LINMA2472 – Algorithms in Data Science

HW 1 – module "Networks" : part 2

Assignment 1 (part2)

Analyzing the graph

- Find degree assortativity of the network and perform community detection using Louvain algorithm. Visualise the results. What can you tell from the them?
- Write the code for k-core decomposition (do not use the preprogrammed instance in *networkx*) and apply it to the network. What can you infer from it?
- Generate the preferential attachment (Barabasi-Albert) network with similar average degree and size. Perform same operations on this network. Describe any differences or similarities you can spot.

Influence Maximization Problem

Take the network of characters you infer from the first part of the assignment. Imagine there is an important rumour to spread in this network. You want it to quickly reach all the people, thus you want to solve the **influence maximization problem**.

- Implement the greedy algorithm from the lectures and identify the set MI of maximal influence of size k = 5% of the nodes.
- Use the **independent cascade model** on this network in order to compare* the outcomes starting from the obtained set MI with similar size set of nodes of largest degrees and a random selection.

(*Comparison can be made by the total size of people reached by a cascade or by the spreading curve : (t,Y(t)) - curve, where t in discrete time and Y(t) is the total average proportion of "infected" people at time t.)

Report guidelines:

- The **deadline is October 20 at 23:59** (see moodle activity)
- Please submit both your code and your report in a single zip file for every group named « group_x_project1_y1_y2_y3 », where x is your group number and yi are the family names of the group members.
- Write in a concise and structured manner. No long sentences, only relevant information.
- Write a report of no more than 10 pages including figures. Annexes are allowed but the 10 main pages must contain all the information you want to convey.
- You may present your data and the preprocessing steps, but remember that this isn't the main goal of the report

- Any numerical result that can be presented in a table should be presented so.
- Round numbers up to 3rd digit, unless it's really necessary. Don't copy-paste 10 digits floats.
- Plots must be easy-to-read. Must include labels on axes, legend if more than one curve is shown, title or a caption, explaining what the plot is about.
- Network properties (k-core shell, community index, etc) can be **visualised** in color. When doing so, it's a good practice to add a colourbar (k-core shell) or a summary of each or most representative communities.