

Relações entre particionamento e centralidade de árvores em função do vetor de Fiedler

Luciano Garim Garcia^I , Carlos Hoppen^{II} 

^IUniversidade do Vale do Rio dos Sinos, São Leopoldo, RS, Brasil

^{II}Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

RESUMO

O propósito deste trabalho é apresentar os conceitos de particionamento e centralidade de arestas para árvores usando o vetor de Fiedler. A partir disso, descrevemos duas heurísticas espectrais. Finalmente, a partir da classificação de árvores do tipo I e tipo II, mostramos em que situações essas heurísticas retornam a mesma partição.

Palavras-chave: Particionamento; Centralidade; Vetor de Fiedler

ABSTRACT

The purpose of this work is to present the concepts of partitioning and centrality of edges for trees using the Fiedler vector. From this, we describe two spectral heuristics. Finally, from the classification of type I and type II trees, we show in which situations these heuristics return the same partition.

Keywords: Partitioning; Centrality; Fiedler Vector

1 INTRODUÇÃO

A área de pesquisa relacionada ao particionamento de grafos vem crescendo nos últimos anos devido a modelos de agrupamento de dados que utilizam grafos como representação de uma rede. Métodos que utilizam o espectro de grafos e seus autovetores associados têm sido amplamente utilizados pela comunidade

científica para esta finalidade Fortunato (2010), Von Luxburg (2007) e Ng *et al.* (2002).

Neste trabalho trataremos de um caso específico de grafos, as árvores. Neste sentido, usaremos o segundo menor autovetor da matriz laplaciana da árvore para descrever heurísticas que particionam a árvore em duas classes. Estas heurísticas serão descritas para dois tipos de árvores, segundo a classificação de Fiedler.

Introduziremos o conceito de centralidade de arestas baseado na numeração característica, e usaremos este conceito para obter uma nova heurística de particionamento. Finalmente, mostraremos que as classes geradas por essa heurística e a heurística de particionamento que usa o sinal do vetor de Fiedler são equivalentes em árvores do tipo II. Para árvores do tipo I, a partição obtida pela centralidade é a mesma obtida pela heurística de Fiedler se as classes geradas são conexas, caso contrário, as partições são distintas.

2 PARTICIONAMENTO DE ÁRVORES

Em geral, o problema de particionamento de grafos têm por objetivo particionar um conjunto de vértices em classes de acordo com algum critério pré-estabelecido. Para modelar esse problema matematicamente é preciso definir funções que medem a qualidade de uma partição reescrevendo-o como um problema de otimização. O domínio dessas funções é o conjunto de todas as partições de V em k classes será denotado por $\mathcal{P}_G(k)$. A Definição 1, formaliza o problema de particionamento de grafos.

Definição 1. Sejam $G = (V, E)$, $k \geq 2$ 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 por objetivo encontrar $P^* \in \mathcal{P}_G(k)$ tal que

$$F(P^*) = \min_{P \in \mathcal{P}_G(k)} F(P). \tag{1}$$

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

Trabalhos de Wei e Cheng (1989) e Shi e Malik (2000), apresentam funções F que tem por objetivo encontrar uma partição de G com restrições em relação a magnitude das classes. Em ambos trabalhos, encontrar uma partição P^* que satisfaz (1) é um problema NP-Difícil. A partir disso, Chan *et al.* (1994) e Shi e Malik (2000) usam uma formulação espectral por meio de matrizes associadas ao grafo para obter soluções aproximadas para $F(P^*)$. A seguir definimos a matriz que usaremos neste trabalho.

Seja $G = (V, E)$ um grafo com um conjunto de vértices $V = \{v_1, \dots, v_n\}$ e um conjunto de arestas E . Denotando por $d(v)$ o grau do vértice v , definimos a matriz laplaciana $L(G) = (a_{ij})$ por:

$$a_{ij} = \begin{cases} -1, & \text{se } \{v_i, v_j\} \in E \\ d(v_i), & \text{se } i = j \\ 0, & \text{c.c.} \end{cases}$$

A matriz $L = L(G)$ é simétrica positiva semidefinida, e portanto seus autovalores são não-negativos. Alternativamente, podemos escrever $L = D - A$, onde D é uma matriz diagonal com o grau dos vértices e A é a matriz de adjacências de G . Sejam $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ os autovalores de L . Como a soma dos elementos de uma linha qualquer de L é zero segue que $L \mathbf{1} = 0 \mathbf{1}$, e a multiplicidade do autovalor λ_1 é igual ao número de componentes conexas do grafo. Quando G é um grafo conexo, L é irredutível e então $\lambda_2 = 0$. O autovalor λ_2 denotado por $\alpha(G)$ é denominado conectividade algébrica, e seu autovetor associado y é conhecido como vetor de Fiedler.

A fim de obter propriedades que relacionam o vetor de Fiedler com árvores, rotulamos os vértices de $T = (V, E)$ de modo que a entrada y_i de \mathbf{y} numere o vértice v_i . Essa associação é conhecida como numeração característica. Desde o trabalho de Pothen *et al.* (1990) essa numeração vem sendo utilizada como ferramenta para particionar grafos. O amplo uso dessas heurísticas se deve ao fato de que o vetor sugere uma partição natural de vértices de acordo com o sinal de suas

componentes Fiedler (1975), tendo a garantia de que pelo menos uma das classes é conexa. Em relação a árvores podemos obter duas caracterizações distintas para a distribuição dessa numeração, e a partir dessa distinção podemos olhar para o problema de particionamento de árvores sob duas perspectivas.

Teorema 1 (Teorema da Monotonicidade de Fiedler). Seja T uma árvore, L sua matriz laplaciana com autovetor y associado ao autovalor λ_2 . Então dois casos podem ocorrer:

Caso A: $y_i \neq 0 \forall i \in V$. Então T contém exatamente uma aresta $e = \{v_p, v_q\}$ tal que $y_p > 0$ e $y_q < 0$. A numeração dos vértices ao longo de qualquer caminho em T que inicia em v_p e não contém v_q cresce, e a numeração dos vértices ao longo de qualquer caminho em T que inicia em v_q e não contém v_p decresce.

Caso B: O conjunto $S = \{v_i \in V, |y_i| = 0\} \neq \emptyset$. Então o grafo $T[S]$ é conexo e existe exatamente um vértice $v_z \in S$ que possui no mínimo um vizinho que não pertence a S . A numeração ao longo de qualquer caminho em T que começa em v_z é crescente, decrescente ou zero.

No Teorema da Monotonicidade de Fiedler (1975), quando o caso B ocorre, a árvore é classificada do tipo I, e o vértice v_z é chamado de vértice característico; quando o caso A ocorre, a árvore é do tipo II e os vértices v_p e v_q são chamados de vértices característicos. A seguir descrevemos a heurística espectral que usa o sinal do vetor de Fiedler.

Algoritmo 1: Heurística de particionamento de Fiedler para árvores

Entrada: Matriz Laplaciana de T

- (1) Compute o vetor de Fiedler;
- (2) Particione os vértices de T de acordo com:
 $S = \{v_i \in V, |y_i| \leq 0\}$.

Saída: Classes S e \bar{S} , onde \bar{S} é o complementar de S .

Para árvores do tipo II, as duas classes obtidas pela heurística são conexas, e os vértices de cada classe possuem o mesmo sinal. Para árvores do tipo I, apesar

das classes possuírem vértices com numeração de mesmo sinal, a partição obtida pode não gerar duas classes conexas. Trabalhos como de Kannan *et al.* (2004) e Guattery e Miller (1998) estudam a qualidade das partições geradas por estas heurísticas para grafos em geral. Ressaltamos que neste trabalho não nos dedicaremos a analisar a qualidade destas partições, mas sim em obter relações entre elas.

2.1 Centralidade de Arestas para Árvores

Em certos contextos é interessante determinar a importância que uma aresta possui em relação as demais. As medidas de centralidade são parâmetros usados para quantificar esta importância. A definição a seguir formaliza esse conceito.

Definição 2. Seja $G = (V, E)$ um grafo. A centralidade de arestas de G é dada por uma função $C : E(G) \rightarrow \mathbb{R}_+$, onde para cada $e = \{v_i, v_j\} \in E$ existe um valor real não negativo associado. Dizemos que uma aresta $e = \{v_i, v_j\}$ possui maior centralidade que uma aresta $f = \{v_j, v_k\}$ se $C_{ij} > C_{jk}$.

Inicialmente trataremos de centralidade para grafos e nos restringiremos a árvores logo após. Para motivar a nossa definição de centralidade, veremos o que ocorre com a conectividade algébrica quando uma aresta é removida. Dado um grafo $G = (V, E)$, considere $\tilde{G} = G - e$ o grafo obtido pela remoção de uma aresta $e = \{v_i, v_j\}$. Denote a laplaciana resultante por $\tilde{L}(i, j)$.

Sabendo que $\lambda_2(L) = \min_{\|x\|^2=1, x \perp \mathbf{1}} x^t L x$ Gould (2012). Denote y como vetor de Fiedler de L , calculando $y^t \tilde{L}(i, j) y$ nos dá uma cota superior para $\lambda_2(\tilde{L}(i, j))$:

$$\lambda_2(\tilde{L}(i, j)) \leq y^t \tilde{L}(i, j) y = \lambda_2(L) - (y_i - y_j)^2.$$

Vale ressaltar que para qualquer grafo conexo G , existe no mínimo uma remoção de aresta tal que $\lambda_2(\tilde{L}(i, j)) < \lambda_2(L)$, caso contrário, $y_i = y_j, \forall i, j \in V$ e isso violaria as restrições $\|y\|^2 = 1$ e $\sum_i^n y_i = 0$. Consequentemente, existe no mínimo

uma remoção de aresta que nos leva a um decréscimo na conectividade algébrica Chen e Hero (2014).

Motivado pelo fato do decréscimo na conectividade algébrica quando removemos uma aresta $\{v_i, v_j\}$, definiremos a centralidade de arestas do vetor de Fiedler, usando a numeração característica associada ao termo $(y_i - y_j)^2$.

A definição a seguir formaliza esse conceito.

Definição 3. Seja y o vetor de Fiedler de um grafo G . A centralidade de arestas do vetor de Fiedler é dada por uma função $C : E(G) \rightarrow \mathbb{R}_+$, onde para cada $e = \{v_i, v_j\} \in E$ existem y_i e y_j componentes de y tal que $C_{ij} = (y_i - y_j)^2$. Dizemos que uma aresta $e = \{v_i, v_j\}$ possui maior centralidade que uma aresta $f = \{v_j, v_k\}$ se $C_{ij} > C_{jk}$.

No caso de árvores, podemos usar a remoção da aresta de maior centralidade para obter uma partição de T . Desse modo, propomos a seguinte heurística:

Como a heurística acima remove apenas uma aresta, temos a garantia de que as classes geradas são sempre conexas. Mas isso não garante que a numeração característica tenha o mesmo sinal em uma classe. A seguir veremos resultados que comparam as heurísticas descritas aqui.

Algoritmo 2: Heurística de particionamento por centralidade

Entrada: Matriz Laplaciana de T

- (1) Compute o vetor de Fiedler;
- (2) Encontre a aresta $\{v_i, v_j\}$ que maximiza C_{ij} ;
- (3) Remova a aresta $\{v_i, v_j\}$.

Saída: Classes S e S^c conexas.

3 METODOLOGIA E RESULTADOS

Nesta seção estamos interessados em mostrar que a heurística de particionamento por centralidade gera uma partição na árvore removendo a aresta

incidente a vértices característicos em árvores do tipo II. Analogamente, para árvores do tipo I a partição obtida provém da remoção da aresta de maior centralidade que é incidente ao vértice característico. Para estas finalidades usaremos um resultado técnico, obtido de Bapat (2010), que nos permitirá construir as provas para os dois tipos de árvores.

Lema 1. Seja T uma árvore, L sua matriz laplaciana com autovetor y associado ao autovalor λ_2 , e $e = \{v_i, v_j\}$ uma aresta. Então:

$$y_i - y_j = -\lambda_2 \sum_k y_k,$$

onde a soma é tomada para todos os vértices $k \in C_e(v_j)$, onde $C_e(v_j)$ é a componente de $T - e$ que contém v_j .

Demonstração. Assuma que, após uma reordenação, a aresta e possui extremidades s e $s+1$, além disso, que as duas componentes de $T - e$ possuem os conjuntos de vértices $\{v_1, \dots, v_s\}$ e $\{v_{s+1}, \dots, v_n\}$, respectivamente. Seja u um vetor de ordem $n \times 1$, com $u_i = 1, i = 1, \dots, s$ e $u_i = 0, s + 1, \dots, n$. Note que $u^t L = [0, \dots, 0, 1, -1, 0, \dots, 0]$, onde 1 e -1 ocorrem nas posições de s e $s + 1$, respectivamente. Assim, $u^t L y = y_s - y_{s+1}$. Consequentemente, de $u^t L y = \lambda_2 u^t y$, concluímos que:

$$y_s - y_{s+1} = \lambda_2 \sum_{k=1}^s y_k.$$

Como y é ortogonal a $\mathbf{1}$, a expressão do lado direito é igual a $-\lambda_2 \sum_{k=s+1}^n y_k$.

A partir do uso do Lema 1 provaremos o Teorema 2, nos restringindo a árvores do tipo II.

Teorema 2. Seja T uma árvore do tipo II, L sua matriz laplaciana com autovetor y associado ao autovalor λ_2 , então:

$$\max_{\{v_i, v_j\} \in E(T)} (y_i - y_j)^2.$$

é atingido quando v_i e v_j são os vértices característicos de T . Consequentemente, a partição gerada pelo algoritmo de Fiedler é a mesma gerada pelo algoritmo da Centralidade.

Demonstração. Seja $e = \{v_r, v_s\}$ a aresta incidente aos vértices característicos v_r e v_s . Note que $T - e$ gera duas componentes $C_e(v_r)$ e $C_e(v_s)$. Pelo Teorema 1 caso A, a numeração característica em cada uma dessas componentes possui o mesmo sinal. Pelo Lema,

$$(y_r - y_s)^2 = \lambda_2^2 \left[\sum_{k \in C_e(v_r)} y_k \right]^2. \quad (2)$$

note que a soma (2) contém todas as entradas de y que possuem o mesmo sinal, pois numeram os vértices em $C_e(v_r)$. Afirmamos que a centralidade de $\{v_r, v_s\}$ é a máxima. Caso não fosse, existiria um aresta $f = \{v_t, v_u\}$ tal que $T - f$ geraria duas componentes $C_e(v_t)$ e $C_e(v_u)$ e pelo Lema 1,

$$(y_t - y_u)^2 = \lambda_2^2 \left[\sum_{k \in C_e(v_t)} y_k \right]^2. \quad (3)$$

Mas a soma (3) é sempre menor que a soma (2), pois na componente $C_e(v_t)$ podem haver vértices com numeração positiva e negativa, ou pode não incluir todos os vértices com numeração característica de mesmo sinal.

Como observado no Teorema 2, a aresta de centralidade máxima é justamente a aresta que é removida pela heurística de Fiedler. Consequentemente, existe uma equivalência entre as partições obtidas e, isso ocorre quando tratamos de árvores do tipo II.

A partir do uso do Lema 1 provaremos o Teorema 3, nos restringindo a árvores do tipo I.

Teorema 3. *Seja T uma árvore do tipo I, L sua matriz laplaciana com autovetor y associado ao autovalor λ_2 , então:*

$$\max_{\{v_i, v_j\} \in E(T)} (y_i - y_j)^2, \quad (4)$$

é atingido quando v_i ou v_j é o vértice característico de T .

Demonstração. Seja v_r o vértice característico de T . Se $T - \{v_r\}$ possui p componentes, então existem p arestas incidentes a v_r . Pelo Teorema 1, cada componente possui numeração característica com mesmo sinal. Então, para cada aresta incidente a v_r e a um vizinho v_s , pelo Lema 1, temos:

$$(y_r - y_s)^2 = \lambda_2^2 \left[\sum_{k \in C_e(v_s)} y_k \right]^2. \quad (5)$$

Se v_s possui numeração zero, então a contribuição de $C_e(v_s)$ para a soma (5) é zero, logo não é máxima. Se v_s possui numeração não-nula, então a soma (5) inclui todos os vértices com numeração característica de mesmo sinal nesta componente.

Analogamente, para as outras $p - 1$ componentes, as arestas incidentes a v_r possuem maior centralidade que as demais. Conseqüentemente, entre as p arestas de maior centralidade em T , existe uma que maximiza (4) e é incidente a v_r .

Quando a heurística de Fiedler gera classes conexas, essas classes são as mesmas obtidas pelo heurística da centralidade. Entretanto, se as classes obtidas pela heurística de Fiedler são desconexas, as duas heurísticas diferem em relação a partição obtida. Essas partições distintas ocorrem quando tratamos de árvores do tipo I, e o vértice característico possui mais de uma componente com o mesmo sinal.

4 CONCLUSÕES

O presente trabalho teve por objetivo analisar as partições geradas por heurísticas espectrais que usam o vetor de Fiedler. Mostramos que é possível obter as mesmas classes para as duas heurísticas quando tratamos de árvores do tipo II. Para árvores do tipo I, vimos a igualdade das partições depende das classes serem conexas na heurística de Fiedler.

Atualmente, estamos desenvolvendo o conceito de centralidade baseado na numeração característica para grafos mais gerais. Esperamos ser possível, novamente, mostrar alguma equivalência entre algumas partições obtidas, agora usando o Teorema da Monotonicidade de Fiedler para grafos. Além disso, seria interessante investigar como usar a centralidade para obter um k particionamento do grafo, e a partir disso, analisar a qualidade do particionamento obtido comparando com outras heurísticas existentes.

5 REFERÊNCIAS

BAPAT, R. B. (2010). **Graphs and matrices**, vol 27. Springer.

CHAN, P. K., SCHLAG, M. D., ZIEN, J. Y. (1994). **Spectral k -way ratio-cut partitioning and clustering**. IEEE Transactions on computer-aided design of integrated circuits and systems, 13(9), 1088–1096.

CHEN, P. Y., HERO, A. O. (2014). **Local fiedler vector centrality for detection of deep and overlapping communities in networks**. Em: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp. 1120–1124.

FIEDLER, M. (1975). A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. **Czechoslovak Mathematical Journal**, 25(4), 619–633.

FORTUNATO, S. (2010). Community detection in graphs. **Physics reports**, 486(3-5), 75–174.

GOULD, S. H. (2012). **Variational methods for eigenvalue problems: an introduction to the methods of Rayleigh, Ritz, Weinstein, and Aronszajn**. Courier Corporation.

GUATTERY, S., MILLER, G. L. (1998). On the quality of spectral separators. **SIAM Journal on Matrix Analysis and Applications**, 19(3), 701–719.

KANNAN, R., VEMPALA, S., VETTA, A. (2004). On clusterings: Good, bad and spectral. **Journal of the ACM (JACM)**, 51(3), 497–515.

NG, A. Y., JORDAN, M. I., WEISS, Y., *et al.* (2002). **On spectral clustering: Analysis and an algorithm**. **Advances in neural information processing systems**, 2, 849–856.

POTHEN, A., SIMON, H. D., LIOU, K. P. (1990). Partitioning sparse matrices with eigenvectors of graphs. **SIAM journal on matrix analysis and applications**, 11(3), 430–452.

SHI, J., MALIK, J. (2000). **Normalized cuts and image segmentation**. IEEE Transactions on pattern analysis and machine intelligence, 22(8), 888–905.

VON LUXBURG, U. (2007). **A tutorial on spectral clustering**. Statistics and computing, 17(4), 395–416.

WEI, Y. C., CHENG, C. K. (1989). **Towards efficient hierarchical designs by ratio cut partitioning**. Em: 1989 IEEE International Conference on Computer-Aided Design. Digest of Technical Papers, IEEE, pp. 298–301.

CONTRIBUIÇÃO DE AUTORIA

1 – Luciano Garim Garcia:

Doutorando em Matemática Aplicada no Programa de Pós Graduação em Matemática Aplicada da Universidade Federal do Rio Grande do Sul.

<https://orcid.org/0000-0003-0990-0999> - lucianogarim@gmail.com

Contribuição: Conceituação | Análise Formal | Investigação | Metodologia | Validação | Escrita –primeira redação.

2 – Carlos Hoppen:

Professor Associado do Departamento de Matemática Pura e Aplicada na Universidade Federal do Rio Grande do Sul.

<https://orcid.org/0000-0002-7581-1583> - choppen@ufgrs.br

Contribuição: Conceituação | Análise Formal | Validação | Escrita – revisão e edição.

Como citar este artigo

GARCIA, L. G.; HOPPEN, C. Relações entre particionamento e centralidade de árvores em função do vetor de Fiedler. **Ciência e Natura**, Santa Maria, v. 43, Ed. Esp. X ERMAC, e18, p. 1-11, 2021. DOI 10.5902/2179460X64895. Disponível em: <https://doi.org/10.5902/2179460X64895>. Acesso em: 5 nov. 2021.