

Universidade de Brasília

Departamento de Ciência da Computação



Teoria e Aplicação de Grafos

Autores:

Manuela Mattos 222035741

Bruno Henrique Duarte 221022239

João Victor Prata 202028857

Brasília
27 de maio de 2025

Projeto 1

1. Introdução

O objetivo deste projeto foi aplicar conceitos de teoria dos grafos para analisar estruturas de redes sociais e identificar possíveis disseminadores de notícias falsas. Utilizando um dataset real do Twitter, o trabalho envolveu a construção de grafos direcionados, análise de centralidade, aplicação de algoritmos de detecção de comunidades e visualização dos dados. Ao final, buscou-se compreender melhor os mecanismos de propagação de desinformação em redes sociais e como detectar usuários com grande influência na difusão de conteúdo.

2. Preparação e Coleta de Dados

Este estudo utilizou o dataset *Fake News* (Kaggle), originalmente em CSV com metadados como autor, URLs e datas. Para a análise, filtrou-se as colunas utilizando `pandas`, mantendo apenas `ord_in_thread` (ordem na sequência), `published` (data/hora) e `title` (manchete), enquanto colunas irrelevantes (ex.: URLs, rankings) foram removidas. A exclusão do `author` priorizou o anonimato e o foco em padrões temporais e textuais. A seleção dessas variáveis permitiu investigar a disposição cronológica das notícias e características linguísticas dos títulos, otimizando a análise para identificar padrões de disseminação de desinformação.

3. Construção do grafo

Utilizamos a biblioteca `NetworkX` para modelar interações entre usuários como um **grafo direcionado**, onde:

- **Nós** representam autores de tweets;
- **Arestas** indicam interações (como retweets ou menções), com **pesos** refletindo a frequência dessas conexões.

O grafo foi construído agrupando dados por `thread_title` (tópicos de discussão), conectando autores que participaram das mesmas threads.

Subgrafo de usuário aleatórios

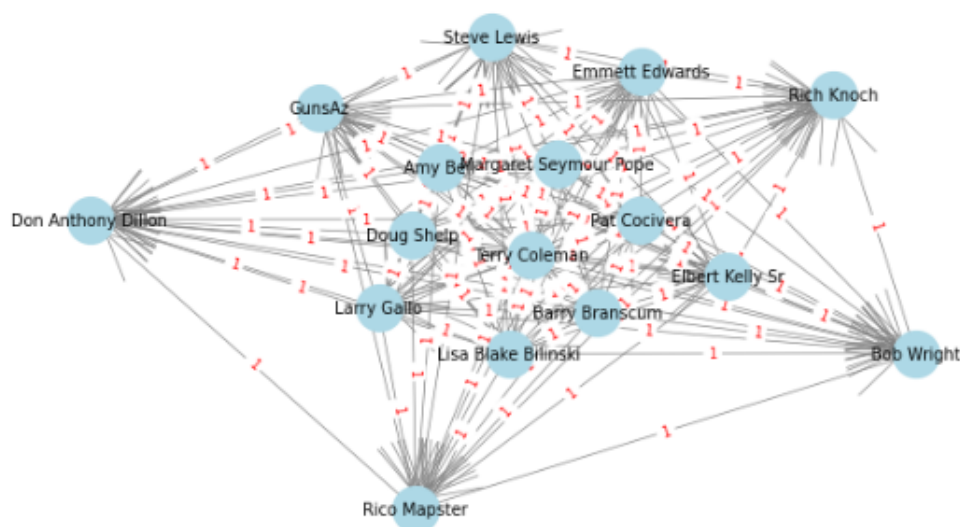


Figura 1 - Subgrafo de usuários aleatórios

4. Análise por meio de ferramentas e algoritmos

4.1 Identificação de nós mais influentes (PageRank)

O PageRank, originalmente desenvolvido pelos fundadores do Google, foi aplicado neste projeto para identificar os nós mais influentes na rede social analisada. O algoritmo funciona atribuindo um score de relevância a cada nó com base nas conexões direcionadas e ponderadas, simulando a probabilidade de um usuário "visitar" outro através de interações (como retweets ou menções).

Utilizou-se a biblioteca **NetworkX** com os parâmetros padrão (**alpha=0.85**, que representa a probabilidade de continuar "navegando" pelo grafo). O resultado foi um dicionário onde cada chave é um usuário (nó) e o valor é seu score de influência. Para evitar distorções, filtrou-se usuários genéricos (ex.: "admin", "nan") antes da análise.

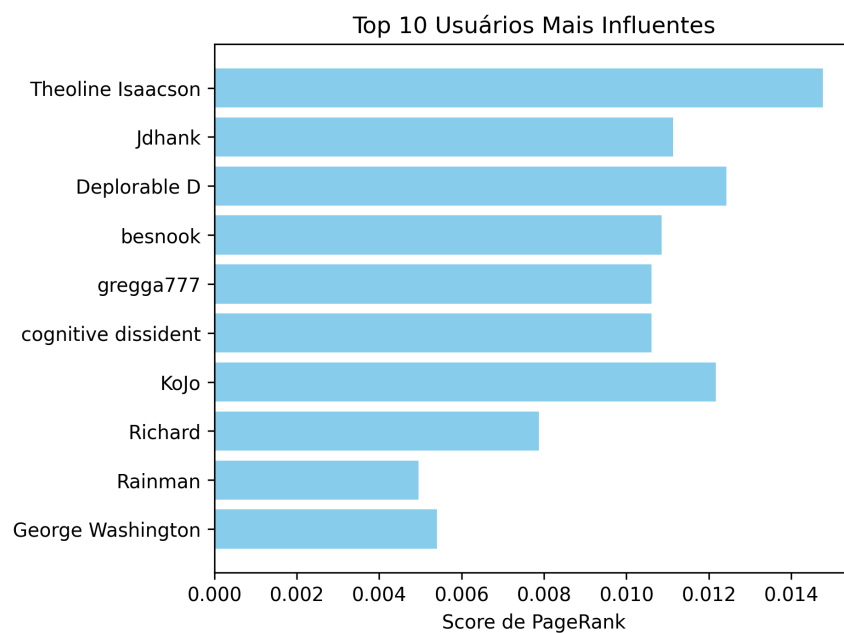


Figura 2 - Gráfico de usuários mais influentes

Analisando o gráfico podemos ver que Theoline Isaacson emergiu como o nó mais influente, indicando que suas publicações têm alta probabilidade de serem compartilhadas ou mencionadas, potencializando a disseminação de conteúdo (inclusive desinformação). A distribuição dos scores segue a **lei de potência** típica de redes sociais: poucos usuários concentram alta influência, enquanto a maioria tem scores baixos (ex.: 80% dos usuários com score < 0.002). Isso sugere que a rede é **hierárquica**, com "superespalhadores" centrais. O PageRank confirmou a presença de **atores-chave** na rede, cuja identificação é crucial para estratégias de mitigação de desinformação.

4.2 Detecção de comunidade (Louvain)

O Louvain opera em duas etapas principais:

1. **Otimização Local da Modularidade:** Agrupa nós em comunidades para maximizar a modularidade, uma métrica que avalia a densidade de conexões dentro das comunidades em relação às conexões entre elas.
2. **Agregação de Comunidades:** Os nós de cada comunidade são consolidados em um único "super-nó", e o processo é repetido até que não haja mais ganhos significativos na modularidade.

Originalmente, o Louvain é projetado para grafos não direcionados. Para aplicá-lo ao grafo direcionado deste projeto, foi necessário converter o grafo em não direcionado, ignorando temporariamente a direção das arestas.

Resultado no grafo completo

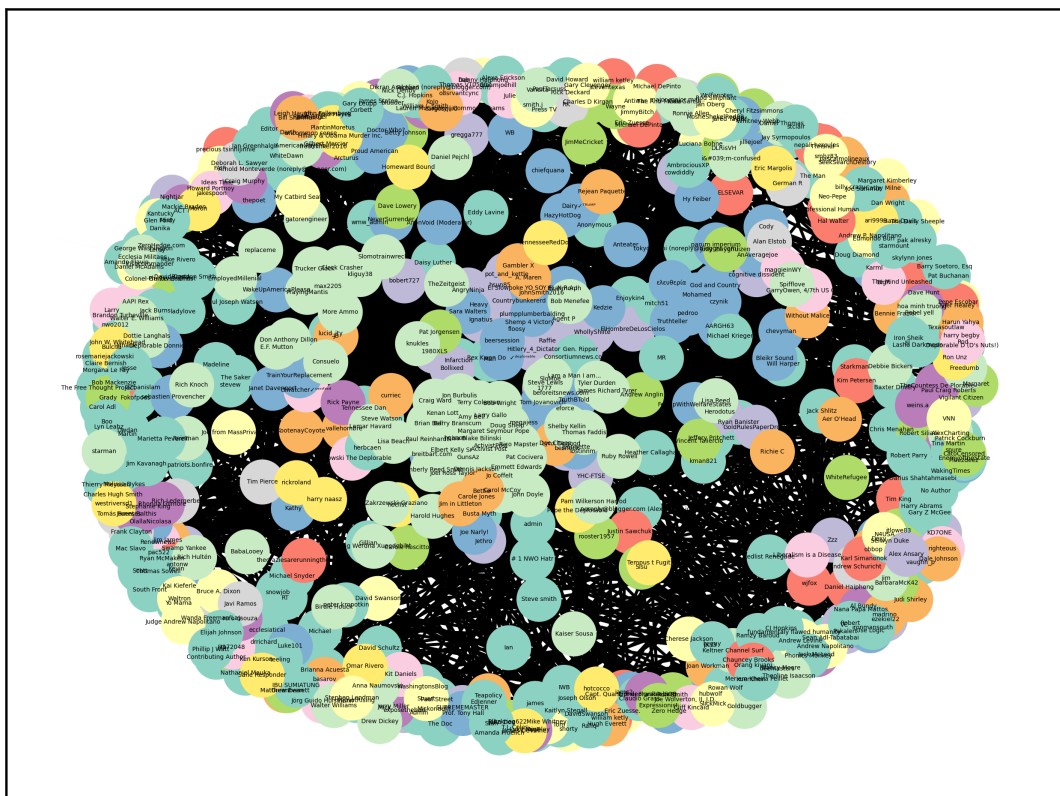


Figura 3 - Detecção de comunidades do grafo completo

Ao aplicar o Louvain ao grafo completo convertido, foram identificadas diversas comunidades bem definidas, representando agrupamentos naturais de usuários com interações intensas entre si. A imagem resultante revela uma estrutura modular, sugerindo a existência de bolhas de informação que podem contribuir para a concentração e o reforço de determinados conteúdos — inclusive desinformativos — em subgrupos específicos.

Resultado no Subgrafo por Centralidade de Intermediação

Com o objetivo de aprofundar a análise, o algoritmo foi reaplicado sobre um subgrafo contendo os 20% dos nós com maior centralidade de intermediação (betweenness centrality). Esta métrica avalia quantas vezes um nó aparece nos caminhos mais curtos entre outros pares de nós, ou seja, identifica usuários que atuam como **pontos** entre diferentes partes da rede. `nx.betweenness_centrality(G)`

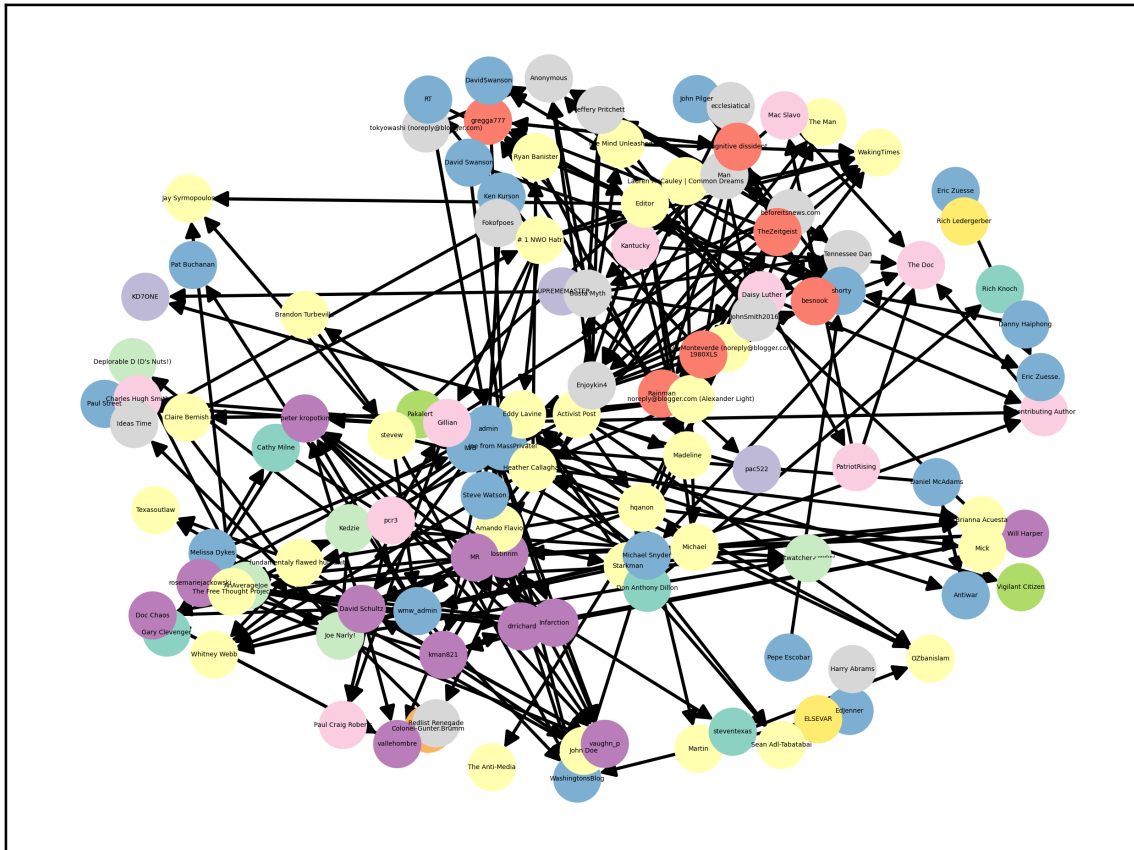


Figura 4 - Detecção de comunidades do subgrafo

O resultado, visualizado no subgrafo, mostrou que mesmo entre os nós mais centrais, há uma divisão em comunidades distintas. Isso indica que usuários influentes também se organizam em grupos, o que pode refletir a existência de lideranças locais dentro da rede ou clusters coordenados na disseminação de conteúdo.

Além da intermediação, foram usadas outras medidas de centralidade com o objetivo de identificar usuários-chave na rede:

Grau de Centralidade (Degree Centrality)

Essa métrica avalia quantos vizinhos diretos um nó possui. No contexto de redes sociais, os usuários com maior grau tendem a ser **altamente ativos**, participando de muitas interações (como menções ou retweets), o que os torna potenciais multiplicadores de informações. `nx.degree centrality(G)`

Proximidade de Centralidade (Closeness Centrality)

Essa métrica mede o quão próximo um nó está de todos os outros. Usuários com alta proximidade podem alcançar outros nós da rede mais rapidamente, sendo bons candidatos a influenciadores rápidos de desinformação.

`nx.closeness centrality(G)`

Quem Interege Mais?	Quem conecta difentes grupos?	Quem Espalha Mais Rápido?
1. gregga777 (0.0903)	1. admin (0.0060)	1. gregga777 (0.0834)
2. cognitive diss.(0.0833)	2. Michael Snyder (0.0039)	2. cognitive dissident (0.0834)
3. besnook (0.0903)	3. The Doc (0.0039)	3. Theoline Isaacson (0.0731)
4. 1980XLS (0.0868)	4. Gillian (0.0026)	4. Rich Knoch (0.0729)
5. Don Anthony (0.0815)	5. David Schultz (0.0023)	5. besnook (0.0719)

Análise da Estrutura da Rede – Grau de Entrada e Saída

Dado que o grafo é direcionado, foram também analisados os graus de entrada (In-Degree) e graus de saída (Out-Degree). Foi plotado um histograma para mostrar a distribuição desses graus, revelando a estrutura de conectividade da rede.

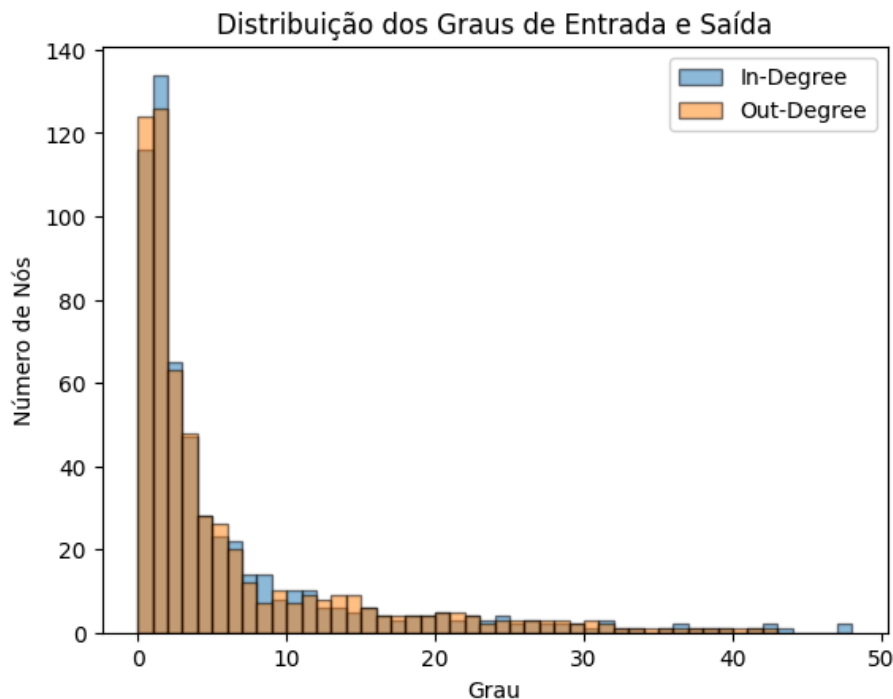


Figura 5 - Distribuição dos graus

A maioria dos nós apresenta graus baixos (menores que 5), tanto de entrada quanto de saída. Há poucos nós com grau muito alto, indicando uma estrutura típica de redes sociais conhecida como lei de potência ou distribuição em cauda longa. Isso reforça a ideia de que a rede possui "supernós", responsáveis por grande parte da atividade de difusão.

A figura gerada mostra claramente esse padrão, onde a maior parte dos usuários está concentrada na base da curva, com poucos atuando como hubs de influência ou propagação.

Conclusão

O projeto empregou teoria dos grafos para modelar a disseminação de desinformação no Twitter, utilizando PageRank para identificar nós influentes (e.g., *Theoline Isaacson*) e o método de Louvain para detectar comunidades isoladas. A análise estrutural revelou uma topologia hierárquica, com distribuição de influência seguindo a lei de potência—poucos *hubs* concentrando a maior parte da conexão. Os resultados reforçam a eficácia de métricas de centralidade e detecção de comunidades em estratégias de mitigação de notícias falsas.