

Projeto 1 - Análise Estrutural de Redes Sociais utilizando Métricas de Grafos

Leandro Beloti Kornelius - 211020900

¹Dep. Ciência da Computação – Universidade de Brasília (UnB)
Teoria e Aplicação de Grafos, Turma 01, 2025/2

1. Introdução

O estudo de grafos constitui uma das bases fundamentais da Ciência da Computação, permitindo representar e compreender sistemas compostos por entidades interconectadas. Formalmente, um grafo é um conjunto de nós (ou vértices) conectados por arestas, que modelam relações ou interações entre os elementos. Essa estrutura abstrata é amplamente empregada em contextos reais, como redes de comunicação, sistemas biológicos, rotas de transporte e, especialmente, redes sociais, nas quais os vértices representam indivíduos e as arestas simbolizam vínculos de amizade ou interação. A análise desses grafos possibilita compreender não apenas a forma da rede, mas também os padrões de comportamento coletivo e de influência presentes nas conexões.

Para além da estrutura topológica, métricas específicas permitem quantificar aspectos importantes da rede e de seus vértices. Entre as mais relevantes estão as medidas de centralidade, que indicam a importância ou influência de cada nó sob diferentes perspectivas: o grau de centralidade revela o número de conexões diretas, a centralidade de intermediação (betweenness) identifica vértices que atuam como pontes na comunicação entre diferentes partes do grafo, a centralidade de proximidade (closeness) mede a eficiência com que um nó pode alcançar os demais, e a centralidade de autovetor (eigenvector) avalia a relevância de um nó considerando a importância dos seus vizinhos. Além dessas métricas, a detecção de comunidades, como a realizada pelo algoritmo de Louvain, permite identificar agrupamentos de nós densamente conectados internamente, revelando subestruturas e padrões de organização social.

Essas análises possuem aplicações diretas e de grande impacto em diversas áreas. No contexto das redes sociais online, compreender a estrutura do grafo e as métricas associadas possibilita identificar usuários influentes, prever a propagação de informações, detectar grupos coesos e até mesmo analisar comportamentos emergentes em comunidades virtuais. Tais ferramentas são essenciais para estratégias de marketing digital, monitoramento de desinformação, estudos sociológicos e otimização de sistemas de recomendação. Assim, o estudo e a aplicação das métricas de grafos reforçam a importância da modelagem matemática e computacional como instrumento para interpretar e intervir em fenômenos complexos do mundo real.

2. Implementação

Todas as seções descritas abaixo, podem ser verificadas com mais detalhes no repositório do github.

<https://github.com/LeandroKornelius/Projeto1-TAG>

Neste relatório foram descritas e mostradas apenas as etapas mais importantes para o funcionamento adequado do grafo e métricas solicitadas.

2.1. Etapa 1 - Coleta de Dados

O primeiro passo do projeto consistiu em preparar o conjunto de dados a ser utilizado para a análise da rede social. O dataset base, disponibilizado pelo repositório SNAP da Universidade de Stanford, contém a rede de amizades do Facebook com 4.039 nós e 88.234 arestas, representando usuários anonimizados e suas conexões. Essa etapa é essencial, pois define a base estrutural sobre a qual todas as métricas de análise de grafos serão calculadas. Trabalhar com redes reais, ainda que anonimizadas, garante que as conclusões obtidas sejam mais próximas de situações concretas de interação social.

Para atender à especificação do projeto, foram selecionados 2.000 nós aleatórios do grafo original, juntamente com todas as suas arestas correspondentes, de modo a criar um subgrafo induzido representativo. A amostragem aleatória foi realizada de forma reprodutível, utilizando uma semente fixa (`seed=42`), assegurando que o experimento possa ser replicado com os mesmos resultados. O processo envolveu a leitura do arquivo de arestas original, identificação dos vértices, sorteio dos nós e gravação de um novo arquivo (`subgraph.txt`) contendo apenas as conexões relativas aos vértices aleatoriamente selecionados.

2.2. Etapa 2 - Construção do Grafo

Após a obtenção do conjunto de dados reduzido, foi realizada a construção do grafo utilizando a biblioteca NetworkX, uma das mais consolidadas para análise e manipulação de grafos em Python. Essa etapa corresponde à criação da estrutura de dados que representa a rede, transformando o arquivo de texto (`subgraph.txt`) em um objeto manipulável dentro do ambiente do Jupyter Notebook. A representação do grafo em memória é essencial para o cálculo das métricas e a visualização posterior, pois permite aplicar diretamente os algoritmos de análise disponíveis na biblioteca.

Durante essa fase, também foram aplicadas verificações básicas sobre a estrutura resultante — como o número de nós, arestas e possíveis componentes desconexas — permitindo avaliar a integridade do subgrafo.

O código diferenciou entre o grafo completo (G) e a versão conectada ($G\text{-conn}$), utilizada nas métricas que exigem conectividade, como as centralidades de proximidade e de autovetor. Essa distinção é importante para garantir a validade matemática dos cálculos e evitar resultados inconsistentes.

2.3. Etapa 3 - Extração de Métricas Relevantes do Grafo

A terceira etapa do projeto focou na detecção de comunidades, um processo essencial para compreender a estrutura modular da rede. Comunidades são agrupamentos de nós com alta densidade de conexões internas e poucas conexões externas, refletindo subgrupos de usuários mais fortemente interligados. Para isso, foi utilizado o algoritmo de Louvain.

Na implementação, o particionamento foi obtido pela função `community-louvain.best-partition`, que associa cada nó a uma comunidade específica. Em seguida, os resultados foram incorporados como atributo no grafo, permitindo visualizar os agrupamentos em

cores distintas. O cálculo da modularidade global foi realizado para quantificar a qualidade da divisão encontrada. Essa análise revelou como a rede social se organiza internamente, indicando potenciais clusters de usuários com interesses ou interações mais intensas entre si.

Ademais, nesta etapa foram calculadas quatro das principais métricas de centralidade: grau, intermediação (betweenness), proximidade (closeness) e autovetor (eigenvector). Cada uma dessas medidas oferece uma perspectiva distinta sobre a importância de um nó dentro da rede. O grau de centralidade avalia a conectividade direta do nó; a intermediação mede sua capacidade de servir como intermediário em caminhos mínimos; a proximidade quantifica a eficiência do nó para alcançar outros vértices; e a centralidade de autovetor considera não apenas o número de conexões, mas também a importância dos vizinhos conectados. Juntas, essas métricas permitem uma visão multifacetada da influência e relevância estrutural dos usuários.

Na implementação, as funções nativas do NetworkX foram utilizadas para o cálculo de cada métrica, aplicadas principalmente sobre o grafo conectado (G-conn) para garantir a consistência dos resultados. Após o cálculo, os valores foram armazenados em um dataframe em que os valores foram armazenados como atributos de cada nó, o que facilita tanto a visualização, manipulação e exportação.

2.4. Etapa 4 - Visualização das Medidas Extraídas de Forma Explícita

A etapa final do projeto concentrou-se na visualização dos resultados obtidos, buscando tornar a análise mais intuitiva e interpretável. No ambiente Jupyter Notebook presente no repositório do projeto no GitHub, foram gerados diversos gráficos utilizando as bibliotecas Matplotlib e NetworkX, com o intuito de destacar tanto as comunidades detectadas quanto as métricas de centralidade analisadas.

Cada nó foi colorido conforme sua comunidade, e os vértices mais relevantes em cada métrica foram rotulados e ampliados, o que facilita a identificação visual das regiões mais influentes da rede.

A primeira figura do notebook apresenta o grafo amostrado de 2.000 nós, com as comunidades detectadas pelo algoritmo de Louvain indicadas por diferentes cores. Os vértices isolados, destacados em laranja, representam usuários sem conexões dentro da amostra selecionada, funcionando como pontos sem arestas incidentes.

Na segunda figura, são mostrados os cinco nós com maior centralidade de grau, destacados em vermelho sobre o mapa de comunidades. Essa métrica mede o número de conexões diretas de cada nó, permitindo identificar usuários com maior conectividade local, que atuam como centros de interação dentro de suas comunidades.

A terceira figura evidencia os cinco nós com maior centralidade de intermediação, métrica que identifica vértices que atuam como pontes entre diferentes comunidades. Os nós destacados em vermelho estão posicionados entre grupos de cores distintas, confirmando sua função de conectar regiões que, de outra forma, permaneceriam separadas. Esses usuários são cruciais para a propagação de informações entre comunidades e para a coesão global da rede.

Na quarta figura, são realçados os cinco nós com maior centralidade de proximidade, indicando os vértices que, em média, estão mais próximos de todos os outros na rede. Esses

nós possuem maior eficiência na disseminação de informações, alcançando rapidamente qualquer outro vértice.

Por fim, a quinta figura mostra os cinco nós com maior centralidade de autovetor, métrica que considera não apenas a quantidade, mas também a importância dos vizinhos conectados. Os nós destacados concentram-se em uma região densa e central do grafo, possivelmente pertencente à maior comunidade detectada, o que evidencia sua relevância estrutural dentro da rede.

3. Etapa 5 - Análises

Diante dos resultados, nota-se que o algoritmo de Louvain identificou múltiplas comunidades bem definidas, com modularidade elevada e pouca sobreposição entre si. Esse padrão é característico de redes sociais reais, nas quais grupos de amigos, colegas ou interesses comuns tendem a se formar de modo autônomo. Além disso, a presença de vértices isolados e pequenos subgrupos periféricos reforça a aleatoriedade da amostra.

As métricas de centralidade mostraram papéis distintos de influência na estrutura da rede. A centralidade de grau destacou usuários com alto número de conexões diretas, geralmente posicionados no interior das comunidades e atuando como líderes locais. Já a centralidade de intermediação evidenciou vértices que funcionam como pontes estruturais entre grupos diferentes, essenciais para o fluxo de informações entre comunidades. Por sua vez, a centralidade de proximidade indicou nós estrategicamente posicionados, capazes de alcançar rapidamente os demais vértices, desempenhando papel relevante na eficiência da comunicação interna da rede.

Por fim, a centralidade de autovetor revelou usuários cuja importância deriva não apenas de sua quantidade de conexões, mas da influência dos nós com os quais se relacionam. Esses vértices se concentram nas regiões mais densas e coesas do grafo, indicando núcleos de prestígio dentro da estrutura social analisada.

Quando interpretadas em conjunto, essas métricas mostram que o poder de influência não se limita à quantidade de conexões, mas também depende da posição topológica e do contexto comunitário de cada indivíduo.

4. Conclusão

O estudo demonstrou como a aplicação de conceitos e métricas de Teoria dos Grafos permite compreender, de forma quantitativa e visual, a dinâmica de uma rede social real. A identificação de comunidades por Louvain, combinada com as medidas de centralidade, evidenciou diferentes camadas de influência, desde líderes locais até intermediários globais, revelando a complexidade da estrutura social subjacente. A visualização dos resultados reforça a noção de que a organização das redes humanas reflete conceitos estudados na disciplina.

Em síntese, o projeto cumpriu seu objetivo ao integrar modelagem, análise e interpretação de dados em um contexto realista de rede social. Os resultados obtidos demonstram a relevância prática da Teoria dos Grafos para a compreensão de fenômenos de influência, disseminação de informação e coesão social.