Social Networks and Online Markets

Homework 1

Due: 8/6/2025, 23:59

Instructions

You must hand in the homeworks electronically and before the due date and time.

The first homework has to be done by each **person individually**.

Handing in: You must hand in the homeworks by the due date and time by an email to `aris@diag.uniroma1.it` that will contain as attachment **(not links to some file-uploading server!)** a .zip or .pdf file with your answers.
After you submit, you will receive an acknowledgement email that your homework has been received and at what date and time. If you have not received an acknowledgement email within 2 days after the deadline then contact Aris.

The solutions for the theoretical exercises must contain your answers either typed up or hand written clearly and scanned.

For information about collaboration, and about being late check the web page.

Problem 1. [**Only for those who choose the project**] Show that for the Watts-Strogatz small-world model for $p = 0$ and as $k,n \to \infty$ the clustering coefficient approaches 3/4. Recall that n is the number of nodes, k is the degree, and p the rewiring probability.

Problem 2. In this homework you need to implement some of the models that we have seen, and measure experimentally some of the graph properties. Of course, each model has its own parameters.

1. The Erdös–Rènyi G_{np} random graph model. Parameters:
   - n: number of nodes
   - p: probability of an edge to exist
2. The Barabási–Albert preferential attachment model. Parameters:
   - n: number of nodes
   - ℓ: number of neighbors that a newly arrived node comes with.

   Assume that the inistial graph is a single node. If it makes your life easier, if multiple edges fall on the same node you can ignore the multiple edges.

The goal is to understand these models, for different parameter combinations. Therefore, for each of these models, you should experiment for different values of the parameters. However, the parameter n should always be high (at least $10K$ but it can be up to the order of millions depending on the power of your computer).

For each of the parameter combinations, compute and report in an organized way:

- Degree distribution (you should display it with a plot)
- Diameter
- Clustering coefficient

For each set of parameters create different graphs and check if the behavior and the values you obtain are the same for each of these graphs.

You are allowed to use a library (such as NetworkX in case you use Python) to handle the graph or compute the graph functions. However, you should implement yourselves the graphs and not use library or other code for that. If you have any questions about what is allowed, feel free to ask Aris.

Now calculate the same three parameters for some of the graphs in the SNAP library: `http://snap.stanford.edu/data/index.html`. Compare the output of your models with those of the real networks and comment on them.

Notice that none of these models has values for the above three values that are realistic. By taking inspiration from them models, try to create a new model that has more realistic values for of these parameters. To create the model you may do whatever you want (e.g. you can combine them, or use some of the other concepts we said in the class).

You should hand in a zip file containing the code of your program results and a report, in pdf format, which will contain the results of your findings. In the report, for each model, you should display a table with the different combinations and the values that you obtain, and for each parameter combination a plot depicting the degree distribution.

As an advice for when you test your programs, first try with smaller values of n to make sure that your program works (e.g., 500 or 1000), then you should try the higher ones.

Problem 3. [Only for those who choose the project]

1. Recall that, given an undirected graph $G = (V,E)$, the densest subgraph $D \subseteq V$ is the set that maximizes
$$\frac{|E \cap (D \times D)|}{|D|}$$
and the sparsest cut $S \subseteq V$ is the set that minimizes
$$\frac{|E \cap (S \times (V \setminus S))|}{\min\{|S|, |V \setminus S|\}}.$$
Give an example of a graph such that $D = S$ and of a graph such that $D \notin \{S, V \setminus S\}$.

2. Let $G = (V, E)$ be a connected, undirected graph with n vertices and Laplacian matrix L. Let λ_2 be the second smallest eigenvalue of L, and let $\mathbf{v}_2 \in \mathbb{R}^n$ be the corresponding eigenvector, normalized so that $\|\mathbf{v}_2\|_2 = 1$.

   Let the coordinates of
$$\mathbf{v}_2(i_1) \geq \mathbf{v}_2(i_2) \geq \cdots \geq \mathbf{v}_2(i_n).$$
   For each $k = 1, 2, \ldots, n-1$, define the subset
$$S_k := \{i_1, i_2, \ldots, i_k\},$$
   and let $E(S_k) \subseteq E$ be the set of edges with both endpoints in S_k.

   Prove that there exists some $k \in \{1, \ldots, n-1\}$ such that
$$|E(S_k)| \geq \frac{\lambda_2}{2}.$$

Problem 4. You are a company that wants to monitor what users post in some site. The users in the site are connected through a network. To find out what users' post you can *follow* some users; when you follow a user, you know what the user posts as well as what her friends post. Assume that you have a budget of k users that you can follow. Provide an approximation algorithm for selecting the set of (at most k) users to follow, such that you maximize the number of users that you know what they post. Assume that you have complete knowledge of the network.