LINMA2472 – Algorithms in Data Science

HW 1 – module "Networks"

This is the first part of the first assignment (the next part will follow after next week's lecture). This assignment is to be completed in groups of 2 or 3, please form your groups on moodle (using the activity called "Group choice for assignment 1"). If you need help to look for teammates, use the "Teammate finder" forum. If you have any practical questions, please email us on remi.delogne@uclouvain.be, bastien.massion@uclouvain.be or brieuc.pinon@uclouvain.be .

The exact deadline for this assignment (this first half with the next part to come) will be confirmed later. Instructions for the next part will follow shortly.

Goal: Build a co-occurrence network of characters

Please choose one of the following options:

- Find an appealing *book* (for example, use the Project Gutenberg (gutenberg.org) to find the text), parse the textual information in order to reconstruct the co-occurrence network of characters. For example, two characters can be linked if they appear in the same paragraph (but feel free to explore other setups).

- Find a *screenplay* from your favorite movie (there are many resources can be found by Googling, for example, https://thescriptsavant.com/free-moviescreenplays-am/). Convert the .pdf to text using any online tool and parse the textual information to reconstruct the co-occurrence network of characters, where two characters can be linked if they appear in the same scene. Scenes are usually distinguished in bold notation.

You are free to choose any manuscript as long as it has many different characters (ideally more than 50). There are examples of text processing tools with Python on moodle in the "Python tutorial" activity. You are however allowed to use any technique with python by the time everything is clearly stated in your code.

In the second part of the homework, you will then apply some of the methods taught in the course to your co-occurrence graph and produce a report to explain your results. For this first half however, there is nothing to hand in, this is just a first step to help you with the second part. Further instructions will be provided later.