

HEURÍSTICA DE GERAÇÃO DE COLUNAS PARA DETECÇÃO DE COMUNIDADES POR DENSIDADE

Rafael de Santiago, Ronaldo Rafael Etur de Moraes, Augusto C. Pluschkat

¹Laboratório de Inteligência Aplicada – Universidade do Vale do Itajaí (UNIVALI)
Itajaí – SC – Brasil

rsantiago@univali.br, rafaetur@gmail.com, acpluchkat@hotmail.com

***Abstract.** Modularity Density Maximization is an optimization problem with the objective of solving non-overlapping instances of the clustering problem know as Community Detection in networks without having the resolution limit of the Modularity Maximization method. In the literature, it is not know of any linear programming model that can give the exact solution for instances with more than 40 vertices. This limitation occurs because of the high number of variables needed to solve the problem. There is a mathematical optimization method know as Column Generation that can be applied to linear optimization problems to reduce the number of variables needed to solve it. Thus, the objective of this work is to apply the Column Generation heuristic to the Modularity Density optimization problem to solve instances with more than 40 vertices.*

1. Introdução

Um dos problemas que costumam ser abordados no uso de grafos é seu particionamento em grupos conhecidos como comunidades, onde os vértices apresentam forte ligação. Tal problema é denominado na literatura como o problema de detecção de comunidades. Um dos principais métodos para quantificar a qualidade das comunidades é a Máximização de Modularidade desenvolvida por Newman e Girvan (2004), sendo considerado como um problema de otimização NP-Difícil [Brandes et al. 2008].

Fortunato e Barthélemy (2007) reportaram uma importante degeneração para este problema, conhecida como o limite de resolução. Tal degeneração é caracterizada pelo fato de soluções ótimas para o problema não detectarem comunidades de tamanhos pequenos. Para resolver o problema da degeneração, Li et al. (2008) desenvolveram uma alternativa para a Máximação de Modularidade denominada de *Modularity Density*.

No trabalho de Costa (2015), através do uso de programação matemática, foi reportado um algoritmo exato para o problema de *Modularity Density*. No entanto, foi possível chegar à solução ótima em apenas instâncias com até 40 vértices por conta do grande número de variáveis geradas na linearização do modelo de Li et al. (2008). Para lidar com essa limitação, Costa sugeriu a aplicação de uma técnica de programação matemática conhecida como Geração de Colunas frente ao problema de *Modularity Density*.

A técnica de Geração de Colunas foi proposta por Dantzig e Wolfe (1960) para resolver problemas matemáticos com grande quantidade de variáveis, possibilitando que seja aplicado em problemas lineares de grandes dimensões. Aloise (2010) aplicou a técnica com sucesso ao problema de Máximização de Modularidade. O método funciona dividindo o problema em um problema mestre e um problema auxiliar. O problema mestre representa o

problema a ser resolvido, e possui um grande número de variáveis, representadas por colunas. Para resolver o problema de maneira eficaz, o problema mestre é reduzido e solucionado, sendo a solução desse problema reduzido utilizada em um problema auxiliar através de técnicas de programação linear. O problema auxiliar gera colunas que são adicionadas ao problema mestre reduzido caso haja uma melhora na solução, assim se aproximando da solução do problema mestre inicial. Esses passos são iterados até que não se possa encontrar colunas que melhorem a solução.

2. Solução Proposta

Inspirado pelo trabalho de Costa (2015) e pelo modelo de Geração de Colunas de Aloise (2010), este projeto visa desenvolver e aplicar a técnica de Geração de Colunas para o problema de detecção de comunidades conhecido como *Modularity Density*, tendo como objetivo conseguir encontrar a solução ótima para instâncias com mais de 40 vértices. Pretende-se avaliar o número de variáveis, tempo e qualidade das soluções obtidas e comparar os resultados com outros modelos de programação matemática encontrados na literatura sobre o problema de *Modularity Density*.

Pode-se definir os seguintes passos para atingir o objetivo do trabalho: (i) criar o modelo de geração de colunas; (ii) desenvolver uma heurística para geração de variáveis em cada iteração; (iii) avaliar parâmetros e componentes do Gerador de Colunas, tais como o problema auxiliar e problema mestre reduzido; e (iv) dar publicidade aos resultados obtidos.

3. Considerações Finais

O projeto encontra-se em sua fase inicial, tendo sido finalizada a implementação básica do modelo utilizado no algoritmo de Geração de Colunas. Ainda é necessário aprimorá-lo para que seja possível passar do limite de 40 vértices e cumprir o objetivo de pesquisa, tendo o foco concentrado na busca de uma solução heurística inicial para o algoritmo e em heurísticas para diminuir o tempo necessário para encontrar colunas que melhorem a solução. Caso o projeto obtenha sucesso, será possível avançar o estado da arte da literatura na área de detecção de comunidades através de soluções exatas no problema de otimização de *Modularity Density*.

Referências

- Aloise, D. et al. (2010). Column generation algorithms for exact modularity maximization in networks. In *Physical Review E*, page 046112. American Physical Society.
- Brandes, U. et al. (2008). On Modularity Clustering. In *IEEE Transactions on Knowledge and Data Engineering*, pages 172–188. IEEE Computer Society.
- Costa, A. (2015). MILP formulations for the modularity density maximization problem. In *European Journal of Operational Research*, pages 14–21. Elsevier B.V.
- Dantzig, G. B. and Wolfe, P. (1960). Decomposition Principle for Linear Programs. In *Operations Research*, pages 101–111. INFORMS.

Fortunato, S. and Barthélemy, M. (2007). Resolution limit in community detection. In *Proceedings of the National Academy of Sciences of the United States of America*, pages 36–41. National Academy of Sciences.

Li, Z. et al. (2008). Quantitative function for community detection. In *Physical Review E*, page 036109. American Physical Society.

Newman, M. and Girvan, M. (2004). Finding and evaluating community structure in networks. In *Physical Review E*, page 026113. American Physical Society.