



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

Exercício: Algoritmos de Detecção de Comunidades

Objetivos: Exercitar os conceitos de detecção de comunidades.

Data da Entrega: 20/01/2025

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) Realizar o “restore” do arquivo (dump) denominado fd_whatsapp_0911_2023.zip no PostgreSQL. Esse arquivo está disponível no link a seguir:
<https://drive.google.com/drive/folders/1kEEnmZUVJEgYTynZjU6qMICbEVd8wKca?usp=sharing>
- b) Remova os trava-zaps.
- c) Remover textos com menos de 5 palavras.
- d) Monte um grafo no para modelar (representar) as relações entre postagens, usuários e grupos do WhatsApp. Para isso, utilize o NetworkX (<https://networkx.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 WhatsApp 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 mensagem para um grupo do qual o usuário j faz parte. O peso dessa aresta é a quantidade de mensagens enviadas pelo usuário i para os grupos nos quais o usuário j está presente.

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 o usuário i enviou uma

mensagem viral para um grupo do qual o usuário j faz parte e o peso dessa aresta é quantidade de mensagens virais enviadas pelo usuário i para aquele grupo.

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 o usuário i enviou uma mensagem contendo desinformação para um grupo do qual o usuário j faz parte e o peso dessa aresta é quantidade de mensagens virais enviadas pelo usuário i para aquele grupo.

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).

Utilize (se necessário, implemente) os seguintes algoritmos de detecção de comunidades:

- Girvan-Newman
- Edge Betweenness Partition
- Edge Current Flow Betweenness Partition
- K-clique Communities
- Greedy Modularity Communities
- Naïve Greedy Modularity Communities
- Lukes Partitioning
- Label Propagation Communities
- Fast Label Propagation Communities
- Louvain Communities
- Asynchronous Fluid Communities

Para cada um dos três cenários investigados, avalie os algoritmos utilizados por meio das seguintes métricas de qualidade:

- Modularity
- Partition Quality

Para cada um dos três cenários investigados, selecione os três melhores resultados. Em seguida, para cada um dos resultados obtidos (conjunto de comunidades), verifique o percentual de elementos de uma mesma comunidade que pertencem a um mesmo grupo do WhatsApp. Busque identificar se existe uma relação entre as comunidades encontradas e os grupos do WhatsApp.

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