# LINMA2472 – Algorithms in Data Science

## HW 1 – module "Networks": part 2

Take the network of characters you constructed in the first part of the homework. In this part of the homework. This second part will solely focus on this network.

**Analysing the communities of the graph**

- Visualise the network and make preliminary observations about it (is it highly connected, can you already notice communities, …).
- Find degree assortativity of the network. Comment you results.
- Use the Louvain algorithm to detect the various communities of your network. Repeat this experiment using spectral clustering on the Laplacian. You can choose your parameters as you like. The only requirement is to justify your choices.
- Compare the two methods and comment your results. Do your results match your expectations? Are the communities that you discovered related to the story of your book?

**Maximising the influence in the graph**

- Imagine there is an important rumour to spread in your network. You want it to quickly reach all the people, thus you want to solve the **influence maximisation problem**. Implement the greedy algorithm from the lectures and identify the set MI of maximal influence of size k = 5% of the nodes.
- Implement the **independent cascade model** on this network and use it 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.
- Generate a Barabasi-Albert network with the similar average degree and size as your original network. Perform the greedy algorithm again and compare the results you obtain with the results on the original character network. Comment your results.

## Report Guidelines

- The deadline is on the 23rd of October at 2359 hours
- Please submit both your code files and your report (pdf format, preferably in LaTeX) in one single .zip file for every group. Name your file

"group_x_project1_y1_y2_y3", where x is your group number and yi are the family names of every member of the group.

- Write in a concise and structures manner, please avoid long sentences and only include relevant information
- Write a report of no more than 10 pages including figures. Annexes are allowed but the main 10 pages must contain all the information you want to convey
- You may present your data preprocessing steps, but remember that this isn't the main goal of the report
- Any numerical result can be presented in a table should be presented so
- Round number to 3$^{rd}$ digit unless it is really necessary. Co not copy-paste 10 digits floats
- Plots must be clear and easy to read. Do not forget labels on axes, legends, titles, captions and so on
- Feel free to include colours in your plots (but do not forget colour bars or legends