Métodos Espectrais de Partição de Grafos

Guilherme Barbosa dos Santos
Licenciatura em Matemática - UTFPR
guilhermebds360@gmail.com

Prof. João Luis Gonçalves
Departamento Acadêmico de Matemática - UTFPR
jlgoncalves@utfpr.edu.br

Palavras-chave: grafos, comunidades, partição.

Resumo:

A Teoria de Grafos tem grande potencial para a descrição e modelagem matemática de fenômenos associados a redes. Em problemas dessa natureza uma dificuldade recorrente é o tamanho da rede, ou seja o número de vértices e arestas. Para contornar essas dificuldades uma alternativa é decompor o grafo associado ao problema, ou ainda agrupar os vértices com propriedades semelhantes, gerando assim um novo grafo com menos vértices mas que ainda representa bem o fenômeno. Esses agrupamentos de vértices chamaremos de comunidades.

Outro aspecto que não pode ser negligenciado é a boa apresentação dos dados que os grafos oferecem. Nesse sentido a detecção de comunidades tem papel de sintetizar ainda mais essas informações.

Contudo a detecção de comunidades raramente pode ser feita empiricamente, principalmente para grafos grandes. Portanto, faz-se necessário um tratamento analítico do grafo, com rigor matemático. Um tratamento matematicamente rigoroso é um ponto positivo pois exigirá o uso de teorias mais desenvolvidas, como a álgebra linear por exemplo. Assim teremos uma quantidade maior de ferramentas para abordar um tema que não nos é tão familiar.

Este trabalho tem como objetivo apresentar a maximização da modularidade, usando análise espectral do grafo, como uma alternativa a partição ou decomposição de grafos.

A modularidade é a função que mede a diferença entre número de arestas dentro da comunidade e o número esperado de tais arestas, para duas ou mais comunidades. Mais precisamente, podemos definir a modularidade como a função Q, descrita a seguir

$$Q = \frac{1}{2m}\sum_{i,j}[A_{ij} - P_{ij}]\delta_{g_i,g_j},$$

em que

$$\delta_{g_i,g_j} = \begin{cases} 1, & \textit{se i e j estão nos mesmos grupos} \\ 0, & \textit{caso contrário,} \end{cases}$$

g_i é a comunidade a qual o vértice i pertence e m é o número de arestas da rede/grafo.

Sobre P_{ij} temos diferentes opções a considerar e cada opção representará um significado diferente para o termo comunidade. Contudo, P_{ij} deve indicar a probabilidade de existir uma aresta entre os vértices i e j. As duas escolhas mais intuitivas para P são P_{ij} um número randômico entre 0 e 1 e $P_{ij} = \frac{K_i K_j}{2m}$, em que k_i é o grau do vértice i, isto é o número de arestas que tem o vértice i como um de seus extremos. Contudo, é conveniente impor que

$$\sum_{ij} P_{ij} = \sum_{ij} A_{ij} = 2m \ , \ \ P_{ij} = P_{ji} \ \text{ e } \ \sum_{j} P_{ij} = k_i. \tag{1}$$

A é a matriz, sinétrica, de adjacência do grafo em questão, cujas entradas são definidas como

$$A_{ij} = \begin{cases} 1, & \textit{se existe uma aresta entre os vértices i e j} \\ 0, & \textit{caso contrário.} \end{cases}$$

Inicialmente vamos tratar da decompor o grafo em duas comunidades, G_1 e G_2. Para isso, vamos definir um vetor de escolha s em que:

$$s_i = \begin{cases} 1, & \textit{se i pertence a } G_1 \\ -1, & \textit{se i pertence a } G_2. \end{cases}$$

Observe que

$$\delta(g_i, g_j) = \frac{1}{2}(s_i s_j + 1),$$

e assim, podemos reescrever a função modularidade como

$$\begin{aligned} Q &= \frac{1}{4m}\sum_{i,j}[A_{ij} - P_{ij}](s_i s_j + 1) \\ &= \frac{1}{4m}\sum_{i,j}[A_{ij} - P_{ij}]s_i s_j + \frac{1}{4m}\sum_{i,j}[A_{ij} - P_{ij}] \\ &\underset{(\textbf{??})}{=} \frac{1}{4m}\sum_{i,j}[A_{ij} - P_{ij}]s_i s_j \end{aligned}$$

Agora, vamos escrever a função modularidade como

$$Q = \frac{1}{4m}\mathbf{s}^T\mathbf{B}\mathbf{s},$$

em que $B_{ij} = A_{ij} - P_{ij}$ é a matriz de modularidade. Observe que $\mathbf{B}$ é simétrica, então $\mathbf{B}$ tem autovalores reais e autovetores associados a autovalores distintos são ortogonais. Sempre é possível escolher uma base ortonormal para o espaço gerado por autovetores associados a autovalores iguais de $\mathbf{B}$.

Sejam $\mathbf{u}_i$ os autovetores normalizados da matriz $\mathbf{B}$. Podemos escrever s como

$$\mathbf{s} = \sum_{i=1}^{n} a_i \mathbf{u}_i,$$

com $a_i = \mathbf{u}_i^T \mathrm{s}$ e assim

$$Q = \frac{1}{4m}\sum_{i} a_i^2 \beta_i,$$

em que β_i é o autovalor associado ao autovetor u_i.

Note que Q assumiria o valor máximo se s fosse paralelo ao autovetor associado ao maior autovalor.Vamos chamar esse autovetor de $\mathbf{u}^{(1)}$. Contudo, em geral isso não é possível, pois s é um vetor que possui entradas 1 ou -1 apenas.

Assim, a melhor escolha possível para s é tal que

$$s_i = \begin{cases} +1, & \textit{se}\ \mathbf{u}_i^{(1)} \geq 0, \\ -1, & \textit{se}\ \mathbf{u}_i^{(1)} < 0. \end{cases}$$

Já temos resultados preliminares para alguns problemas padrão de detecção de comunidades, contudo estamos estudando novas possibilidades para P e uma função que represente uma modularidade de segunda ordem.

Referências

[1] NEWMAN, M.E.J., Finding community structure in networks using eigenvectors of matrices. **arXiv**, arXiv:physics/0605087v3, 2006.

[2] SCHAUB, M.T., DELVENNE, J.-C., ROSVALL, M., LAMBIOTTE, R., The many facets of community detection in complex networks, **Applied Network Science**. 2:4. 2017.

[3] MESBAHI, M., EGERSTEDT, M., Graph theoretic methods in multiagent networks, Princenton University Press, 2010.