



Departamento de Ciência da Computação (DCC)
Discente: Gustavo Dantas Galote de Lima Padilha
Matrícula: 202220661
Professor: Vinícius Dias
Disciplina: GCC218 - Algoritmos em Grafos

Introdução

Este projeto envolve a análise do estudo do gráfico da web de Stanford, uma representação complexa das páginas da Universidade de Stanford e suas interconexões via hiperlinks. Utilizando o algoritmo HITS (Hyperlink-Induced Topic Search), nosso objetivo é identificar as páginas mais influentes nesta rede, classificando-as como 'hubs', importantes fontes de links, e 'autoridades', páginas frequentemente referenciadas por outras. Este método, amplamente empregado em motores de busca e análises de redes sociais, oferece insights valiosos sobre como a informação e a influência se propagam na web.

Apresentação da Base de Dados

Este projeto envolve a análise do gráfico da [web de Stanford](#), que representa páginas da Universidade de Stanford e suas conexões de hiperlink. Este grafo é direcionado, com os vértices representando as páginas da web e as arestas representando os hiperlinks entre elas.

Limitação do Escopo e Definição do Problema

O escopo deste projeto é estritamente delimitado ao estudo da influência e da interconectividade dentro do gráfico da web de Stanford, utilizando especificamente o algoritmo HITS para analisar e identificar hubs e autoridades. Esta abordagem foca em duas questões principais: quais páginas da web de Stanford são centrais em termos de fornecer links para outras páginas (hubs) e quais são amplamente referenciadas por outras páginas (autoridades). O objetivo é mapear a estrutura de influência dentro da rede, proporcionando uma visão detalhada de como a informação e a autoridade são distribuídas.

Entrada Esperada

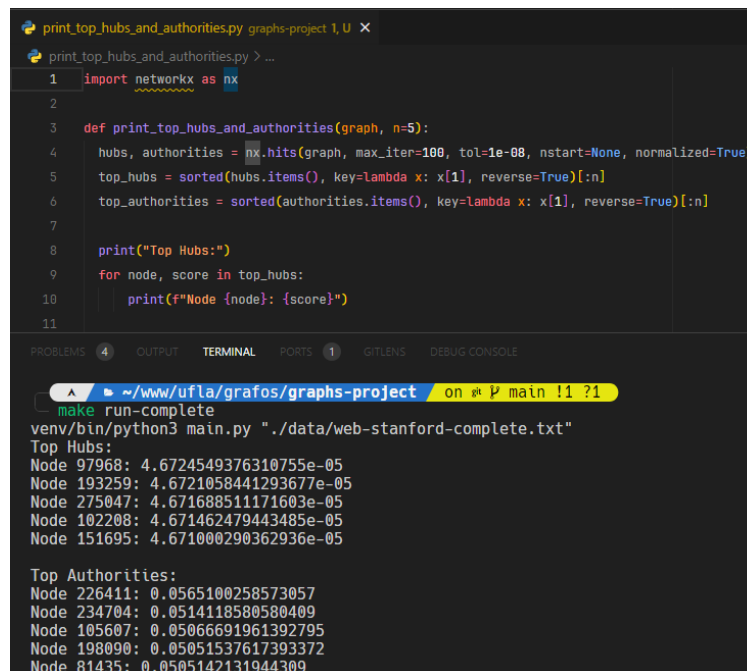
- Grafo: O gráfico da web de Stanford, representado em um formato apropriado para processamento, como um arquivo de texto contendo pares de vértices, onde cada par representa uma ligação de hiperlink entre duas páginas.
- Parâmetros do Algoritmo: Configurações para o algoritmo HITS, como número de iterações e tolerância para a convergência.

Saída Esperada

- Listas de Hubs e Autoridades: Duas listas classificadas, uma contendo os principais hubs e a outra as principais autoridades, com seus respectivos scores calculados pelo algoritmo HITS.
- Visualizações: Representações gráficas que ilustram os hubs e autoridades mais significativos, possivelmente destacando-os em um subconjunto do gráfico original.

Exemplos de saídas

Exemplo completo (base de dados completa que, pela quantidade, foi utilizado o terminal)

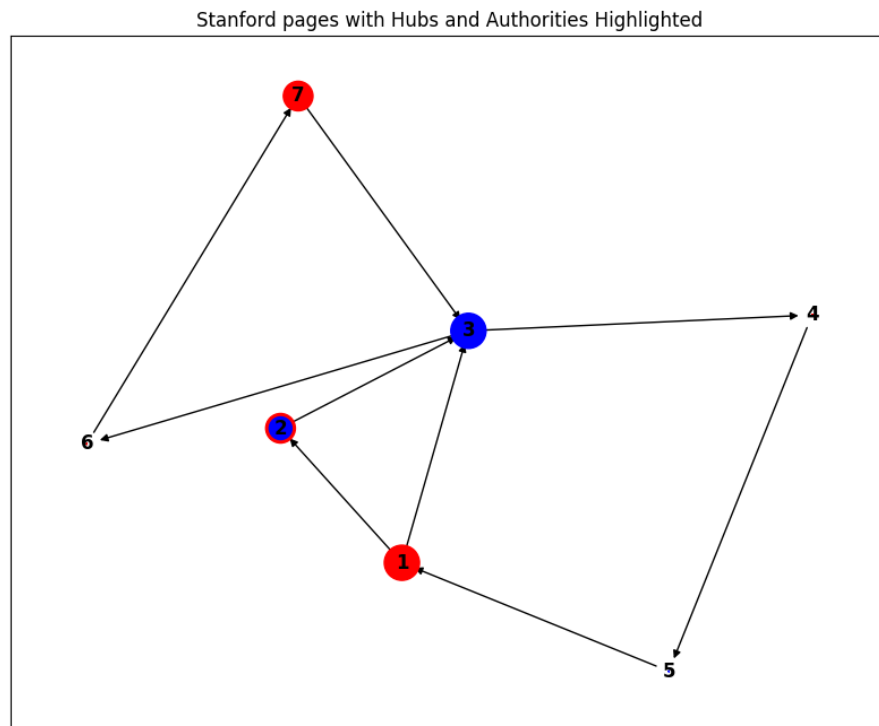


```

print_top_hubs_and_authorities.py graphs-project 1, U X
print_top_hubs_and_authorities.py > ...
1 import networkx as nx
2
3 def print_top_hubs_and_authorities(graph, n=5):
4     hubs, authorities = nx.hits(graph, max_iter=100, tol=1e-08, nstart=None, normalized=True)
5     top_hubs = sorted(hubs.items(), key=lambda x: x[1], reverse=True)[:n]
6     top_authorities = sorted(authorities.items(), key=lambda x: x[1], reverse=True)[:n]
7
8     print("Top Hubs:")
9     for node, score in top_hubs:
10         print(f"Node {node}: {score}")
11
PROBLEMS 4 OUTPUT TERMINAL PORTS 1 GITLENS DEBUG CONSOLE
~/www/ufla/grafos/graphs-project on * P main !1 ?1
make run-complete
venv/bin/python3 main.py "./data/web-stanford-complete.txt"
Top Hubs:
Node 97968: 4.6724549376310755e-05
Node 193259: 4.6721058441293677e-05
Node 275047: 4.671688511171603e-05
Node 102208: 4.671462479443485e-05
Node 151695: 4.671000290362936e-05
Top Authorities:
Node 226411: 0.0565100258573057
Node 234704: 0.0514118580580409
Node 105607: 0.05066691961392795
Node 198090: 0.05051537617393372
Node 81435: 0.0505142131944309

```

Exemplo com dados de amostra (sendo possível ser plotado)



Resultados

Após a execução do algoritmo HITS no gráfico da web de Stanford, os resultados revelam informações fascinantes sobre a estrutura de influência e conectividade da rede. Os 'Top Hubs' identificados, como os nós 97968, 193259, 275047, 102208 e 151695. Estes valores indicam que estas páginas são pontos cruciais de conexão na rede, atuando como grandes referenciadores para outras páginas.

Em contrapartida, os 'Top Authorities' destacados, incluindo os nós 226411, 234704, 105607, 198090 e 81435, com scores variando de 0.056 a 0.050, emergem como as páginas mais referenciadas, estabelecendo-as como fontes de informação altamente influentes e confiáveis dentro da rede da universidade.

Considerações finais

O projeto enfrentou desafios, especialmente no que diz respeito ao manejo de grandes conjuntos de dados e à interpretação de resultados complexos, mas os insights adquiridos são valiosos. Eles fornecem uma possível base para futuras investigações em análises de redes e oferecem uma perspectiva para acadêmicos e profissionais interessados na dinâmica das redes da web. Por fim, ressalto a importância da teoria dos grafos na compreensão das estruturas digitais e o potencial dos algoritmos de análise de rede em revelar as camadas ocultas de conexões e influências no ambiente digital.