

Trabalho Prático

Redes Complexas

O objetivo desse trabalho é se familiarizar com ferramentas e APIs que serão necessárias, além de exercitar os principais conceitos dados em aula. O trabalho é **individual**. Você deve entregar uma documentação que contenha código, explicações e testes através do moodle. O trabalho vale **75 pontos**. Você deve submeter um PDF contendo cada etapa requisitada abaixo e também um arquivo compactado contendo seus códigos.

1) Crawler de dados:

- a. Construa um coletor de um grafo de uma rede social ou de algum sistema Web que possua uma rede complexa em seus dados. Uma lista de redes sociais se encontra abaixo.
http://en.wikipedia.org/wiki/List_of_social_networking_websites . Você deve escolher uma rede social com alguma característica peculiar. Você pode coletar redes como o Twitter, mas, nesse caso, você deve ter uma proposta clara em relação ao grafo que pretende construir. Procure uma API do sistema escolhido se existir e forneça uma breve descrição das chamadas que você utilizou na coleta. Você pode ainda montar uma rede que seja necessariamente uma rede social, como uma rede que relaciona chamadas de software, uma rede jogadores de futebol, etc. Nesse caso, procure obter bases de dados grandes e mostre a fonte onde obteve os dados.
- b. Identifique um grafo interessante a ser estudado nos dados coletados. Caso o grafo seja uma amostra, discuta suas limitações e procure fornecer alguma estimativa da fração do grafo coletado. O grafo não necessariamente precisa ser um grafo de amizades. Isso vai depender do sistema coletado e do tipo de análise proposta.

2) Análise de uma rede complexa: Para esta questão, você deve realizar uma análise de um grafo obtido no passo anterior do trabalho. Sugerimos o uso da biblioteca networkX em python. Em todas as questões abaixo, forneça interpretações dos resultados buscando comparar os resultados com a estrutura típica de outras redes antes estudadas.

- a) Forneça uma breve explicação da rede analisada.
- b) Calcule e plote a distribuição do grau dos nodos desse grafo. Discuta se a curva parece uma Power Law. Faça uma regressão linear na tentativa de obter o coeficiente alfa. Calcule também o grau médio do grafo.
- c) Calcule o número de componentes do grafo.
- d) Calcule o coeficiente de clusterização de cada nodo e plote a distribuição. Calcule também o coeficiente de clusterização global do grafo.
- e) Plote a distribuição do tamanho dos componentes do grafo. Se o grafo possuir apenas um componente não é preciso plotar.

- f) Calcule o overlap da vizinhança para cada par de nodo e plote a distribuição dos valores encontrados.
- g) Calcule a distância média e a distribuição das distâncias de todos os nodos da rede.
- h) Calcule o betweenness dos nodos e das arestas desse grafo e plote a distribuição. O que acontece com o grafo à medida que você retira os nodos e arestas com maior betweenness?
- i) Implemente uma abordagem para encontrar arestas com grandes chances de serem pontes. Implemente o conceito de *bridge span* e escolha um valor de *bridge span* interessante para o seu grafo.
- j) Calcule a assortatividade do grafo. Plote o gráfico da assortatividade além de apresentar o coeficiente de Pearson.
- k) Plote uma visualização do grafo ou de parte, ex. um componente, do grafo caso ele seja muito grande. Para essa tarefa você pode utilizar qualquer ferramenta. Uma sugestão é usar o gephi (<https://gephi.org/>).

3) Considere as redes com sinais abaixo. Note que essas redes são direcionadas e diferentes das redes discutidas no capítulo 5 do livro do Kleiberg.

<http://snap.stanford.edu/data/soc-sign-epinions.html>

<http://snap.stanford.edu/data/soc-sign-Slashdot081106.html>

<http://snap.stanford.edu/data/soc-sign-Slashdot090216.html>

<http://snap.stanford.edu/data/soc-sign-Slashdot090221.html>

Para cada uma delas, identifique os diferentes tipos de triângulos formados nesses grafos e verifique se a maior parte encontrada são triângulos balanceados ou desbalanceados. As redes são direcionadas. A abordagem utilizada para a análise de uma direcionada com sinais é parte da questão. Discuta e interprete seus resultados.

- 4) Faça um código para gerar grafos aleatórios $G=(N,p)$ e realize as seguintes tarefas.
 - a. Gere 3 grafos com $N=50$ e $P=0.2$ e mostre a visualização desses grafos.
 - b. Gere um grafo aleatório maior e plote a distribuição dos graus, ou seja, no eixo x plote o grau k dos nodos e no eixo y a contagem de nodos com cada grau k (ou sua porcentagem/probabilidade). Comente se curva se assemelha com as distribuições Binomial/Poisson esperadas.
- 5) Implemente um dos modelos epidemiológicos descritos no artigo dos Zoombies, que é um dos artigos para os quais você deve fazer uma resenha. Você deve implementar o modelo com um código em uma das linguagens: python, C, C++, ou Java. Simule um processo de propagação/contágio e plote um gráfico semelhante aos do artigo gerado a partir dos resultados da implementação do seu modelo.

- 6) Cada aluno deve submeter um projeto final disciplina contendo no máximo 6 páginas que descreva um problema dentro de um dos temas da disciplina, motivação e relevância do estudo e a forma como pretende abordar o problema. Seu projeto pode usar parte dos resultados obtidos no trabalho prático e pode usar os mesmos dados. Porém, note que o projeto deve ser escrito como um artigo científico, apresentando questões de pesquisa ou hipóteses, desenvolvendo uma metodologia para respondê-las e buscando resultados e evidências que buscam responder as questões elencadas.