

Rede Neural para a Detecção de Comunidades em Grafos

Rafael de Oliveira Schmitt, Rafael de Santiago

Laboratório de Inteligência Aplicada – Universidade do Vale do Itajaí (UNIVALI)
Caixa Postal 360 – 88.302-202 – Itajaí – SC – Brasil

`schmittrafa@gmail.com.br, rsantiago@univali.br`

***Abstract.** In the literature, there are several approaches to the problem of identifying communities in graphs, due to their importance in a wide range of scientific fields. A less explored approach is the application of expert systems, which can be trained to recognize patterns. This work proposes the use of artificial neural networks to recognize the patterns that characterize the communities, allowing their identification.*

1. Introdução

O estudo de comunidades em grafos tem um grande impacto em diversas disciplinas, uma vez que ele nos permite entender o comportamento de grupos de animais e pessoas, assim como associações de palavras e interações entre proteínas [Fortunato 2010].

Domínios da inteligência artificial apresentam grande influência no estudo sobre comunidades em grafos, como pode ser observado nos seguintes trabalhos [Tian et al. 2015, Xie 2013, Yang et al. 2016]. Grande foco desse estudo está em como detectar comunidades, tarefa essa que pode ser abordada de várias maneiras, visto que não há definições concretas, resultando em várias maneiras de definir comunidades pertinentes [Fortunato 2010].

2. Solução Proposta

Este trabalho propõe aplicar redes neurais para os fins de detecção de comunidades, fazendo uso de sua propriedade de reconhecimento de padrões [Haykin 1998]. A rede definida trabalhará com a identificação de arestas entre comunidades, indicando as arestas que podem ser removidas, necessitando de um pós-processamento para a concretização das comunidades formadas.

Para os treinamentos e testes, será utilizado um algoritmo de *benchmark* para a geração de grafos com estruturas de comunidades [Lancichinetti et al. 2008], onde serão testados a necessidade de cada entrada modelada para a rede, assim como, a capacidade da rede trabalhar com grafos de diferentes densidades e tamanhos.

Para a validação dos resultados, o algoritmo de *benchmark* disponibiliza um arquivo de comunidades o qual será utilizado como o resultado esperado presente no treino supervisionado, e será utilizado a correlação de Matthews para a quantificação das previsões corretamente realizadas pela rede neural.

3. Resultados Preliminares

Foi implementada uma rede neural alimentada adiante (*feedforward*) empregando o algoritmo de retropropagação (*backpropagation*) para realizar o aprendizado da rede.

A rede implementada foi modelada para identificar as arestas que se encontram entre as comunidades, devido a simplicidade e generalização das características das arestas. Para isso foram necessárias cinco entradas referentes ao par de vértices, sendo elas, o grau de cada vértice, a similaridade de Jaccard e o número de caminhos alternativos entre os vértices, sendo esses caminhos limitados a distâncias de 2 e 3 arestas. O treinamento foi de acordo com o parâmetro μ , presente no algoritmo *benchmark*, o qual indica a densidade do grafo.

Os testes preliminares indicam que a rede é incapaz de trabalhar com grafos que possuam densidade μ igual ou maior que 0,5, visto que a identificação de estruturas de comunidades só faz sentido em grafos esparsos [Fortunato 2010]. A rede também é limitada no sentido de ser incapaz de identificar comunidades sobrepostas, devido ao seu aspecto de trabalhar com arestas.

Nota-se também que os resultados melhoraram conforme o tamanho da rede aumentava, porém faltam testes para comprovar essa característica.

4. Considerações Finais

A possibilidade do uso de redes neurais para identificar comunidades em grafos é tentadora, por suas capacidades de generalização de informações. Graças a isso, podemos utilizar uma rede simples em diferentes grafos sem a necessidade de refazer o treinamento ou de alterar completamente a topologia da rede.

Grafos esparsos possuem comunidades bem definidas, tornando possível a utilização da rede neural, porém a geração de alguns dos parâmetros de entradas da rede são computacionalmente custos, fazendo-se necessária a validação da importância de cada parâmetro de entrada.

Para a continuidade no trabalho, é pretendida a validação das características da rede como: tamanhos de grafos que podem ser trabalhados, assim como a importância de cada parâmetro de entrada.

Referências

- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3-5):75–174.
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition.
- Lancichinetti, A., Fortunato, S., and Radicchi, F. (2008). Benchmark graphs for testing community detection algorithms. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 78(4):1–6.
- Tian, F., Gao, B., Cui, Q., Chen, E., and Liu, T.-y. (2015). Learning Deep Representations for Graph Clustering. *Cvpr*, pages 1293–1299.
- Xie, J. (2013). Overlapping Community Detection in Networks : The State-of-the-Art. 45(4).
- Yang, L., Cao, X., He, D., Wang, C., Wang, X., and Zhang, W. (2016). *Modularity Based Community Detection with Deep Learning*.