera exatamente a partição da Figura 4.3b, mostrando que o algoritmo fornece a resposta correta para o grafo de Petersen se for escolhido um autovetor apropriado.

Agora veremos o que ocorre quando aplicamos o algoritmo para um grafo $G = (V, E)$ irregular e ponderado, conforme representado na Figura 4.5a e cujos pesos são dados na Tabela 4.3. Testando todas as condutâncias das partições do conjunto de vértices deste grafo em k classes, é possível verificar que a que está

representada na Figura 4.5b é a que gera $\phi_G(2) = 0.439$. O peso das arestas são diferenciados nas figuras conforme as suas cores e tracejados: quanto mais próxima a tonalidade da linha que representa a aresta é do preto e o comprimento dos seus traços espaçados forem maiores, maior é o peso da aresta, enquanto que quanto mais próximo do branco e menores forem os traços, menor é o peso.

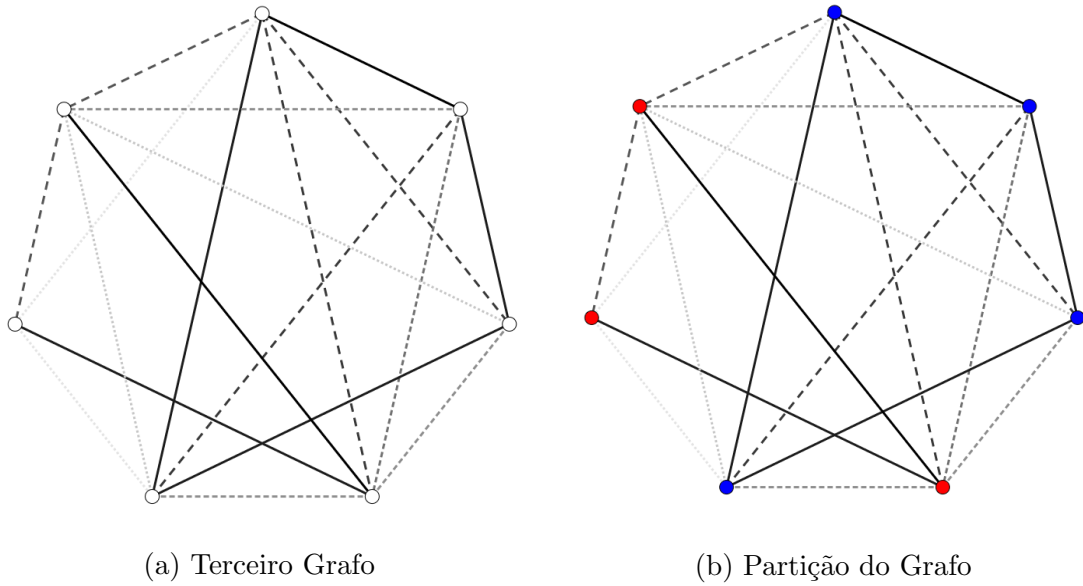


Figura 4.5: Exemplo de grafo e sua respectiva partição gerada pelo algoritmo, que também gera a condutância de G

Tabela 4.3: Pesos das Arestas e Grau dos Vértices de G

$w_{AB} = 0.9$	$w_{AB} = 0.9$	$w_{AC} = 0.7$	$w_{AD} = 0.7$	$w_{AE} = 0.8$	$w_{AF} = 0.1$	$w_{AG} = 0.6$
$w_{AC} = 0.7$	$w_{BC} = 0.8$	$w_{BC} = 0.8$	$w_{BD} = 0.5$	$w_{BE} = 0.7$	$w_{BF} = 0.0$	$w_{BG} = 0.4$
$w_{AD} = 0.7$	$w_{BD} = 0.5$	$w_{CD} = 0.4$	$w_{CD} = 0.4$	$w_{CE} = 0.8$	$w_{CF} = 0.0$	$w_{CG} = 0.2$
$w_{AE} = 0.8$	$w_{BE} = 0.7$	$w_{CE} = 0.8$	$w_{DE} = 0.4$	$w_{DE} = 0.4$	$w_{DF} = 0.8$	$w_{DG} = 0.9$
$w_{AF} = 0.1$	$w_{BF} = 0.0$	$w_{CF} = 0.0$	$w_{DF} = 0.8$	$w_{EF} = 0.1$	$w_{EF} = 0.1$	$w_{EG} = 0.2$
$w_{AG} = 0.6$	$w_{BG} = 0.4$	$w_{CG} = 0.2$	$w_{DG} = 0.9$	$w_{EG} = 0.2$	$w_{FG} = 0.6$	$w_{FG} = 0.6$
$d_A = 3.8$	$d_B = 3.3$	$d_C = 2.9$	$d_D = 3.7$	$d_E = 3.0$	$d_F = 1.6$	$d_G = 2.9$

Para a matriz laplaciana normalizada deste grafo, encontramos o auto-vetor

$$\mathbf{u}^{(2)} = (0.019, 0.308, 0.371, -0.373, 0.321, -0.579, -0.393)$$

associado a $\lambda_2^{(\mathcal{L})}$. Além disso, o vetor $\mathbf{x} = D^{-1/2}\mathbf{u}^{(2)}$ é

$$\mathbf{x} = (0.097, 0.170, 0.218, -0.194, 0.186, -0.458, -0.231).$$

A ordenação de suas coordenadas é $x_3 \geq x_5 \geq x_2 \geq x_1 \geq x_4 \geq x_7 \geq x_6$. Logo temos que os S_i e seus respectivos $\alpha_G(S_i)$ são dados conforme a Tabela 4.4.

Tabela 4.4: Resultados dos $\alpha_G(S_i)$

Conjunto S_i	$\alpha_G(S_i)$	-	Conjunto S_i	$\alpha_G(S_i)$	-	Conjunto S_i	$\alpha_G(S_i)$
$S_1 = \{C\}$	1,000	-	$S_3 = S_2 \cup \{B\}$	0,500	-	$S_5 = S_4 \cup \{D\}$	0,733
$S_2 = S_1 \cup \{E\}$	0,729	-	$S_4 = S_3 \cup \{A\}$	0,439	-	$S_6 = S_5 \cup \{G\}$	1,000

Neste exemplo, o algoritmo retornou como resposta o conjunto $S_4 = \{A, B, C, E\}$ e, portanto, obteve a partição da Figura 4.5b que gera $\phi_G(2)$, mostrando que para grafos irregulares e ponderados ele também pode ter um bom desempenho.

4.3 Desigualdade de Cheeger de Ordem k

Versões da Desigualdade de Cheeger de ordens superiores a 2 podem ser enunciadas conforme o seguinte teorema.

Teorema 4.5 (Desigualdade de Cheeger de Ordem k). *Seja $G = (V, E)$ e \mathcal{L} sua matriz laplaciana normalizada, cujos autovalores são $\lambda_1^{(\mathcal{L})} \leq \dots \leq \lambda_n^{(\mathcal{L})}$. Dado um $k \geq 2$ inteiro, vale*

$$\frac{\lambda_k^{(\mathcal{L})}}{2} \leq \phi_G(k) \leq C \cdot f(k) \cdot \sqrt{\lambda_k^{(\mathcal{L})}}, \quad (4.7)$$

onde $C > 0$ é uma constante absoluta e $f : \{2, \dots, n-1\} \rightarrow \mathbb{R}$ é uma função.

Neste trabalho, será apresentada a prova, demonstrada por Trevisan [33], de que $f(k) = k^{3.5}$ em (4.7). No entanto, vale ressaltar que são conhecidas cotas superiores mais fortes do que esta cujas provas são de mais difícil compreensão, como é o caso de quando $f(k) = k^2$, conforme Lee, Gharan e Trevisan [22].

Para as provas das cotas superior e inferior de (4.7) consideraremos por simplicidade um grafo G d -regular e não ponderado, mas o resultado pode ser estendido para grafos irregulares e com pesos, conforme Trevisan [33].

4.3.1 Demonstração para a Cota Inferior

O lema a seguir prova de forma direta a desigualdade $\frac{\lambda_k^{(\mathcal{L})}}{2} \leq \phi_G(k)$.

Lema 4.6. *Sejam $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}$ vetores de suporte disjunto e \mathcal{L} a matriz laplaciana normalizada de um grafo não ponderado G d -regular. Então*

$$R_{\mathcal{L}} \left(\sum_{i=1}^k \mathbf{x}^{(i)} \right) \leq 2 \cdot \max_{i=1, \dots, k} R_{\mathcal{L}}(\mathbf{x}^{(i)}).$$

Demonstração. Primeiro, é fácil verificar que $\|\sum_i \mathbf{x}^{(i)}\|^2 = \sum_i \|\mathbf{x}^{(i)}\|^2$. Queremos mostrar que, para cada aresta $\{u, v\}$, vale

$$\left(\sum_{i=1}^k x_u^{(i)} - x_v^{(i)} \right)^2 \leq 2 \sum_{i=1}^k (x_u^{(i)} - x_v^{(i)})^2.$$

Pela definição de vetores com suporte disjunto, existem j tal que $x_u^{(i)} = 0$ para $i \neq j$ e ℓ tal que $x_v^{(i)} = 0$ para $i \neq \ell$. Se $j = \ell$, então

$$\left(\sum_{i=1}^k x_u^{(i)} - x_v^{(i)} \right)^2 = (x_u^{(j)} - x_v^{(j)})^2 = \sum_{i=1}^k (x_u^{(i)} - x_v^{(i)})^2$$

e, se $j \neq \ell$, então

$$\left(\sum_{i=1}^k x_u^{(i)} - x_v^{(i)} \right)^2 = (x_u^{(j)} - x_v^{(\ell)})^2 \leq 2(x_u^{(j)})^2 + 2(x_v^{(\ell)})^2 = 2 \sum_{i=1}^k (x_u^{(i)} - x_v^{(i)})^2.$$

Com isso, temos que

$$\begin{aligned}
R_{\mathcal{L}}\left(\sum_{i=1}^k \mathbf{x}^{(i)}\right) &= \frac{\sum_{\{u,v\} \in E} \left(\sum_{i=1}^k x_u^{(i)} - x_v^{(i)}\right)^2}{d \|\sum_i \mathbf{x}^{(i)}\|^2} \leq 2 \cdot \frac{\sum_{i=1}^k \sum_{\{u,v\} \in E} \left(x_u^{(i)} - x_v^{(i)}\right)^2}{d \sum_i \|\mathbf{x}^{(i)}\|^2} \\
&\leq 2 \cdot \max_{i=1, \dots, k} R_{\mathcal{L}}(\mathbf{x}^{(i)}), \tag{4.8}
\end{aligned}$$

onde em (4.8) foi utilizado o Lema 2.19. \square

Sejam S_1, \dots, S_k os conjuntos que satisfazem $\phi_G(S_i) \leq \phi_G(k)$, para qualquer $i = 1, \dots, k$, e \mathbf{X} o conjunto de todas as combinações lineares dos vetores $\mathbf{1}_{S_i}$. Primeiro note que, para qualquer escolha de $\alpha_i \in \mathbb{R}$ os vetores $\alpha_i \mathbf{1}_{S_i}$ são de suporte disjunto e, por isso, satisfazem o Lema 4.6. Além disso, é possível verificar que, utilizando o item (2) da Proposição 2.15, temos $R_{\mathcal{L}}(\alpha_i \mathbf{1}_{S_i}) = \frac{\text{Cut}(S_i, \overline{S_i})}{\text{vol}(S_i)} = \phi_G(S_i)$. Portanto

$$2 \cdot \max_{i=1, \dots, k} R_{\mathcal{L}}(\alpha_i \mathbf{1}_{S_i}) = 2 \cdot \max_{i=1, \dots, k} \phi_G(S_i) \leq 2 \cdot \phi_G(k).$$

Por outro lado, como $\mathbf{1}_{S_i}$ são ortogonais entre si, então \mathbf{X} tem dimensão k . Logo, pelo Teorema 2.9, temos que para $\mathbf{x} = \alpha_1 \mathbf{1}_{S_1} + \dots + \alpha_k \mathbf{1}_{S_k}$, com $\alpha_i \neq 0$, que satisfaz $\max_{\mathbf{y} \in \mathbf{X} \text{ e } \mathbf{y} \neq 0} R_{\mathcal{L}}(\mathbf{y}) = R_{\mathcal{L}}(\mathbf{x})$ vale

$$\begin{aligned}
\lambda_k^{(\mathcal{L})} &= \min_{W \in \mathcal{W}_k} \max_{\mathbf{y} \in W \text{ e } \mathbf{y} \neq 0} R_{\mathcal{L}}(\mathbf{y}) \leq R_{\mathcal{L}}(\mathbf{x}) = R_{\mathcal{L}}\left(\sum_{i=1}^k \alpha_i \mathbf{1}_{S_i}\right) \leq 2 \cdot \max_{i=1, \dots, k} R_{\mathcal{L}}(\alpha_{S_i} \mathbf{1}_{S_i}) \\
&\leq 2 \cdot \phi_G(k).
\end{aligned}$$

A prova de que vale a desigualdade $\phi_G(k) \leq C \cdot k^{3.5} \cdot \sqrt{\lambda_k^{(\mathcal{L})}}$ fornece um algoritmo que pode gerar k conjuntos que podem estar próximos daqueles que satisfazem $\phi_G(k)$. Nas subseções seguintes, descreveremos o funcionamento do algoritmo em três níveis. O primeiro deles irá explicar em poucos detalhes quais os procedimentos que serão feitos, explicando brevemente o motivo de utilizá-los. O segundo demonstrará a desigualdade $\phi_G(k) \leq C \cdot k^{3.5} \cdot \sqrt{\lambda_k^{(\mathcal{L})}}$ com definições e lemas que serão

enunciados na subseção. O terceiro definirá o algoritmo que será implementado e ocorrerá sua aplicação para exemplos que serão dados.

4.3.2 Cota Superior - Funcionamento do Algoritmo

Definimos a matriz $U \in \mathbb{R}^{n \times k}$ cujas colunas são os autovetores $\mathbf{u}_1, \dots, \mathbf{u}_k$ ortonormais associados aos k menores autovalores da matriz laplaciana normalizada de G . Tomamos os vetores $\mathbf{f}(i)$ como sendo os vetores linha de U , onde a i -ésima linha possui correspondência com o i -ésimo vértice de G . Definimos a *massa* de um vértice v como o valor de $\|\mathbf{f}(v)\|^2$. Depois, projetamos estes vetores na casca esférica unitária de \mathbb{R}^k ; por simplicidade, chamamos as projeções $\bar{\mathbf{f}}(v)$ obtidas por esse procedimento como *projeções de v* .

A ideia inicial é dividir os vértices em k conjuntos de forma que a distância entre as projeções dos vértices em conjuntos distintos seja pelo menos uma constante $\delta > 0$, a massa de cada conjunto (soma das massas dos vértices que pertencem a ele) seja pelo menos $1/2$ e a massa total (soma das massas de todos os conjuntos) seja pelo menos $k - \frac{1}{4}$.

Para isso, primeiro construiremos uma malha de cubos e seus respectivos núcleos, que são cubos de mesmo centro aos da malha, porém de lados menores aos dela e iguais entre si. Alocaremos em um mesmo conjunto T_i os vértices cujas projeções estejam no mesmo núcleo. Primeiro repare que a existência dos núcleos garantem que a distância entre as projeções dos vértices em conjuntos distintos não seja pequena. Repare também que pode ocorrer que algum vértice não fique em nenhum T_i , pois a sua projeção pode ficar próxima à fronteira do cubo. Caso a massa total seja pequena, moveremos a malha a distâncias aleatórias ao longo de cada coordenada de \mathbb{R}^k até que a massa total dos T_i obtidos seja suficientemente grande. Em seguida iremos unir ou excluir conjuntos até que tenhamos k deles cujas respectivas massas sejam pelo menos $1/2$.

Sejam A_1, \dots, A_k os conjuntos formados ao final desse passo. O algoritmo permitirá incluir aos conjuntos A_i alguns vértices cujas projeções não estavam em nenhum núcleo, mas estavam suficientemente próximos da projeção de algum vértice que esteja em algum desses conjuntos, formando conjuntos S_1, \dots, S_k .

4.3.3 Demonstração para a Cota Superior

Nesta seção será provada a desigualdade $\phi_G(k) \leq C \cdot k^{3.5} \cdot \sqrt{\lambda_k^{(\mathcal{L})}}$ com a utilização de definições e lemas que serão enunciados durante a demonstração. Para cada resultado enunciado, são tomados um grafo $G = (V, E)$ não ponderado d -regular com vértices $1, \dots, n$, sua matriz laplaciana normalizada \mathcal{L} de autovalores $\lambda_1^{(\mathcal{L})} \leq \dots \leq \lambda_n^{(\mathcal{L})}$ e seus respectivos autovetores ortonormais $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(n)}$ e fixado um $k \geq 2$ inteiro.

Definição 4.7. *Sejam $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}$ vetores ortonormais cujas coordenadas são indexadas pelo conjunto de vértices de um grafo G . Definimos a função de mapeamento em termos dos $\mathbf{x}^{(i)}$ como $\mathbf{f} : V \rightarrow \mathbb{R}^k$ que satisfaz*

$$\mathbf{f}(v) = (x_v^{(1)}, \dots, x_v^{(k)}).$$

Denotaremos também

$$\bar{\mathbf{f}}(v) = \frac{\mathbf{f}(v)}{\|\mathbf{f}(v)\|}.$$

Um resultado imediato da Definição 4.7 é que

$$\sum_{v \in V} \|\mathbf{f}(v)\| = \sum_{v \in V} \sum_{i=1}^k (x_v^{(i)})^2 = \sum_{i=1}^k \|\mathbf{x}^{(i)}\|^2 = k.$$

Nesta seção, os vetores ortonormais $\mathbf{x}^{(i)}$ enunciados na Definição 4.7 serão os autovetores $\mathbf{u}^{(i)}$ relacionados aos k menores autovalores de \mathcal{L} .

É importante definir também a distância entre dois vértices e entre um vértice e um conjunto de vértices.

Definição 4.8. Definimos a distância entre os vértices u e v como

$$\text{dist}(u, v) = \|\bar{\mathbf{f}}(u) - \bar{\mathbf{f}}(v)\|.$$

Além disso, dado um subconjunto $A \subset V$, definimos a distância entre u e A como

$$\text{dist}(u, A) = \min_{v \in A} \text{dist}(u, v).$$

Outro conceito que será utilizado em determinado momento é do quociente de Rayleigh de uma função de mapeamento.

Definição 4.9. Definimos o quociente de Rayleigh da função \mathbf{f} como

$$R_{\mathcal{L}}(\mathbf{f}) = \frac{\sum_{\{u,v\} \in E} \|\mathbf{f}(u) - \mathbf{f}(v)\|^2}{\sum_v \|\mathbf{f}(v)\|^2 d_v}.$$

O seguinte lema sobre o quociente de Rayleigh de uma função, demonstrada na Subseção 4.3.5, será importante em um determinado momento nesta subseção.

Lema 4.10. Seja o mapeamento $\mathbf{f} : V \rightarrow \mathbb{R}^k$ definido em termos de k vetores ortonormais $\mathbf{x}^{(i)}$, temos que

$$R_{\mathcal{L}}(\mathbf{f}) \leq \max_{i=1, \dots, k} R_{\mathcal{L}}(\mathbf{x}^{(i)}). \quad (4.9)$$

Em particular, se \mathbf{f} é a função de mapeamento definida em termos dos autovetores ortonormais $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}$ associados aos k menores autovalores da matriz laplaciana normalizada de um grafo $\lambda_1^{(\mathcal{L})} \leq \dots \leq \lambda_k^{(\mathcal{L})}$, então, pelo Teorema 2.7, a desigualdade em (4.9) pode ser reescrita como

$$R_{\mathcal{L}}(\mathbf{f}) \leq \lambda_k^{(\mathcal{L})}. \quad (4.10)$$

Começamos a demonstração da desigualdade $\phi_G(k) \leq C \cdot k^{3.5} \cdot \sqrt{\lambda_k^{(\mathcal{L})}}$ tomando uma malha de cubos em \mathbb{R}^k de lados iguais a L e seus respectivos núcleos.

Cada núcleo é um cubo de mesmo centro que o original e de lado igual a $L \cdot \left(1 - \frac{1}{4k^2}\right)$. Alocaremos em um mesmo conjunto T_i os vértices que estejam num mesmo núcleo e definiremos os R_i como os núcleos que contém pelo menos um dos $\mathbf{f}(v)$; seja m a quantidade de conjuntos T_i e R_i . Caso $\sum_{i=1}^m \sum_{v \in T_i} \|\mathbf{f}(v)\|^2 < k - \frac{1}{4}$, tomamos uma k -upla (c_1, \dots, c_k) onde cada $c_i \in [0, L]$ é um número real aleatório. Em seguida, deslocamos a malha pelo vetor (c_1, \dots, c_k) e repetimos o processo de alocar os vértices e mover a malha até que $\sum_{i=1}^m \sum_{v \in T_i} \|\mathbf{f}(v)\|^2 \geq k - \frac{1}{4}$. Os conjuntos T_1, \dots, T_m obtidos ao final deste processo satisfazem as propriedades do seguinte lema, que será demonstrado na Subseção 4.3.5.

Lema 4.11. *Existem $m \geq k$ conjuntos disjuntos de vértices T_1, \dots, T_m tais que*

$$(1) \sum_{i=1}^m \sum_{v \in T_i} \|\mathbf{f}(v)\|^2 \geq k - \frac{1}{4};$$

$$(2) \text{ Para quaisquer vértices } u, v \text{ em conjuntos diferentes, } \text{dist}(u, v) \geq C \cdot k^{-3}, \text{ onde } C \in \mathbb{R}_{>0};$$

$$(3) \text{ Para qualquer conjunto } T_i, \text{ vale } \sum_{v \in T_i} \|\mathbf{f}(v)\|^2 \leq 1 + \frac{1}{4k}.$$

Até o momento comentamos como obter os conjuntos T_1, \dots, T_m que satisfazem o item (1) deste lema. Para os itens (2) e (3), o seguinte lema, que será demonstrado na Subseção 4.3.5, irá auxiliar a garantir que estes conjuntos satisfaçam estas duas propriedades.

Lema 4.12. *Seja $\text{diam}(R) = \max_{\mathbf{u}, \mathbf{v} \in R} \|\mathbf{u} - \mathbf{v}\|$ e $V(R) = \{v \in V : \bar{\mathbf{f}}(v) \in R\}$. Para todo subconjunto R da casca esférica unitária, temos que*

$$\sum_{v \in V(R)} \|\mathbf{f}(v)\|^2 \leq \left(1 - \frac{1}{2}(\text{diam}(R))^2\right)^{-2}.$$

Definindo o lado dos cubos da malha como $L = \frac{1}{\sqrt{5}k}$, dados $u \in T_i$ e $v \in T_j$, com $i \neq j$, temos que $\text{dist}(u, v) \geq \frac{1}{4\sqrt{5}k^3}$, o que garante o item (2) do Lema

4.11. Além disso, podemos concluir que, para qualquer $i = 1, \dots, m$, $\text{diam}(R_i) \leq \frac{1}{\sqrt{5k}}$.
 Pelo Lema 4.12 temos que

$$\begin{aligned} \sum_{v \in T_i} \|\mathbf{f}(v)\|^2 &\leq \left(1 - \frac{1}{2}(\text{diam}(R_i))^2\right)^{-2} \leq \left(1 - \frac{1}{10k}\right)^{-2} \leq \left(1 - \frac{1}{5k}\right)^{-1} \\ &= 1 + \frac{1}{5k-1} \leq 1 + \frac{1}{4k}, \end{aligned}$$

o que garante a condição do item (3) do Lema 4.11.

Partindo dos conjuntos do Lema 4.11, podemos produzir conjuntos A_1, \dots, A_k que satisfazem as propriedades do lema seguinte, cuja demonstração aparece após seu enunciado.

Lema 4.13 (Conjuntos Bem-Separados). *Existem k conjuntos disjuntos de vértices A_1, \dots, A_k tais que*

(1) *Para qualquer conjunto A_i , vale $\sum_{v \in A_i} \|\mathbf{f}(v)\|^2 \geq \frac{1}{2}$;*

(2) *Para quaisquer vértices u, v em conjuntos diferentes, $\text{dist}(u, v) \geq C \cdot k^{-3}$, onde $C \in \mathbb{R}_{>0}$.*

Demonstração. Sejam T_1, \dots, T_m os conjuntos obtidos através do Lema 4.11. Unimos dois a dois os conjuntos T_i e T_j que tenham massa inferior a $1/2$ e repetimos este processo até que tenhamos no máximo um conjunto cuja massa é inferior a $1/2$; chamamos de A_1, \dots, A_t todos os conjuntos obtidos com massa maior ou igual a $1/2$. Como cada conjunto gerado pela união de dois desses conjuntos tem massa menor que 1, podemos afirmar que cada A_i tem massa menor ou igual a $1 + \frac{1}{4k}$. Com isso, é possível afirmar que

$$\frac{1}{2} + t \cdot \left(1 + \frac{1}{4k}\right) \geq \sum_{i=1}^t \sum_{v \in A_i} \|\mathbf{f}(v)\|^2 \geq k - \frac{1}{4},$$

que implica $\frac{t}{4k} \geq k - t - \frac{3}{4}$. Supondo que $t < k$, isto é, $k - t \geq 1$, a desigualdade anterior implica $\frac{t}{4k} \geq \frac{1}{4}$, o que contradiz o fato de $t < k$. Logo, $t \geq k$.

Para formar os k conjuntos mencionados no item (1) do Lema 4.13, basta tomar k conjuntos destes t , seja excluindo alguns deles ou unindo aqueles já existentes. Note que, com isso, a condição (1) é satisfeita e as operações realizadas nos conjuntos T_i para obter os A_j não mudam o fato de que o item (2) do Lema 4.11, que é o mesmo do 4.13, seja satisfeito. \square

Até este ponto da prova, obtivemos k conjuntos A_1, \dots, A_k que, conforme mencionado no final da subseção anterior, utilizaremos para formar os k conjuntos S_1, \dots, S_k desejados. Isto será feito em dois passos. No primeiro, utilizamos os conjuntos A_1, \dots, A_k para formar k vetores $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)}$ de suporte disjunto. Em seguida, escolhemos os k conjuntos S_1, \dots, S_k onde, para cada $i = 1, \dots, k$, S_i é subconjunto com suporte de $\mathbf{y}^{(i)}$. Para realizar o primeiro passo, utilizaremos o seguinte lema, que será demonstrado na Subseção 4.3.5.

Lema 4.14 (Localização). *Seja $\mathbf{f} : V \rightarrow \mathbb{R}^k$ uma função de mapeamento em termos de k vetores $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}$ ortonormais. Dados k conjuntos de vértices A_1, \dots, A_k de vértices tais que, para todo $i = 1, \dots, k$, $\sum_{v \in A_i} \|\mathbf{f}(v)\|^2 \geq \frac{1}{2}$ e, para quaisquer vértices u, v de conjuntos diferentes, $\text{dist}(u, v) \geq \delta$, podemos construir k vetores $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)} \in \mathbb{R}^n$ de suporte disjunto e entradas não negativas tais que, para qualquer $i = 1, \dots, k$, vale que*

$$R_{\mathcal{L}}(\mathbf{y}^{(i)}) \leq C \cdot k \cdot \delta^{-2} \cdot R_{\mathcal{L}}(\mathbf{f}), \quad (4.11)$$

onde $C \in \mathbb{R}_{>0}$.

Conforme a prova deste lema, os vetores $\mathbf{y}^{(i)}$ que satisfazem a desigualdade em (4.11) são definidos como

$$y_v^{(i)} = \begin{cases} \|\mathbf{f}(v)\| & \text{se } v \in A_i \\ 0 & \text{se } \text{dist}(v, A_i) \geq \frac{\delta}{2} \\ \|\mathbf{f}(v)\| \cdot \left(1 - \frac{2}{\delta} \cdot \text{dist}(v, A_i)\right) & \text{caso contrário} \end{cases} .$$

Note que os vetores $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)}$ são de suporte disjunto, pois não há como $y_v^{(i)} > 0$ e $y_v^{(j)} > 0$ para $i \neq j$, caso contrário teríamos que $\text{dist}(v, A_i) < \frac{\delta}{2}$ e $\text{dist}(v, A_j) < \frac{\delta}{2}$, contradizendo a hipótese de que $\text{dist}(u, v) \geq \delta$ para u, v de conjuntos distintos. Utilizando estes vetores de suporte disjunto e $\delta = \frac{1}{4\sqrt{5}k^3}$ como hipóteses do Lema 4.14, obtemos a desigualdade

$$\begin{aligned} R_{\mathcal{L}}(\mathbf{y}^{(i)}) &\leq C^* \cdot k \cdot \left(\frac{1}{4\sqrt{5}k^3} \right)^{-2} \cdot R_{\mathcal{L}}(\mathbf{f}) = \frac{C^2}{2} \cdot k^7 \cdot R_{\mathcal{L}}(\mathbf{f}) \\ &\leq \frac{C^2}{2} \cdot k^7 \cdot \lambda_k^{(\mathcal{L})}, \end{aligned} \quad (4.12)$$

onde em (4.12) foi utilizada a desigualdade em (4.10) e definimos, de forma conveniente, $C^* = \frac{C^2}{160}$.

Ainda sobre os vetores $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)}$, utilizamos o seguinte lema.

Lema 4.15. *Seja $\mathbf{y} \in \mathbb{R}^n$ vetor com entradas não negativas. Então existe $0 < t \leq \max\{y_j : j \in V\}$ tal que*

$$\phi_G(\{j : y_j \geq t\}) \leq \sqrt{2R_{\mathcal{L}}(\mathbf{y})}.$$

Em outras palavras, o Lema 4.15 afirma que, para cada $i = 1, \dots, k$, existe um subconjunto não vazio $S_i \subseteq \text{supp}(\mathbf{y}^{(i)})$ que satisfaz

$$\phi_G(S_i) \leq \sqrt{2R_{\mathcal{L}}(\mathbf{y}^{(i)})}. \quad (4.13)$$

Seja $\ell = 1, \dots, k$ tal que $\phi(S_\ell) = \max_{i=1, \dots, k} \phi_G(S_i)$, pela definição de condutância de G de ordem k podemos afirmar que

$$\phi_G(k) \leq \phi_G(S_\ell) \leq \sqrt{2R_{\mathcal{L}}(\mathbf{y}^{(\ell)})} \leq \sqrt{2 \cdot \frac{C^2}{2} \cdot k^7 \cdot \lambda_k^{(\mathcal{L})}} = C \cdot k^{3.5} \cdot \sqrt{\lambda_k^{(\mathcal{L})}},$$

concluindo a prova da desigualdade da cota superior de $\phi_G(k)$.

Alguns lemas utilizados para provar a Desigualdade de Cheeger de ordem k indicam que alguns vértices do grafo poderão não pertencer a nenhum dos conjuntos S_1, \dots, S_k obtidos no final de sua prova. No Lema 4.11, a projeção de algum

vértice poderá não estar em nenhum dos núcleos da malha e , conseqüentemente, em nenhum dos conjuntos S_1, \dots, S_k . Além disso, o Lema 4.15 garante a existência de um conjunto S_i que satisfaz a desigualdade em (4.13) e que é subconjunto do suporte do vetor $\mathbf{y}^{(i)}$, o que também indica que algum vértice que estava no suporte poderá ficar de fora no conjunto S_i . Não fica claro como seria possível adicionar estes vértices a um dos conjuntos S_1, \dots, S_k sem aumentar consideravelmente o valor da condutância.

4.3.4 Detalhes sobre o Algoritmo

Assim como a prova da cota superior da Desigualdade de Cheeger de ordem 2, a de ordem k demonstrada na subseção anterior também fornece um algoritmo que tem como resultado final uma coleção de k conjuntos S_1, \dots, S_k disjuntos e não-vazios utilizada para concluir o resultado. Como o resultado garante um limite superior definido para a condutância desta coleção, é possível que esta não esteja tão distante da menor condutância de ordem k . Portanto, este resultado teórico é uma possível evidência de que o algoritmo espectral fornecido pela demonstração tenha uma boa qualidade para o problema de condutância.

Nesta subseção, descreveremos esquematicamente os passos de um algoritmo que satisfaz as propriedades da prova da subseção anterior. Cabe mencionar que alguns dos passos da prova são descritos em um alto nível de generalidade e, por essa razão, o algoritmo poderia ser escrito de outras maneiras. Depois de descrevê-lo, ele será aplicado em exemplos para que possa entender o seu funcionamento.

Para facilitar a compreensão, subdividimos o algoritmo em quatro partes. Na primeira delas, nos preocupamos em definir as funções de empacotamento.

Algoritmo 2a: Empacotamento

Input: Matriz \mathcal{L} de G

Output: Vetores $\mathbf{f}(v_1), \dots, \mathbf{f}(v_n) \in \mathbb{R}^k$

- 1 Calcular os autovetores ortonormais $\{\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(k)}\}$ associados a $\lambda_1^{(\mathcal{L})} \leq \dots \leq \lambda_k^{(\mathcal{L})}$;
 - 2 Para cada $i = 1, \dots, n$, definir os vetores $\mathbf{f}(v_i) = (u_i^{(1)}, \dots, u_i^{(k)}) \in \mathbb{R}^k$.
-

No segundo algoritmo, queremos encontrar, conforme o Lema 4.11, $m \geq k$ conjuntos que respeitem certas propriedades.

Algoritmo 2b: Conjuntos Bem-Separados 1

Input: Vértices v_1, \dots, v_n e matriz \mathcal{L} de G e vetores

$$\mathbf{f}(v_1), \dots, \mathbf{f}(v_n) \in \mathbb{R}^k \text{ tais que } \sum_{i=1}^n \|\mathbf{f}(v_i)\|^2 = k$$

Output: Conjuntos T_1, \dots, T_m de vértices v_1, \dots, v_n de G

- 1 Para cada $i = 1, \dots, n$, definir os vetores $\bar{\mathbf{f}}(v_i) = \frac{\mathbf{f}(v_i)}{\|\mathbf{f}(v_i)\|} \in \mathbb{R}^k$;
 - 2 Criar em \mathbb{R}^k uma malha infinita de cubos de tamanho $L = \frac{1}{\sqrt{5}k}$;
 - 3 Para cada cubo da malha, definir o seu respectivo núcleo como um cubo de mesmo centro e tamanho $L \cdot (1 - \frac{1}{4k^2})$;
 - 4 **repeat**
 - 5 Tomar $\mathbf{c} = (c_1, \dots, c_k)$ tal que, para cada $j = 1, \dots, k$, c_j é um número aleatório no intervalo $[0, L]$;
 - 6 Deslocar a malha pelo vetor \mathbf{c} ;
 - 7 Sejam R_1, \dots, R_m os núcleos que contém pelo menos um dos vetores $\bar{\mathbf{f}}(v_i)$, definimos T_1, \dots, T_m os conjuntos de vértices tais que $T_j = \{v_i \in V : \bar{\mathbf{f}}(v_i) \in R_j\}$;
 - 8 $M = \sum_{j=1}^m \sum_{v_i \in T_j} \|\mathbf{f}(v_i)\|^2$.
 - 9 **until** $M \geq k - 1/4$;
-

Os conjuntos T_1, \dots, T_m produzidos pelo Algoritmo 2b satisfazem o Lema 4.11. Partindo desses conjuntos, o terceiro algoritmo produz conjuntos que satisfazem as propriedades dos conjuntos A_1, \dots, A_k do Lema 4.13.

Algoritmo 2c: Conjuntos Bem-Separados 2

Input: Conjuntos T_1, \dots, T_m de vértices v_1, \dots, v_n de G e suas respectivas massas e k inteiro fixado

Output: Conjuntos A_1, \dots, A_k de vértices v_1, \dots, v_n de G

- 1 Definir os conjuntos X_1, \dots, X_p como os T_i ($i = 1, \dots, m$) de massa $\geq 1/2$ e Y_1, \dots, Y_{m-p} como os de massa $< 1/2$;
- 2 **while** $p < k$ **do**
 - 3 Definir $i^* \neq j^*$ tais que $\text{dist}(Y_{i^*}, Y_{j^*}) = \min_{i \neq j} \text{dist}(Y_i, Y_j)$;
 - 4 $Z = Y_{i^*} \cup Y_{j^*}$;
 - 5 **if** massa de $Z \geq 1/2$ **then**
 - 6 $X_{p+1} = Z$;
 - 7 Redefinir os conjuntos Y_1, \dots, Y_{m-p-2} , removendo Y_{i^*} e Y_{j^*} ;
 - 8 $p = p + 1$;
 - 9 **else**
 - 10 $Y_{i^*} = Z$;
 - 11 Redefinir os conjuntos Y_1, \dots, Y_{m-p-1} , removendo Y_{j^*} ;
 - 12 **end**
 - 13 $m = m - 1$;
- 14 **end**
- 15 **while** $m > k$ **do**
 - 16 Definir i^* que satisfaça $\phi_G(X_{i^*}) = \max_{i=1, \dots, p} \phi_G(X_i)$;
 - 17 Definir j^* que satisfaça $\phi_G(X_{i^*} \cup Y_{j^*}) = \min_{j=1, \dots, m-p} \phi_G(X_{i^*} \cup Y_j)$;
 - 18 **if** $\phi_G(X_{i^*}) \geq \phi_G(X_{i^*} \cup Y_{j^*})$ **then**
 - 19 $X_{i^*} = X_{i^*} \cup Y_{j^*}$;
 - 20 Redefinir os conjuntos Y_1, \dots, Y_{m-p-1} , removendo Y_{j^*} ;
 - 21 $m = m - 1$;
 - 22 **else**
 - 23 $m = k$;
 - 24 **end**
- 25 **end**
- 26 **if** $p > k$ **then**
 - 27 Redefinir os conjuntos X_1, \dots, X_k como aqueles que possuem menor condutância;
- 28 **end**
- 29 Para cada $i = 1, \dots, k$, $A_i = X_i$.

Como a demonstração do Lema 4.13 apresentada na subseção anterior se preocupou em realizar operações que garantem que tenhamos k conjuntos de massa maior ou igual a $1/2$, estes podem não ter sido bem escolhidos no ponto de vista do algoritmo espectral. Por essa razão, a forma como escolhemos os k conjuntos A_1, \dots, A_k no Algoritmo 2c difere daquele mencionado na demonstração do Lema 4.13. A decisão de definir estes conjuntos conforme os passos do Algoritmo 2c foi tomada com base no resultado dos exemplos que serão apresentados nesta subseção.

No quarto e último algoritmo, como os conjuntos A_1, \dots, A_k obtidos no Algoritmo 2c satisfazem a hipótese do Lema 4.14, eles serão utilizados para produzir os k vetores $\mathbf{y}_1, \dots, \mathbf{y}_k$ de suporte disjunto que satisfazem a desigualdade do Lema 4.14. Em seguida, cada um destes vetores será utilizado no Lema 4.15 para obter um dos k conjuntos S_1, \dots, S_k finais.

Algoritmo 2d: Localizaçãõ

Input: Conjuntos A_1, \dots, A_k de vértices v_1, \dots, v_n de G e vetores de $\mathbf{f}(v_1), \dots, \mathbf{f}(v_n) \in \mathbb{R}^k$

Output: Conjuntos S_1, \dots, S_k de vértices v_1, \dots, v_n de G

- 1 Para cada $i = 1, \dots, n$ e $j = 1, \dots, k$, definir os vetores $\mathbf{y}^{(j)} = (y_1^{(j)}, \dots, y_n^{(j)})$ de tal forma que

$$y_i^{(j)} = \begin{cases} \|\mathbf{f}(v_i)\| & \text{se } v_i \in A_j \\ 0 & \text{se } \text{dist}(v_i, A_j) \geq \frac{1}{8\sqrt{5}k^3} ; \\ \|\mathbf{f}(v_i)\| \cdot [1 - 8\sqrt{5}k^3 \cdot \text{dist}(v_i, A_j)] & \text{caso contrário} \end{cases}$$

for $j < k$ **do**

- 2 Definir $p(j)$ como a quantidade de coordenadas não-nulas de $\mathbf{y}^{(j)}$;
- 3 Definir $\alpha_{j,1} < \dots < \alpha_{j,p(j)}$ como todas as possíveis coordenadas não nulas de $\mathbf{y}^{(j)}$;
- 4 Definir os conjuntos $S_{j,t} = \{v_i : y_i^{(j)} \geq \alpha_{j,t}\}$, onde $t = 1, \dots, p(j)$;
- 5 Calcular as funções $\phi_G(S_{j,t})$, onde $t = 1, \dots, p(j)$;
- 6 $S_j = \{S_{j,t} : \min_{t=1, \dots, p(j)} \phi_G(S_{j,t})\}$.
- 7 **end**
-

O Lema 4.15 garante a existência dos k conjuntos que satisfazem a desigualdade em (4.13), mas não define exatamente como eles são. Porém, pelos passos definidos no Algoritmo 2d, basta tomar dentre todos os candidatos o que possuir a menor condutância para garantir que este satisfaça a desigualdade.

Para entender o seu funcionamento, aplicaremos o algoritmo para $k = 2$ no grafo G da Figura 4.6a.

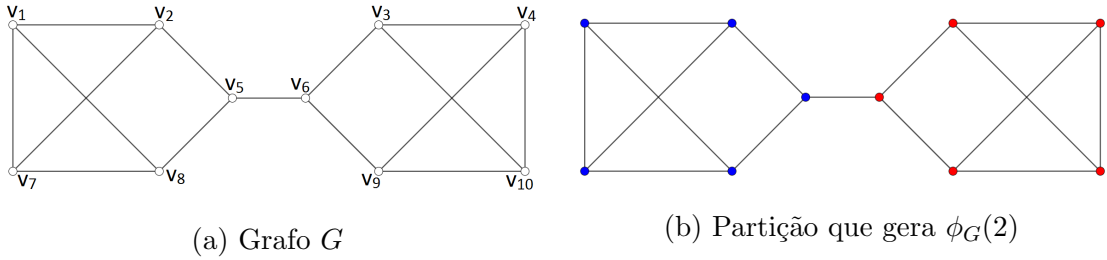
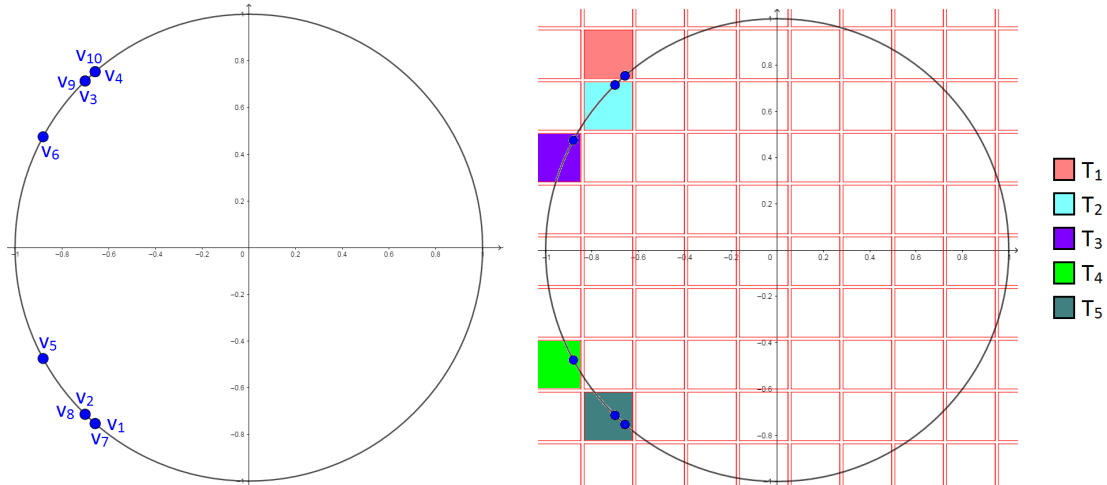


Figura 4.6: Exemplo de Grafo

É fácil verificar que a partição dada pela Figura 4.6b gera $\phi_G(2) = \frac{1}{15} = 0.067$, pois, sabendo que $\phi_G(2) = \min_{\text{vol}(S) \leq \frac{\text{vol}(V)}{2}} \phi_G(S)$, o menor numerador e maior denominador possíveis para $\phi_G(S)$, onde $S \subset V$ e $\text{vol}(S) \leq \frac{\text{vol}(V)}{2}$, são 1 e $\frac{\text{vol}(V)}{2} = \frac{3 \cdot |V|}{2} = 15$, respectivamente.

Suponha que já tenham sido realizados os passos que definem as projeções dos vértices na casca esférica unitária, realizados conforme Algoritmo 2a e primeiro passo do 2b. Na Figura 4.7a, podemos visualizar as projeções dos $f(v_i)$ na casca esférica unitária, onde, para cada projeção, está indicado o vértice correspondente.



(a) Projeções na Casca Esférica Unitária

(b) Malha de Cubos

Figura 4.7: Malha de Cubos

Iniciando o Algoritmo 2b a partir do segundo passo, inserimos a malha de cubos com seus respectivos núcleos no plano onde estão as projeções e ela será movimentada até que a massa total seja maior que $k - \frac{1}{4} = \frac{7}{4}$. Uma das posições possíveis da malha para que essa condição ocorra é a que aparece na Figura 4.7b, onde o vetor escolhido para mover a malha é, aproximadamente, $\mathbf{c} = (0.06, 0.05)$. Note que nela todos os vetores estão em algum núcleo e, portanto, nessa posição a massa total é igual a $2 > \frac{7}{4}$. Com esta posição, obtivemos os conjuntos $T_1 = \{v_4, v_{10}\}$,

$T_2 = \{v_3, v_9\}$, $T_3 = \{v_6\}$, $T_4 = \{v_5\}$, $T_5 = \{v_1, v_2, v_7, v_8\}$. Podemos observar na Figura 4.8 como os conjuntos T_i estão distribuídos no grafo.

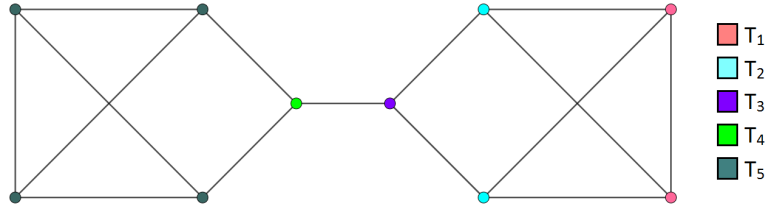


Figura 4.8: Conjuntos T_1, T_2, T_3, T_4 e T_5

Com o auxílio da Tabela 4.5, que mostra as normas ao quadrado de cada vetor $\mathbf{f}(v_i)$, podemos verificar que os conjuntos T_1, T_2, T_3, T_4 e T_5 possuem, respectivamente, massas iguais a $0.4630 < 1/2$, $0.4079 < 1/2$, $0.1291 < 1/2$, $0.1291 < 1/2$ e $0.8710 \geq 1/2$. Portanto, definimos $X_1 = T_5$, $Y_1 = T_1$, $Y_2 = T_2$, $Y_3 = T_3$ e $Y_4 = T_4$. Conforme primeiro laço do Algoritmo 2c, como há somente um conjunto de massa $\geq 1/2$, unimos dois conjuntos Y_{i^*} e Y_{j^*} que tenham a menor distância entre si dentre os de massa $< 1/2$. Pela Figura 4.7b, é fácil verificar que este dois conjuntos são $Y_1 = T_1$ e $Y_2 = T_2$. Como $Y_1 \cup Y_2$ tem massa $0.8710 \geq 1/2$, obtemos dois conjuntos de massa $\geq 1/2$, $X_1 = T_5$ e $X_2 = Y_1 \cup Y_2$, e redefinimos $Y_1 = T_3$ e $Y_2 = T_4$. Ao final deste laço, teremos os conjuntos distribuídos conforme Figura 4.9.

Tabela 4.5: Valor das Normas ao Quadrado dos $\mathbf{f}(v_i)$

Função	$\ \cdot\ ^2$	Função	$\ \cdot\ ^2$	Função	$\ \cdot\ ^2$
$\mathbf{f}(v_1)$	0.2315	$\mathbf{f}(v_5)$	0.1291	$\mathbf{f}(v_9)$	0.2040
$\mathbf{f}(v_2)$	0.2040	$\mathbf{f}(v_6)$	0.1291	$\mathbf{f}(v_{10})$	0.2315
$\mathbf{f}(v_3)$	0.2040	$\mathbf{f}(v_7)$	0.2315		
$\mathbf{f}(v_4)$	0.2315	$\mathbf{f}(v_8)$	0.2040		

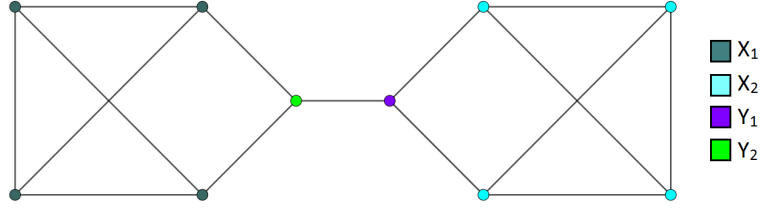


Figura 4.9: Conjuntos X_1 , X_2 , Y_1 e Y_2

No segundo laço, verificamos se, unindo o conjunto de maior condutância entre X_1 e X_2 a um dos conjuntos Y_1 e Y_2 diminuimos o valor da condutância. Consultando a Tabela 4.6, vemos que $\phi_G(X_1) = \phi_G(X_2) = 0.167$ e $\phi_G(X_1 \cup Y_2) = \phi_G(X_2 \cup Y_1) = 0.067 < 0.167$ são os menores valores de condutância nesta situação. Portanto, redefinimos os conjuntos $X_1 = X_1 \cup Y_2$ e $X_2 = X_2 \cup Y_1$ e obtemos $A_1 = X_1 = \{v_1, v_2, v_5, v_7, v_8\}$ e $A_2 = X_2 = \{v_3, v_4, v_6, v_9, v_{10}\}$ como conjuntos finais do Algoritmo 2c. O terceiro laço não será necessário, pois já obtemos exatamente $k = 2$ conjuntos de massa $\geq 1/2$.

Tabela 4.6: Valor das Condutâncias

$\phi_G(X_1) = 0.167$	$\phi_G(X_2) = 0.167$
$\phi_G(X_1 \cup Y_1) = 0.333$	$\phi_G(X_2 \cup Y_1) = 0.067$
$\phi_G(X_1 \cup Y_2) = 0.067$	$\phi_G(X_2 \cup Y_2) = 0.333$

Com os conjuntos A_1 e A_2 obtidos no Algoritmo 2c, formamos os vetores de suporte disjunto

$$\mathbf{y}^{(1)} = (0, 0, 0.4516, 0.4811, 0, 0.3593, 0, 0, 0.4516, 0.4811)$$

e

$$\mathbf{y}^{(2)} = (0.4811, 0.4516, 0, 0, 0.3593, 0, 0.4811, 0.4516, 0, 0)$$

definidos no primeiro passo do Algoritmo 2d. Sabendo que em cada um dos dois vetores as possíveis coordenadas são $0.4811 > 0.4516 > 0.3593$, definimos três subcon-

juntos do suporte de cada vetor e calculamos a sua condutância, conforme resultados da Tabela 4.7.

Tabela 4.7: Condutância dos $S_{i,t}$

Conjuntos ($i = 1$)	$\phi_G(S_{1,t})$	Conjuntos ($i = 2$)	$\phi_G(S_{2,t})$
$S_{1,1} = \{v_3, v_4, v_6, v_9, v_{10}\}$	0.067	$S_{2,1} = \{v_1, v_2, v_5, v_7, v_8\}$	0.067
$S_{1,2} = \{v_3, v_4, v_9, v_{10}\}$	0.167	$S_{2,2} = \{v_1, v_2, v_7, v_8\}$	0.167
$S_{1,3} = \{v_4, v_{10}\}$	0.667	$S_{2,3} = \{v_1, v_7\}$	0.667

Por obterem a menor condutância dentre os $S_{1,t}$ e $S_{2,t}$, definimos $S_1 = S_{1,1} = \{v_3, v_4, v_6, v_9, v_{10}\}$ e $S_2 = S_{2,1} = \{v_1, v_2, v_5, v_7, v_8\}$, encerrando o algoritmo. Note que estes dois conjuntos são os mesmos da Figura 4.6b, que geram $\phi_G(2)$.

Uma observação importante é que o algoritmo obtido com a construção da prova da subseção anterior não é único, pois alguns passos podem ser definidos de diferentes formas. No nosso caso, o Algoritmo 2c realiza alguns passos que não são os mesmos utilizados na demonstração do Lema 4.13. No momento em que há k conjuntos de massa $\geq 1/2$, o Algoritmo 2c começa a unir estes k conjuntos com os de massa $\geq 1/2$ caso a condutância diminua, enquanto que na demonstração do Lema 4.13 os conjuntos de massa $< 1/2$ são unidos uns com os outros até obtermos no máximo um conjunto de massa $< 1/2$. Caso o Algoritmo 2c procedesse desta maneira, é possível verificar que o resultado final do algoritmo não seria alguma cuja condutância seja igual a $\phi_G(2) = 0.067$.

A seguir veremos a aplicação do algoritmo para $k = 3$ no exemplo do grafo G da Figura 4.10a. A coleção de conjuntos $\{P_1, P_2, P_3\}$ que gera $\phi_G(3) = 0.067$ está representada na Figura 4.10b, onde $P_1 = \{v_1, v_2, v_5, v_{10}, v_{11}\}$ (vértices azuis), $P_2 = \{v_3, v_4, v_7, v_8, v_9, v_{12}, v_{13}\}$ (vértices vermelhos) e $P_3 = \{v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}\}$ (vértices verdes). É possível notar que esta coleção não é uma partição, mas, por simplicidade, iremos nos referir a ela assim.

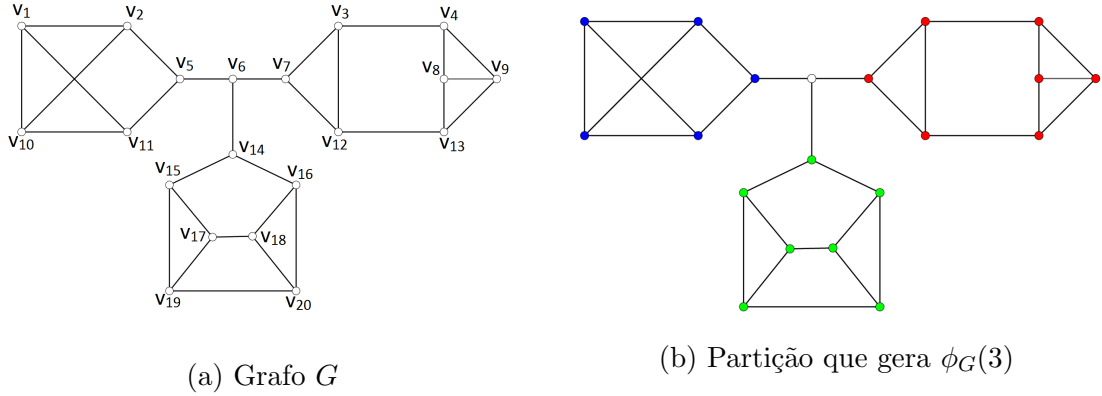


Figura 4.10: Segundo Exemplo

As coordenadas de cada projeção $\bar{\mathbf{f}}(v_i)$ são dadas pela Tabela 4.8. Elas serão importantes para o passo onde serão calculadas as distâncias entre os X_j , caso necessite.

Tabela 4.8: Projeções dos Vértices

Projeções	Projeções
$\bar{\mathbf{f}}(v_1) = (-0.4774, 0.0683, -0.8760)$	$\bar{\mathbf{f}}(v_{11}) = (-0.5016, 0.0684, -0.8624)$
$\bar{\mathbf{f}}(v_2) = (-0.5016, 0.0684, -0.8624)$	$\bar{\mathbf{f}}(v_{12}) = (-0.6417, -0.7125, 0.2839)$
$\bar{\mathbf{f}}(v_3) = (-0.6417, -0.7125, 0.2839)$	$\bar{\mathbf{f}}(v_{13}) = (-0.5559, -0.7619, 0.3323)$
$\bar{\mathbf{f}}(v_4) = (-0.5559, -0.7619, 0.3323)$	$\bar{\mathbf{f}}(v_{14}) = (-0.7509, 0.5839, 0.3085)$
$\bar{\mathbf{f}}(v_5) = (-0.6164, 0.0681, -0.7845)$	$\bar{\mathbf{f}}(v_{15}) = (-0.6071, 0.6715, 0.4250)$
$\bar{\mathbf{f}}(v_6) = (-0.9757, 0.0473, -0.2140)$	$\bar{\mathbf{f}}(v_{16}) = (-0.6071, 0.6715, 0.4250)$
$\bar{\mathbf{f}}(v_7) = (-0.7877, -0.5885, 0.1824)$	$\bar{\mathbf{f}}(v_{17}) = (-0.5655, 0.6894, 0.4528)$
$\bar{\mathbf{f}}(v_8) = (-0.5367, -0.7713, 0.3423)$	$\bar{\mathbf{f}}(v_{18}) = (-0.5655, 0.6894, 0.4528)$
$\bar{\mathbf{f}}(v_9) = (-0.5367, -0.7713, 0.3423)$	$\bar{\mathbf{f}}(v_{19}) = (-0.5655, 0.6894, 0.4528)$
$\bar{\mathbf{f}}(v_{10}) = (-0.4774, 0.0683, -0.8760)$	$\bar{\mathbf{f}}(v_{20}) = (-0.5655, 0.6894, 0.4528)$

Na Figura 4.11 estão representadas as segundas e terceiras coordenadas de cada projeção e uma possível malha de cubos que satisfaz massa total maior que

$3 - \frac{1}{4} = \frac{11}{4}$. Sobre a primeira coordenada, há quatro intervalos distintos entre os núcleos que contém pelo menos uma das projeções; na Figura 4.11 as projeções que estão em intervalos distintos são representados por diferentes tonalidades azuis. Com isso podemos concluir que a projeção $\bar{\mathbf{f}}(v_5)$ não está no mesmo núcleo que $\bar{\mathbf{f}}(v_1)$, $\bar{\mathbf{f}}(v_2)$, $\bar{\mathbf{f}}(v_{10})$ e $\bar{\mathbf{f}}(v_{11})$, e $\bar{\mathbf{f}}(v_{14})$ não está no mesmo núcleo que $\bar{\mathbf{f}}(v_{15})$ e $\bar{\mathbf{f}}(v_{16})$, pois apesar de parecerem estar no mesmo núcleo na perspectiva das segunda e terceira coordenada, na perspectiva da primeira não estão. A figura também indica que os vetores $\mathbf{f}(v_4)$ e $\mathbf{f}(v_{13})$ estão em um núcleo diferente de $\mathbf{f}(v_8)$ e $\mathbf{f}(v_9)$.

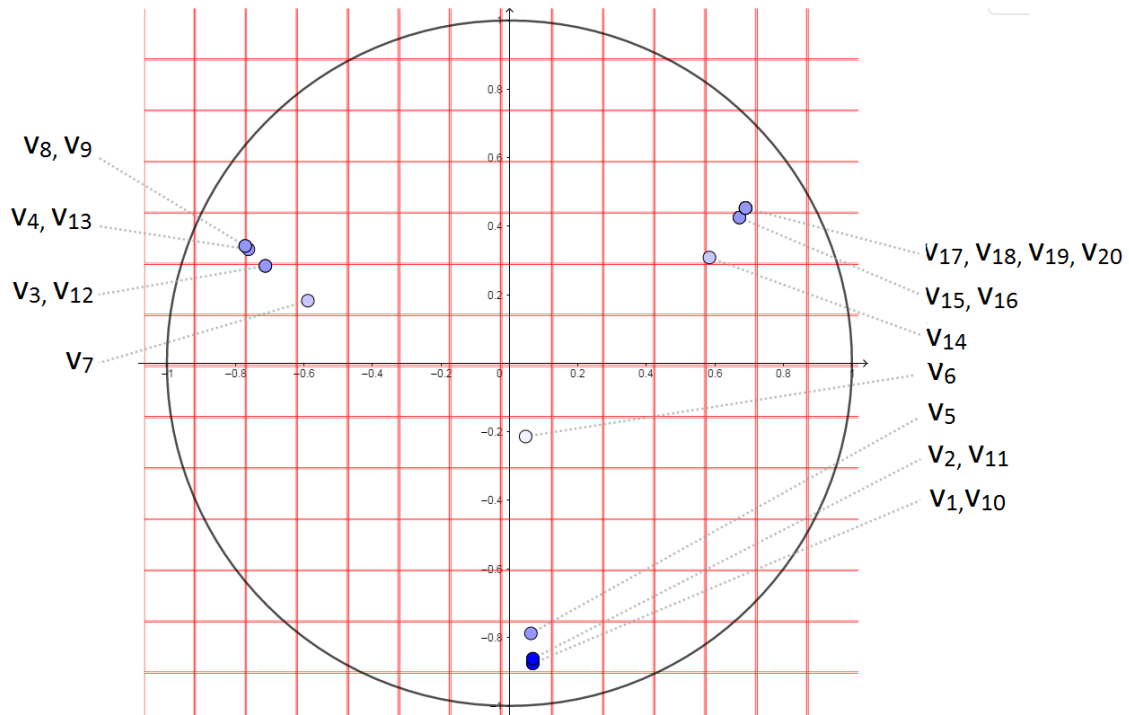


Figura 4.11: Malha do Segundo Exemplo

Com esta malha formamos os conjuntos $X_1 = \{v_{17}, v_{18}, v_{19}, v_{20}\}$ e $X_2 = \{v_1, v_2, v_{10}, v_{11}\}$ de massa $\geq 1/2$ e $Y_1 = \{v_8, v_9\}$, $Y_2 = \{v_4, v_{13}\}$, $Y_3 = \{v_{15}, v_{16}\}$, $Y_4 = \{v_{14}\}$, $Y_5 = \{v_3, v_{12}\}$, $Y_6 = \{v_7\}$, $Y_7 = \{v_6\}$ e $Y_8 = \{v_5\}$ de massa $< 1/2$. Conforme terceira parte do algoritmo, como temos dois conjuntos de massa $\geq 1/2$ iremos unir aqueles de massa $< 1/2$. Calculando todas as possíveis distâncias entre os Y_j , encontramos que $\text{dist}(Y_1, Y_2)$ é a menor delas e, portanto, unimos Y_1 e Y_2 para

formar um novo conjunto. Como a massa de $Y_1 \cup Y_2$ é $0.6708 \geq 1/2$, redefinimos os conjuntos de massas $\geq 1/2$ e $< 1/2$ obtendo $X_1 = \{v_{17}, v_{18}, v_{19}, v_{20}\}$, $X_2 = \{v_4, v_8, v_9, v_{13}\}$ e $X_3 = \{v_1, v_2, v_{10}, v_{11}\}$, $Y_1 = \{v_{15}, v_{16}\}$, $Y_2 = \{v_{14}\}$, $Y_3 = \{v_3, v_{12}\}$, $Y_4 = \{v_7\}$, $Y_5 = \{v_6\}$ e $Y_6 = \{v_5\}$.

Como temos três conjuntos de massa igual a $1/2$, passamos a analisar se há como diminuir a maior condutância dentre os X_i unindo a ele algum Y_j . Podemos verificar que $\phi_G(X_1) = 0.333 > \phi_G(X_2) = \phi_G(X_3) = 0.167$. Além disso, $\phi_G(X_1 \cup Y_1) = 0.111$ é a menor das condutâncias de X_1 unido com algum Y_j . Como $\phi_G(X_1 \cup Y_1) < \phi_G(X_1)$, redefinimos os conjuntos $X_1 = \{v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}\}$, $X_2 = \{v_4, v_8, v_9, v_{13}\}$ e $X_3 = \{v_1, v_2, v_{10}, v_{11}\}$, $Y_1 = \{v_{14}\}$, $Y_2 = \{v_3, v_{12}\}$, $Y_3 = \{v_7\}$, $Y_4 = \{v_6\}$ e $Y_5 = \{v_5\}$.

Repetimos este passo até encontrar $X_1 = \{v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}\}$, $X_2 = \{v_3, v_4, v_8, v_9, v_{12}, v_{13}\}$, $X_3 = \{v_1, v_2, v_5, v_{10}, v_{11}\}$ e $Y_1 = \{v_6\}$. Podemos verificar que $\phi_G(X_3) = 0.067 > \phi_G(X_1) = \phi_G(X_2) = 0.048$. Além disso, temos que $\phi_G(X_3 \cup Y_1) > \phi_G(X_3)$. Logo, descartamos Y_1 para o restante do algoritmo e obtemos $A_1 = X_1$, $A_2 = X_2$ e $A_3 = X_3$.

Prosseguindo com o algoritmo até o fim, encontramos que a resposta final dele serão os próprios conjuntos A_1 , A_2 e A_3 que é aquela que gera $\phi_G(3)$ conforme Figura 4.10b.

Para provar que os conjuntos A_1, A_2 e A_3 geram $\phi_G(3) = \frac{1}{15}$, mostraremos que não existem outros três conjuntos $R_1, R_2, R_3 \subset V$ disjuntos tais que a maior condutância entre eles seja menor que $\frac{1}{15}$. Conforme demonstração para $\phi_G(2)$ na Seção 4.2.2, os únicos conjuntos cujos cortes com seus complementares é igual a 1 são exatamente A_1, A_2 e A_3 . Com isso, supondo que R_1 tenha a maior condutância entre R_1, R_2 e R_3 , temos que $\text{Cut}(R_1, \overline{R_1}) > 1$ e, conseqüentemente, $|S| > 10$ para que $\phi(R_1) < \frac{1}{15}$. No entanto, a única possibilidade para que as condutâncias de R_2 e R_3 sejam menores ou iguais que as de R_1 é se $|R_2| = |R_3| = 5$ e

$\text{Cut}(R_2, \overline{R_2}) = \text{Cut}(R_3, \overline{R_3}) = 1$. Como vimos que só há um subconjunto de V de tamanho 5 cujo corte com seu complementar é igual a 1, chegamos a uma contradição. Logo os conjuntos A_1, A_2 e A_3 geram $\phi_G(3) = \frac{1}{15}$.

Como último exemplo, retomamos o grafo da Figura 4.5a. Testando todas as condutâncias das partições do conjunto de vértices deste grafo, é possível verificar que a que está representada na Figura 4.12b é a que gera $\phi_G(3) = \frac{1}{15} = 0.0729$.

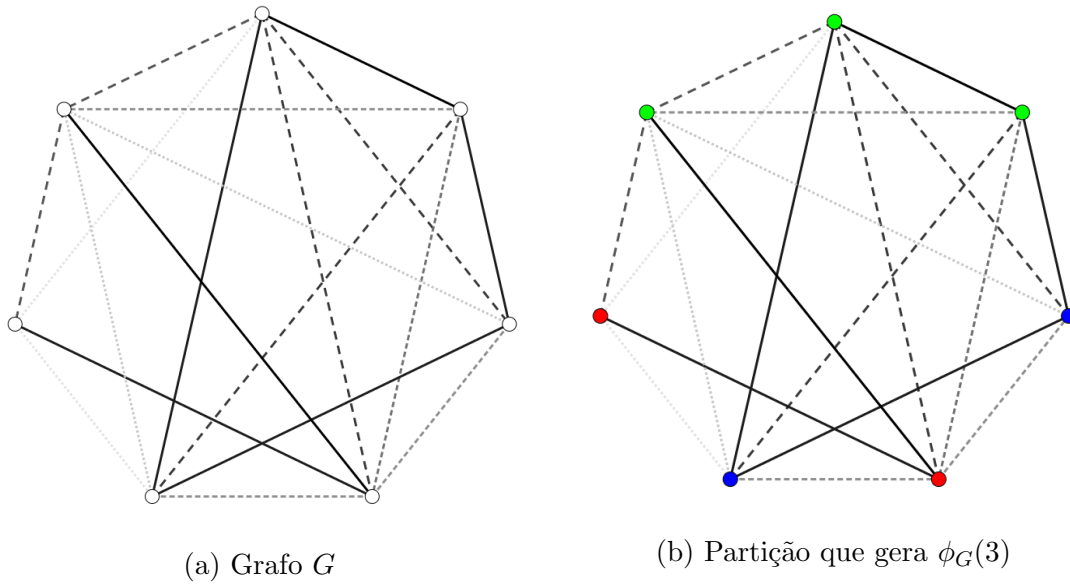


Figura 4.12: Partição que gera $\phi_G(3)$

Os únicos conjuntos que poderão ser gerados com a malha de cubos é $T_1 = \{A\}$, $T_2 = \{B\}$, $T_3 = \{C\}$, $T_4 = \{D\}$, $T_5 = \{E\}$, $T_6 = \{F\}$ e $T_7 = \{G\}$, pois, neste exemplo, não há como duas projeções $\bar{\mathbf{f}}(u)$ e $\bar{\mathbf{f}}(v)$, para $u \neq v$, estarem no mesmo núcleo. Além disso, temos neste caso $\|\mathbf{f}(A)\|^2 = 0.424$, $\|\mathbf{f}(B)\|^2 = 0.375$, $\|\mathbf{f}(C)\|^2 = 0.399$, $\|\mathbf{f}(D)\|^2 = 0.331$, $\|\mathbf{f}(E)\|^2 = 0.507$, $\|\mathbf{f}(F)\|^2 = 0.530$ e $\|\mathbf{f}(G)\|^2 = 0.435$ como as massas dos conjuntos T_i .

Como há dois conjuntos T_i de massa $< 1/2$, tentamos unir, entre os demais, aqueles cuja projeção sejam mais próximas; neste caso, T_1 e T_2 . Como

$T_1 \cup T_2$ tem massa $\geq 1/2$, definimos $X_1 = \{T_1 \cup T_2\}$, $X_2 = T_5$, $X_3 = T_6$, $Y_1 = T_3$, $Y_2 = T_4$, $Y_3 = T_7$.

Como $\phi_G(X_2) = \phi_G(X_3) = 1 > \phi_G(X_1) = 0.746$, verificamos se há possibilidade de unir aos X_2 e X_3 os Y_j consultando as condutâncias da Tabela 4.9.

Tabela 4.9: Primeiro Teste das Condutâncias

* $\phi_G(X_2 \cup Y_1) = 0.729$	$\phi_G(X_2 \cup Y_2) = 0.881$	$\phi_G(X_2 \cup Y_3) = 0.932$
$\phi_G(X_3 \cup Y_1) = 1.000$	* $\phi_G(X_3 \cup Y_2) = 0.698$	$\phi_G(X_3 \cup Y_3) = 0.733$

* União que gerou menor condutância para cada X_i .

Logo, redefinimos $X_1 = \{A, B\}$, $X_2 = \{C, E\}$, $X_3 = \{D, F\}$ e $Y_1 = \{G\}$. Além disso, temos $\phi_G(X_1) = 0.746 > \phi_G(X_2) = 0.729 > \phi_G(X_3) = 0.698$. Verificando a condutância de $X_1 \cup Y_1$ temos $\phi_G(X_1 \cup Y_1) = 0.620 < \phi_G(X_1) = 0.746$. No fim, geramos os k conjuntos $A_1 = \{A, B, G\}$, $A_2 = \{C, E\}$ e $A_3 = \{D, F\}$.

Ordenando as coordenadas de cada vetor de suporte disjunto

$$\begin{aligned} \mathbf{y}^{(1)} &= (0.651, 0.612, 0, 0, 0, 0, 0.659), \\ \mathbf{y}^{(2)} &= (0, 0, 0.631, 0, 0.712, 0, 0) \text{ e} \\ \mathbf{y}^{(3)} &= (0, 0, 0, 0.575, 0, 0.728, 0) \end{aligned}$$

temos os conjuntos $S_{1,1} = \{A, B, G\}$, $S_{1,2} = \{A, G\}$, $S_{1,3} = \{G\}$, $S_{2,1} = \{C, E\}$, $S_{2,2} = \{E\}$, $S_{3,1} = \{D, F\}$, $S_{3,2} = \{F\}$. Além disso, $\phi_G(S_{1,1}) = 0.62 < \phi_G(S_{1,2}) = 0.821 < \phi_G(S_{1,3}) = 1$, $\phi_G(S_{2,1}) = 0.729 < \phi_G(S_{2,2}) = 1$ e $\phi_G(S_{3,1}) = 0.698 < \phi_G(S_{1,2}) = 1$ e, portanto, temos como resultado do algoritmo que $S_1 = S_{1,1}$, $S_2 = S_{2,1}$, $S_3 = S_{3,1}$, que é a partição que gera $\phi_G(3)$ conforme Figura 4.12b.

4.3.5 Demonstração dos Lemas Auxiliares

A seguir veremos as provas dos lemas que foram utilizados para provar a cota superior da Desigualdade de Cheeger de ordem k . Em todos os lemas, consideramos um grafo $G = (V, E)$ d -regular não-ponderado e \mathbf{f} sua função de mapeamento.

O seguinte lema utiliza o quociente de Rayleigh da função de mapeamento definida na Subseção 4.3.3.

Lema 4.10. *Seja o mapeamento $\mathbf{f} : V \rightarrow \mathbb{R}^k$ definido em termos de k vetores ortonormais $\mathbf{x}^{(i)}$, temos que*

$$R_{\mathcal{L}}(\mathbf{f}) \leq \max_{i=1, \dots, k} R_{\mathcal{L}}(\mathbf{x}^{(i)}).$$

Demonstração.

$$\begin{aligned} R_{\mathcal{L}}(\mathbf{f}) &= \frac{\sum_{\{u,v\} \in E} \sum_i (x_u^{(i)} - x_v^{(i)})^2}{d \sum_v \sum_i (x_v^{(i)})^2} = \frac{\sum_i \sum_{\{u,v\} \in E} (x_u^{(i)} - x_v^{(i)})^2}{dk} \\ &= \frac{1}{k} \sum_i \frac{\sum_{\{u,v\} \in E} (x_u^{(i)} - x_v^{(i)})^2}{d} = \frac{1}{k} \sum_i R_{\mathcal{L}}(\mathbf{x}^{(i)}) \\ &\leq \max_{i=1, \dots, k} R_{\mathcal{L}}(\mathbf{x}^{(i)}). \end{aligned}$$

□

O Lema 4.13 é uma consequência do que vem a seguir. Por conveniência, enunciamos o resultado novamente.

Lema 4.11. *Existem $m \geq k$ conjuntos disjuntos de vértices T_1, \dots, T_m tais que*

$$(1) \sum_{i=1}^m \sum_{v \in T_i} \|\mathbf{f}(v)\|^2 \geq k - \frac{1}{4};$$

$$(2) \text{ Para quaisquer vértices } u, v \text{ em conjuntos diferentes, } \text{dist}(u, v) \geq C \cdot k^{-3}, \text{ onde } C \in \mathbb{R}_{>0};$$

$$(3) \text{ Para qualquer conjunto } T_i, \text{ vale } \sum_{v \in T_i} \|\mathbf{f}(v)\|^2 \leq 1 + \frac{1}{4k}.$$

Na Seção 4.3.3 provamos que, com a construção da malha de cubos de lado $L = \frac{1}{\sqrt{5}k}$, é possível obter os conjuntos T_1, \dots, T_m que satisfazem os itens (2) e (3). Em relação ao item (1), foi demonstrado no capítulo 2 que o modelo probabilístico definido consegue garantir que, construindo em \mathbb{R}^k uma malha de cubos de lado L cujos núcleos distam pelo menos um valor δ , para um conjunto de vetores $\mathbf{f}(v_1), \dots, \mathbf{f}(v_n)$ que satisfazem $\sum_{i=1}^n \|\mathbf{f}(v_i)\|^2 = k$ vale $\sum_{\mathbf{f}(v_i) \in C(x)} \|\mathbf{f}_i\|^2 \geq k \cdot \left(1 - \frac{k\delta}{L}\right)$, onde $C(x)$ é a união das regiões internas de cada núcleo gerado por uma certa posição da malha. É possível notar que os somatórios $\sum_{\mathbf{f}(v_i) \in C(x)} \|\mathbf{f}(v_i)\|^2$ e $\sum_{i=1}^m \sum_{v \in T_i} \|\mathbf{f}(v)\|^2$ se equivalem e, além disso, tomando $\delta = \frac{L}{4k^2}$, encontramos $1 - \frac{k\delta}{L} = k - \frac{1}{4}$.

Lema 4.12. *Para todo subconjunto R da casca esférica unitária, temos que*

$$\sum_{v \in V(R)} \|\mathbf{f}(v)\|^2 \leq \left(1 - \frac{1}{2}(\text{diam}(R))^2\right)^{-2}.$$

Demonstração. Para quaisquer vetores \mathbf{z} e \mathbf{w} unitários, temos

$$\langle \mathbf{z}, \mathbf{w} \rangle = 1 - \frac{1}{2}\|\mathbf{z} - \mathbf{w}\|^2.$$

Além disso, considere $X \in \mathbb{R}^{k \times n}$ cujas colunas são os $\mathbf{f}(v)$. Então temos que

$$\sum_{v \in V} \langle \mathbf{f}(v), \mathbf{w} \rangle^2 = \|X^T \mathbf{w}\|^2 = \mathbf{w}^T X X^T \mathbf{w} = \mathbf{w}^T \mathbf{w} = 1.$$

Seja $\mathbf{w} \in R$, então

$$\begin{aligned} 1 &\geq \sum_{v \in V(R)} \langle \mathbf{f}(v), \mathbf{w} \rangle^2 = \sum_{v \in V(R)} \|\mathbf{f}(v)\|^2 \cdot \langle \bar{\mathbf{f}}(v), \mathbf{w} \rangle^2 \\ &= \sum_{v \in V(R)} \|\mathbf{f}(v)\|^2 \cdot \left(1 - \frac{1}{2}\|\bar{\mathbf{f}}(v) - \mathbf{w}\|^2\right)^2 \\ &\geq \sum_{v \in V(R)} \|\mathbf{f}(v)\|^2 \cdot \left(1 - \frac{1}{2}(\text{diam}(R))^2\right)^2. \end{aligned}$$

□

Lema 4.16.

$$\|\mathbf{f}(v)\| \cdot \text{dist}(u, v) \leq 2\|\mathbf{f}(u) - \mathbf{f}(v)\|.$$

Demonstração.

$$\begin{aligned}
\|\mathbf{f}(v)\| \cdot \text{dist}(u, v) &= \|\mathbf{f}(v)\| \cdot \left\| \frac{\mathbf{f}(u)}{\|\mathbf{f}(u)\|} - \frac{\mathbf{f}(v)}{\|\mathbf{f}(v)\|} \right\| = \left\| \mathbf{f}(u) \cdot \frac{\|\mathbf{f}(v)\|}{\|\mathbf{f}(u)\|} - \mathbf{f}(v) \right\| \\
&\leq \left\| \mathbf{f}(u) \cdot \frac{\|\mathbf{f}(v)\|}{\|\mathbf{f}(u)\|} - \mathbf{f}(u) \right\| + \|\mathbf{f}(u) - \mathbf{f}(v)\| \quad (4.14) \\
&= \left\| \mathbf{f}(u) \cdot \left(\frac{\|\mathbf{f}(v)\|}{\|\mathbf{f}(u)\|} - 1 \right) \right\| + \|\mathbf{f}(u) - \mathbf{f}(v)\| \\
&= \|\mathbf{f}(u)\| \cdot \left| \frac{\|\mathbf{f}(v)\|}{\|\mathbf{f}(u)\|} - 1 \right| + \|\mathbf{f}(u) - \mathbf{f}(v)\| \\
&= \left| \|\mathbf{f}(u)\| - \|\mathbf{f}(v)\| \right| + \|\mathbf{f}(u) - \mathbf{f}(v)\| \\
&\leq 2\|\mathbf{f}(u) - \mathbf{f}(v)\|, \quad (4.15)
\end{aligned}$$

onde em (4.14) e (4.15) foi utilizada a Desigualdade Triangular (Corolário 2.18). \square

Lema 4.14 (Localização). *Seja $\mathbf{f} : V \rightarrow \mathbb{R}^k$ uma função de mapeamento em termos de k vetores $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}$ ortonormais. Dados k conjuntos de vértices A_1, \dots, A_k de vértices tais que, para todo $i = 1, \dots, k$, $\sum_{v \in A_i} \|\mathbf{f}(v)\|^2 \geq \frac{1}{2}$ e, para quaisquer vértices u, v de conjuntos diferentes, $\text{dist}(u, v) \geq \delta$, podemos construir k vetores $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(k)} \in \mathbb{R}^n$ de suporte disjunto e entradas não negativas tais que, para quaisquer $i = 1, \dots, k$, vale*

$$R_{\mathcal{L}}(\mathbf{y}^{(i)}) \leq C \cdot k \cdot \delta^{-2} \cdot R_{\mathcal{L}}(\mathbf{f}),$$

onde $C \in \mathbb{R}_{>0}$.

Demonstração. Queremos mostrar que os vetores $\mathbf{y}^{(i)}$ tais que $y_v^{(i)} = \tau_i(v) \cdot \|\mathbf{f}(v)\|$ e

$$\tau_i(v) = \begin{cases} 0 & \text{se } \text{dist}(v, A_i) \geq \frac{\delta}{2} \\ 1 - \frac{2}{\delta} \cdot \text{dist}(v, A_i) & \text{se } \text{dist}(v, A_i) < \frac{\delta}{2}, \end{cases}$$

onde para $v \in A_i$ vale $\text{dist}(v, A_i) = 0$, satisfazem as condições propostas no Lema de Localização.

No denominador de $R_{\mathcal{L}}(\mathbf{y}^{(i)})$ temos

$$d \cdot \sum_{v \in V} \tau_i(v)^2 \cdot \|\mathbf{f}(v)\|^2 \geq d \cdot \sum_{v \in A_i} \|\mathbf{f}(v)\|^2 \geq d \cdot \frac{1}{2}.$$

A contribuição da uma aresta $\{u, v\}$ do numerador é o quadrado de

$$\begin{aligned}
|y_v^{(i)} - y_u^{(i)}| &= |\tau_i(v) \cdot \|\mathbf{f}(v)\| - \tau_i(u) \cdot \|\mathbf{f}(u)\|| \\
&= |\tau_i(v) \cdot \|\mathbf{f}(v)\| - \tau_i(v) \cdot \|\mathbf{f}(u)\| + \tau_i(v) \cdot \|\mathbf{f}(u)\| - \tau_i(u) \cdot \|\mathbf{f}(u)\|| \\
&\leq |\tau_i(v) \cdot \|\mathbf{f}(v)\| - \tau_i(v) \cdot \|\mathbf{f}(u)\|| + |\tau_i(v) \cdot \|\mathbf{f}(u)\| - \tau_i(u) \cdot \|\mathbf{f}(u)\|| \\
&\leq \tau_i(v) \cdot \|\mathbf{f}(v) - \mathbf{f}(u)\| + \|\mathbf{f}(u)\| \cdot |\tau_i(v) - \tau_i(u)| \tag{4.16}
\end{aligned}$$

$$\begin{aligned}
&\leq \|\mathbf{f}(v) - \mathbf{f}(u)\| + \|\mathbf{f}(u)\| \cdot |\text{dist}(v, A_i) - \text{dist}(u, A_i)| \\
&\leq \|\mathbf{f}(v) - \mathbf{f}(u)\| + \|\mathbf{f}(u)\| \cdot \frac{2}{\delta} \cdot \text{dist}(v, u) \\
&\leq \|\mathbf{f}(v) - \mathbf{f}(u)\| \cdot \left(1 + \frac{4}{\delta}\right) \tag{4.17}
\end{aligned}$$

onde em (4.16) foi utilizada a Desigualdade Triangular (Corolário 2.18) e em (4.17) utilizamos o Lema 4.16.

Note que

$$\sum_{v \in V} \|\mathbf{f}(v)\|^2 = \sum_{v \in V} \sum_{i=1}^k (x_v^{(i)})^2 = \sum_{i=1}^k \|\mathbf{x}^{(i)}\|^2 = k.$$

Portanto, o numerador de $R_{\mathcal{L}}(\mathbf{y}^{(i)})$ é

$$\begin{aligned}
\sum_{\{u,v\} \in E} |y_v^{(i)} - y_u^{(i)}|^2 &\leq \left(1 + \frac{4}{\delta}\right)^2 \cdot \sum_{\{u,v\} \in E} \|\mathbf{f}(v) - \mathbf{f}(u)\|^2 \\
&\leq C^* \cdot \delta^{-2} \cdot \sum_{\{u,v\} \in E} \|\mathbf{f}(v) - \mathbf{f}(u)\|^2 \\
&= C^* \cdot k \cdot \delta^{-2} \cdot R_{\mathcal{L}}(\mathbf{f}).
\end{aligned}$$

Considerando $C = \frac{2 \cdot C^*}{d}$, temos

$$R_{\mathcal{L}}(\mathbf{y}^{(i)}) \leq C \cdot k \cdot \delta^{-2} \cdot R_{\mathcal{L}}(\mathbf{f}).$$

□

Lema 4.15. *Seja $\mathbf{y} \in \mathbb{R}^n$ vetor com entradas não negativas. Então existe $0 < t \leq \max\{y_j : v_j \in V\}$ tal que*

$$\phi_G(\{v_j \in V : y_j \geq t\}) \leq \sqrt{2R_{\mathcal{L}}(\mathbf{y})}.$$

Demonstração. Seja o subconjunto $X_r = \{v_j \in V : y_j \geq \sqrt{r}\}$ e $m = \max\{y_j^2 : v_j \in V\}$. Nos passos seguintes, utilizaremos a integral de uma função degrau como uma soma de áreas de retângulos.

Note que $g(r) = \text{vol}(X_r)$ é uma função degrau e, além disso, o valor do grau d_j contribui para a soma que resulta $\text{vol}(X_r)$ enquanto $v_j \in X_r$, ou seja, quando $r \in (0, y_j^2]$. Logo a integral de $\text{vol}(X_r)$ em r de 0 a m é igual à soma das áreas dos retângulos $d_j \times y_j^2$, para cada j . Com isso, obtemos a igualdade

$$\int_0^m \text{vol}(X_r) dr = \sum_{v_j \in V} d_j y_j^2.$$

Por outro lado, temos que $f(r) = \text{Cut}(X_r, \overline{X_r})$ também é uma função e degrau. Além disso, para cada $i \neq j$ tal que $y_i \geq y_j$, w_{ij} contribui para a soma de pesos que resulta o $\text{Cut}(X_r, \overline{X_r})$ enquanto $v_i \in X_r$ e $v_j \in \overline{X_r}$, ou seja, quando $r \in (y_j^2, y_i^2]$. Logo a integral de $\text{Cut}(X_r, \overline{X_r})$ em t de 0 a m é igual à soma das áreas dos retângulos $w_{ij} \times |y_i^2 - y_j^2|$, para cada v_i, v_j adjacentes. Portanto temos que

$$\begin{aligned} \int_0^m \text{Cut}(X_r, \overline{X_r}) dr &= \sum_{\{v_i, v_j\} \in E} w_{ij} |y_i^2 - y_j^2| \\ &= \sum_{\{v_i, v_j\} \in E} w_{ij} |y_i - y_j| \cdot (y_i + y_j) \\ &\leq \sqrt{\sum_{\{v_i, v_j\} \in E} w_{ij} (y_i - y_j)^2} \sqrt{\sum_{\{v_i, v_j\} \in E} w_{ij} (y_i + y_j)^2} \quad (4.18) \end{aligned}$$

$$\leq \sqrt{\sum_{\{v_i, v_j\} \in E} w_{ij} (y_i - y_j)^2} \sqrt{2 \sum_{v_i \in V} d_i y_i^2}, \quad (4.19)$$

onde nas desigualdades (4.18) e (4.19) foram utilizados os Teoremas 2.17 (desigualdade de Cheeger) e 2.22, respectivamente. Combinando as desigualdades encontradas e utilizando o Lema 2.20, podemos afirmar que existe um $0 < t \leq m$ tal que

$$\phi_G(\{v_j \in V : y_j \geq t\}) \leq \frac{\int_0^m \text{Cut}(X_r, \overline{X_r}) dr}{\int_0^m \text{vol}(X_r) dr} \leq \sqrt{2 \cdot \frac{\sum_{\{v_i, v_j\} \in E} w_{ij} (y_i - y_j)^2}{\sum_{v_i \in V} d_i y_i^2}}$$

$$= \sqrt{2R_{\mathcal{L}}(\mathbf{y})}, \quad (4.20)$$

onde em a igualdade em (4.20) é atingida pelo item (2) da Proposição 2.15. \square

5 CONSIDERAÇÕES FINAIS

Nesta dissertação, foi possível estudar diferentes estratégias espectrais para problemas de particionamento. No caso dos problemas de minimização das funções RatioCut e NCut, obtemos a relaxação de um problema matemático bem definido e formulado conforme o problema original como passo espectral para uma heurística. Por se tratar de uma aproximação do problema original, a utilização do problema relaxado na heurística traz evidências de que esta, quando utilizada com um passo de *clustering* que geralmente obtém bons resultados, como o algoritmo das k -médias, pode trazer resultados satisfatórios para o problema.

Além disso, estudamos o resultado apresentado por Wang et. al [35], que diz que a melhor partição em duas classes de uma árvore utilizando a função NCut garante que o grafo induzido de cada uma de suas classes é conexo. Eles afirmam que o resultado também vale para $k > 2$ classes, porém não demonstram esse fato e nem citam referências. Neste trabalho, propusemos uma possível estratégia de demonstração para este resultado utilizando indução e grafos que possuem pesos em seus vértices.

Outras estratégias estudadas detalhadamente foram os algoritmos de aproximação fornecidos pelas demonstrações apresentadas por Chung [8] e Trevisan [33]. A demonstração de Chung encontra uma partição $\{S, \bar{S}\}$ que satisfaz $\phi_G(2) \leq \phi_G(S, \bar{S}) \leq 2\sqrt{\phi_G(2)}$. Da mesma forma, a demonstração de Trevisan encontra uma coleção $\{S_1, \dots, S_k\}$ de k conjuntos disjuntos e não vazios que satisfazem $\phi_G(k) \leq \phi_G(S_1, \dots, S_k) \leq C^* \cdot k^{3.5} \sqrt{\lambda_k^{(\mathcal{L})}}$, onde C^* é uma constante absoluta. Vale também mencionar que, para a prova de Trevisan, observamos que há mais de uma maneira de definir o algoritmo. Isto ocorre porque alguns passos da prova não são explícitos, visto que o resultado é válido independente da escolha realizada. Não realizamos uma pesquisa experimental de grande escala, mas observamos que os

passos do algoritmo que implementamos obtiveram bons resultados nos exemplos apresentados.

5.1 Trabalhos Futuros

A compreensão dos resultados relacionados ao problema de particionamento apresentados neste trabalho proporcionou um ambiente promissor para a realização de outros estudos deste tipo. Com o conhecimento adquirido com a realização deste trabalho, será possível investigar novas questões sobre o problema de particionamento ou estudar detalhadamente aqueles que já constam na literatura, sejam resultados tradicionais ou recentes.

O estudo realizado neste trabalho sobre particionamento de árvores, cujo foco foi o Teorema 3.5 apresentado por Wang et. al [35], trouxe possíveis estratégias para demonstrar este resultado para o caso em que tratamos de partições com $k > 2$ classes. Uma das ideias de trabalhos futuros é conseguir provar o resultado do Teorema 3.7 para grafos com pesos nos vértices, levando a apresentar uma demonstração do Teorema 3.5 para o caso em que $k > 2$.

Também gostaríamos de estudar com mais detalhes algumas classes de grafos cuja melhor partição para um problema de particionamento possui propriedades que possibilitam que o problema possa ser resolvido com mais facilidade em relação ao problema para grafos no geral. Um exemplo onde isso ocorre é no estudo de Peng, Sun e Zanetti [30], que provam que um algoritmo espectral combinado com o algoritmo de classificação das k -médias tem garantia de aproximação quando aplicado a grafos *bem particionados*, que são aqueles que possuem uma razão $\Upsilon(k) = \frac{\lambda_{k+1}}{\phi_G(k)}$ elevada, onde λ_{k+1} é o $(k+1)$ -ésimo menor autovalor da matriz laplaciana normalizada do grafo G e $\phi_G(k)$ é a condutância de G de ordem k .

Devido à sua ampla utilização, seria interessante obter mais resultados sobre o desempenho do algoritmo da k -médias como passo de *clustering* de algoritmos espectrais e a influência do método de inicialização deste algoritmo.

Os problemas apresentados por Chung e Trevisan poderão ser abordados de diferentes formas daquelas estudadas neste trabalho, como a aplicação dos algoritmos espectrais em grafos aleatórios e a investigação se existem versões das desigualdades de Cheeger adaptadas para as funções RatioCut e NCut e, caso afirmativo, verificar se é possível obter algoritmos espectrais de bom desempenho a partir de suas demonstrações. Isso mostraria que estratégias como as descritas nas seções 4.2.2 e 4.3.3 também são algoritmos de aproximação. Além disso, outras desigualdades cuja demonstração também fornece algoritmos espectrais poderão ser estudadas. Seja o grafo G com n vértices e $\lambda_1 \leq \dots \leq \lambda_n$ os autovalores da sua matriz laplaciana normalizada, para um k fixo algumas destas desigualdades são $\phi_G(k) \leq C_1 \cdot k^2 \sqrt{\lambda_k}$ (apresentada por Lee, Gharan e Trevisan [22]), $\phi_G(k) \leq C_2 \cdot k \cdot \frac{\lambda_2}{\sqrt{\lambda_k}}$ (apresentada por Kwok, Lau, Lee, Gharan e Trevisan [20]) e $\max_{i=1, \dots, ck} \phi_G(S_i) \leq C_3 \cdot \sqrt{\lambda_k \log k}$ (apresentada por Louis, Raghavendra, Tetali e Vempala [26]), onde $c < 1$ e $C_1, C_2, C_3 > 0$ são constantes reais absolutas.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ANDERSON, W. N., AND MORLEY, T. D. Algebraic connectivity of graphs. *Linear and Multilinear Algebra* 18 (1985), 141–145.
- [2] ARORA, P., DEEPALI, D., AND VARSHNEY, S. Analysis of k-means and k-medoids algorithm for big data. *Procedia Computer Science* 78 (2016), 507–512.
- [3] BAJUSZ, D., RÁCZ, A., AND HÉBERGER, K. Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics* 7, 20 (2015), 1–13.
- [4] BANERJEE, A. *The Spectrum of the Graph Laplacian as a Tool for Analyzing Structure and Evolution of Networks*. PhD thesis, Leipzig, 2008.
- [5] BONDY, J. A., AND MURTY, U. S. R. *Graph theory, vol. 244 of Graduate Texts in Mathematics*, vol. 244. Springer, New York, 2008.
- [6] CHAKRABORTY, S., AND DAS, S. k-means clustering with a new divergence-based distance metric: Convergence and performance analysis. *Pattern Recognition Letters* 100 (2017), 67–73.
- [7] CHEEGER, J. A lower bound for the smallest eigenvalue of the laplacian. *Proceedings of the Princeton conference in honor of Professor S. Bochner* (1969), 195–199.
- [8] CHUNG, F. R. K. *Spectral Graph Theory*. No. 92. CBMS Regional Conference Series in Mathematics, 1994.
- [9] COVER, T. M., AND HART, P. E. Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory* 21, 3 (1967), 21–27.

- [10] COX, I. J., AND ZHONG, Y. Ratio regions: A technique for image segmentation. *Proc. 13th Int'l Conf. Pattern Recognition 2* (1996), 557–564.
- [11] DAMLE, A., MINDEN, V., AND YING, L. Robust and efficient multi-way spectral clustering. *Advances in Neural Information Processing Systems* (2016).
- [12] DE ABREU, N. M. M., DEL-VECCHIO, R. R., AND STEVANOVIĆ, D. Introdução à teoria espectral de grafos com aplicações.
- [13] DHILLON, I. S., GUAN, Y., AND KULIS, B. Kernel k-means, spectral clustering and normalized cuts. *Proc. 10th KDD 100* (2004), 551–556.
- [14] DOS SANTOS, D. F., DE MENDONÇA, L. F., AND TEIXEIRA, M. G. Um algoritmo de agrupamento heterogêneo para formação de grupos de aprendizagem. *Proceeding Series of the Brazilian Society of Applied and Computational Mathematics 3*, 1 (2015).
- [15] FIEDLER, M. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal 23*, 98 (1975), 298–305.
- [16] FRÄNTI, P., AND SIERANOJA, S. How much can k-means be improved by using better initialization and repeats? *Pattern Recognition 93* (2019), 95–112.
- [17] GUTMAN, I., ŠČIĆ, B., TRINAJSTIĆ, N., AND F. WILCOX, J. Graph theory and molecular orbitals. xii. acyclic polyenes. *The Journal of Chemical Physics 62*, 9 (1975).
- [18] HORN, R. A., AND JOHNSON, C. R. *Matrix Analysis*. No. 2 ed. Cambridge University Press, 1985.
- [19] JAMES, B. R. *Probabilidade: Um Curso em Nível Intermediário*. No. 3 ed. IMPA, Rio de Janeiro, 2011.

- [20] KWOK, T. C., LAU, L. C., LEE, Y. T., GHARAN, S. O., AND TREVISAN, L. Improved cheeger’s inequality: Analysis of spectral partitioning algorithms through higher order spectral gap. *In 45th Annual ACM Symposium on Theory of Computing (STOC’13)*, 11–20.
- [21] LANG, S. *Linear Algebra*. No. 3 ed. Springer, 1987.
- [22] LEE, J. R., GHARAN, S. O., AND TREVISAN, L. Multiway spectral partitioning and higher-order cheeger inequalities. *Journal of the ACM* 61, 6 (2014).
- [23] LIMA, E. L. *Análise Real*, vol. 1. Coleção Matemática Universitária, Rio de Janeiro, 1989.
- [24] LIU, H., ZHAO, F., AND JIAO, L. Fuzzy spectral clustering with robust spatial information for image segmentation. *Applied Soft Computing* 12, 11 (2012), 3636–3647.
- [25] LLOYD, S. P. Least squares quantization in pcm. *IEEE Transactions on Information Theory* 28, 2, 129–137.
- [26] LOUIS, A., RAGHAVENDRA, P., TETALI, P., AND VEMPALA, S. Many sparse cuts via higher eigenvalues. *In 44th Annual ACM Symposium on Theory of Computing (STOC’12)*, 1131–1140.
- [27] MACQUEEN, J. Some methods for classification and analysis of multivariate observations. *Fifth Berkeley Symposium* (1967), 282–297.
- [28] MAGNUS, J. R., AND NEUDECKER, H. *Matrix Differential Calculus With Applications in Statistics and Econometrics*. No. 3. John Wiley & Sons, Chichester, 1999.
- [29] NG, A. Y., JORDAN, M. I., AND WEISS, Y. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems* 14 (2002), 849–856.

- [30] PENG, R., SUN, H., AND ZANETTI, L. Partitioning well-clustered graphs: Spectral clustering works! *In Proceedings of The 28th Conference on Learning Theory 40*, 1423–1455.
- [31] SELIM, S. Z., AND ISMAIL, M. A. Least squares quantization in pcm. *IEEE Transactions on Pattern Analysis and Machine Intelligence 6*, 1, 81–87.
- [32] SHI, J., AND MALIK, J. Normalized cuts and image segmentation. *Transactions on Pattern Analysis and Machine Intelligence 22*, 8, 888–905.
- [33] TREVISAN, L. Lecture notes on graph partitioning, expanders and spectral methods, 2017.
- [34] VON LUXBURG, U. A tutorial on spectral clustering. *Statistics and Computing 17*, 4 (2007).
- [35] WANG, J., JIA, Y., HUA, X.-S., ZHANG, C., AND QUAN, L. Normalized tree partitioning for image segmentation. *In Conf. on Computer Vision and Pattern Recognition* (2008).
- [36] WANG, S., AND SISKIND, J. M. Image segmentation with ratio cut. *Transactions on Pattern Analysis and Machine Intelligence 25*, 6 (2003), 675–690.
- [37] WU, Z., AND LEAHY, R. Optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *Transactions on Pattern Analysis and Machine Intelligence 15*, 11 (1993), 1101–1113.
- [38] YARAM, S. Machine learning algorithms for document clustering and fraud detection. *2016 International Conference on Data Science and Engineering (ICDSE)* (2016).

- [39] ZENG, S., HUANG, R., KANG, Z., AND SANG, N. Image segmentation using spectral clustering of gaussian mixture models. *Neurocomputing* 144 (2014), 346–356.