



Universidade Federal de Alagoas
Instituto de Computação
Ciência de Dados



Professor: Bruno Pimentel

Aluno: _____

Lista de Exercícios 2

1. Aquisição de Dados e Leitura

Escreva um script em Python que leia o arquivo da rede de e-mails, disponível no seguinte link: <https://snap.stanford.edu/data/email-Eu-core-temporal.txt.gz>. Este conjunto de dados foi gerado a partir de trocas de e-mails em uma instituição de pesquisa europeia. Para mais informações sobre o conjunto de dados, acesse: <https://snap.stanford.edu/data/email-Eu-core-temporal.html>. Sugestão: para as questões a seguir, use o pacote **NetworkX**: <https://networkx.org/documentation/stable/index.html> .

2. Visualização da Rede

Desconsiderando o fator temporal, visualize a estrutura da rede social direcionada por meio de um gráfico que ilustre sua organização.

3. Análise Global

Utilizando ferramentas de análise de redes sociais, calcule a média dos menores caminhos (*shortest paths*) entre os nós da rede. Com base nesse valor, o que podemos concluir sobre a conectividade e a eficiência da comunicação na rede?

4. Análise Estrutural

Calcule os 5 nós com os maiores valores de Centralidade de Intermediação (*Betweenness Centrality*). Com base nesses nós, interprete os resultados alinhando o conceito dessa métrica ao contexto do conjunto de dados, que representa o envio de e-mails entre diferentes entidades dentro de uma instituição de pesquisa. O que esses nós indicam sobre a estrutura de comunicação na instituição?

5. Análise de Comunidades

Utilizando o algoritmo de detecção de comunidades de Louvain, identifique o nó com o maior grau de centralidade de chegada em cada comunidade. O que podemos afirmar sobre esses nós? Sugestão: use a função `louvain_partitions` do pacote `NetworkX`.

6. Visualização Temporal

Com base nos nós selecionados na questão anterior e considerando o fator temporal presente no conjunto de dados original, apresente graficamente o número de arestas de entrada para cada um desses nós ao longo dos 803 dias.

7. Análise de Séries Temporais

Escolha dois nós (A e B) da questão 5 e decomponha a série temporal correspondente em tendência, sazonalidade e ruído. Compare as tendências e sazonalidades entre os dois nós. Quais conclusões podemos tirar dessa comparação?

8. Mudança de Comportamento

Suponha que a instituição deseja alterar o fluxo de e-mails para reduzir a demanda em A e B. Para isso, crie um novo nó na rede, denominado C. Aleatoriamente, selecione 25% das arestas que vão de i para A (ou seja, arestas do tipo (i, A)) e substitua A por C, resultando em arestas do tipo (i, C). Repita o mesmo processo para as arestas que vão de i para B, substituindo 25% dessas arestas por (i, C).

9. Análise da Mudança

Repita o mesmo procedimento da questão 7 para os nós A, B e C.

10. Tomada de Decisão baseada em Rede Social

- Com base nas observações feitas na questão anterior, é possível afirmar se o fluxo de chegadas de e-mails em A e B sofreu alterações?
- Utilize as métricas de redes sociais discutidas anteriormente e aplique testes de hipóteses para fundamentar matematicamente a sua conclusão.