



Universidade Federal do Ceará – UFC
Centro de Ciências – CC
Mestrado e Doutorado em Ciências da Computação - MDCC
Estruturas de Dados

Exercício: Algoritmos de Predição de Ligações

Objetivos: Exercitar os conceitos predição de ligações.

Data da Entrega: 13/10/2025

OBS 1: Exercício Individual.

OBS 2: A entrega da lista deverá ser executada utilizando-se o SIGAA.

NOME: _____ MATRÍCULA: _____

Questão 1

Crie um arquivo Jupyter Notebook e realize as seguintes operações:

- a) Ler o dataset fakeTelegram.BR_2022.csv, o qual está disponível no link a seguir:
https://drive.google.com/file/d/1c_hLzk85pYw-huHSnFYZM_gn-dUsYRDm/view?usp=sharing
- b) Remova os trava-zaps.
- c) Remover textos com menos de 5 palavras.
- d) Monte os grafos solicitados a seguir.

Utilizando as ferramentas:

1. NetworkX (<https://networkx.org/>)
2. Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>)
3. Gephi (<http://gephi.org>)

Modelar as relações entre os usuários na forma de um grafo pode fornecer informações relevantes sobre padrões de comportamento. Porém, diferente de redes sociais como Twitter ou Facebook, onde existem conexões bem definidas entre os usuários pela relação de seguir (Twitter) ou de amizade (Facebook), no Telegram essas conexões não são explícitas.

Assim, propomos uma modelagem das relações entre usuários do WhatsApp na forma de grafos direcionados e valorados, considerando o envio de mensagens em grupos. Nessa modelagem, **cada nó representa um usuário** e podemos considerar **um grafo para cada tipo de mensagem: mensagem em geral, mensagem viral e mensagem com desinformação**.

Considerando o grafo de mensagens gerais, onde cada nó representa um usuário, existe uma aresta direcionada entre o usuário i e o usuário j se o usuário i enviou uma mesma

mensagem que o usuário j. O peso dessa aresta é a quantidade de mensagens iguais (ou semelhantes) enviadas tanto pelo usuário i quanto pelo usuário j.

Um raciocínio análogo pode ser aplicado para criar um grafo apenas de mensagens virais: existe uma aresta direcionada entre o usuário i e o usuário j se tanto o usuário i quanto o usuário j postaram uma mesma mensagem viral e o peso dessa aresta é quantidade de mensagens virais enviadas tanto pelo usuário i quanto pelo usuário j.

O mesmo procedimento pode ser utilizado para criar o grafo de desinformação: existe uma aresta direcionada entre o usuário i e o usuário j se tanto o usuário i quanto o usuário j postaram uma mesma mensagem contendo desinformação e o peso dessa aresta é quantidade de mensagens com desinformação enviadas tanto pelo usuário i quanto pelo usuário j.

Percebe-se que nos três grafos, a quantidade de nós é a mesma, variando a quantidade de arestas.

Monte uma tabela contendo a quantidade de nós e a quantidade de arestas para cada grafo (mensagens gerais, mensagens virais e mensagens com desinformação).

Verifique se existem grupos isolados e clusters de usuários fortemente conectados. Isso pode ocorrer caso existam usuários engajados que possuem participação ativa em diversos grupos.

Analise a rede e os nós utilizando as medidas discutidas em sala de aula (Giant coeficiente, Density, Diameter, Degree centrality, Clustering coefficient, Closeness centrality, Betweenness centrality).

Identifique os 5 usuários mais ativos.

Identifique os 5 usuários que mais espalham desinformação.

Identifique os 5 usuários mais influentes.

Identifique os 5 usuários mais conectados.

Monte uma tabela contendo a quantidade de nós e a quantidade de arestas para cada grafo (mensagens gerais, mensagens virais e mensagens com desinformação).

Implemente algoritmos de predição de ligações utilizando as métricas a seguir:

- Preferential attachment
- Common neighbors
- Jaccard
- Adamic-Adar
- Graph distance
- Katz

Avalie cada um dos algoritmos implementados (em cada um dos três grafos construídos) utilizando como estratégia a remoção aleatória de 10% das arestas.

A avaliação deste trabalho se dará em duas etapas:

1ª. Vídeo de Apresentação: Cada estudante irá disponibilizar um vídeo (no Youtube) apresentando o todo o código gerado, bem como as ferramentas utilizadas. O estudante pode utilizar slides e notebooks na produção do vídeo.

2ª. Avaliação do Código: O professor da disciplina irá avaliar a qualidade dos códigos (notebooks) gerados pelo estudante, bem como a utilização das ferramentas utilizadas e as análises realizadas.

A avaliação do trabalho irá envolver os seguintes quesitos:

- Qualidade e organização do código (Notebook);
- Clareza da descrição das atividades realizadas e dos resultados obtidos;
- Domínio do Tema;

PS. Não serão aceitos trabalhos que não forem apresentados (por meio de vídeo disponibilizado no Youtube).

PS. Cada estudante será responsável pela disponibilização do ambiente (software e hardware) necessário para a gravação da apresentação do seu trabalho.

Os Notebooks, Arquivos CSVs, Dumps e URL do vídeo deverão ser disponibilizados, em formato .ZIP, no SIGAA. Caso o tamanho do arquivo .ZIP ultrapasse o limite máximo permitido pelo SIGAA, o estudante pode disponibilizar um link para um repositório no Google Drive.

“Se não posso estimular sonhos impossíveis, não devo negar o direito de sonhar com quem sonha.”.

Paulo Freire