

Aplicação de Redes Neurais em Grafos para Aproximação da Métrica de Betweenness Centrality em Redes Sociais

Ana Luiza Almeida Soares¹, Rodrigo César Pedrosa Silva¹

¹Programa de Pós-Graduação em Ciência da Computação

Departamento de Computação

Universidade Federal de Ouro Preto

ana.almeida3@aluno.ufop.edu.br, rodrigo.silva@ufop.edu.br

1. Introdução

As redes sociais têm se mostrado ferramentas poderosas para resolver uma ampla gama de problemas em diferentes áreas. Essas redes não se restringem apenas a plataformas digitais como Facebook, Twitter e Instagram, mas também são aplicáveis em contextos que envolvem interações entre elementos, como a disseminação de informações, a propagação de doenças, e a maximização de influência em campanhas de marketing. Sendo assim, em contextos como o espalhamento de notícias ou epidemias, a compreensão da estrutura da rede é crucial para prever e controlar o fluxo de informações ou patógenos.

Nessas aplicações, o peso das arestas desempenha um papel fundamental, uma vez que muitos algoritmos dependem dessa informação para operar de maneira eficiente. Por exemplo, o algoritmo de *Linear Threshold* utiliza os pesos das arestas para determinar a probabilidade de um nó influenciar seus vizinhos, o que impacta diretamente o resultado do processo de propagação de influência. No entanto, obter essas informações de peso nem sempre é uma tarefa trivial, especialmente em redes grandes e complexas, onde esses valores podem não ser conhecidos ou fáceis de mensurar.

Uma abordagem comum para atribuir pesos às arestas em redes sociais é o uso da métrica de *Betweenness Centrality*. Esta métrica mede a frequência com que uma aresta aparece nos caminhos mínimos entre todos os pares de nós na rede, sendo, portanto, um indicador de sua importância estrutural. No entanto, calcular a *Betweenness Centrality* é um processo computacionalmente custoso, com complexidade $O(n^2m)$ em grafos não ponderados, onde n é o número de nós e m é o número de arestas. Isso torna o uso dessa métrica inviável para redes sociais de grande escala.

Diante desse desafio, o objetivo deste trabalho é utilizar uma *Graph Neural Network* (GNN) para prever a *Betweenness Centrality* das arestas em redes sociais, aproximando seu valor de forma mais eficiente. A proposta é explorar o potencial das GNNs para aprender padrões complexos de conectividade e fornecer estimativas precisas dessa métrica, reduzindo significativamente o tempo de processamento necessário em comparação aos métodos tradicionais.

2. Materiais e Métodos

A metodologia adotada para este trabalho envolve a escolha de três redes sociais, o cálculo da *Betweenness Centrality* dessas redes, a seleção aleatória de algumas arestas com seus respectivos valores de *Betweenness*, e a alimentação de uma *Graph Neural Network* (GNN) com esses dados. Posteriormente, a previsão da GNN será comparada com os

valores reais de *Betweenness*, e o erro obtido será utilizado para avaliar a qualidade da previsão.

As redes sociais escolhidas para a experimentação foram obtidas a partir dos sites disponibilizados pelo Professor Dr. Vander Freitas na disciplina *PCC 121 Redes Complexas*. Essas redes representam páginas do Facebook (verificadas, ou seja, páginas azuis) de diferentes categorias, e os nós representam as páginas enquanto as arestas representam as "curtidas mútuas" entre elas. Abaixo estão as redes selecionadas para o estudo:

1. **Site:** <https://networkrepository.com/fb-pages-politician.php>
Nós: 5.900
Arestas: 41.700
2. **Site:** <https://networkrepository.com/fb-pages-food.php>
Nós: 620
Arestas: 2.100
3. **Site:** <https://networkrepository.com/fb-pages-artist.php>
Nós: 50.500
Arestas: 819.100

3. Resultados Esperados

Espera-se que o erro entre a previsão da *Graph Neural Network* (GNN) e os valores reais de *Betweenness Centrality* não seja significativamente alto, indicando que a rede neural é capaz de aprender e generalizar bem as características da rede para estimar essa métrica de forma precisa. Além disso, espera-se que, ao utilizar a GNN para prever os valores de *Betweenness*, haja uma melhoria considerável no tempo de processamento em comparação com o cálculo tradicional dessa métrica, que é computacionalmente custoso.