Exam

Social and Graph Data Management
Université Paris-Saclay, M2 Data Science

December 15th, 2023

This exam subject consists of 3 exercises and has 3 pages. You must **select 2** of the 3 exercises and only answer those 2 exercises (so you must leave out one of the exercises).

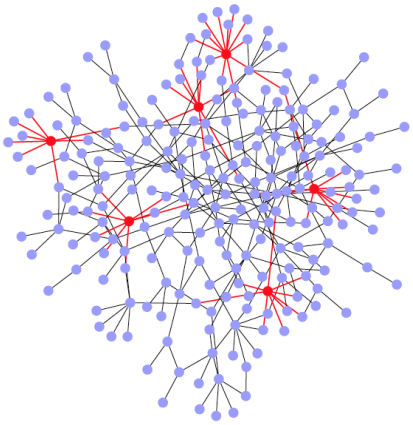
The exam is *strictly personal*: any communication or influence between students, or use of outside help, is prohibited.

Exercise 1 – Communities (10 points)

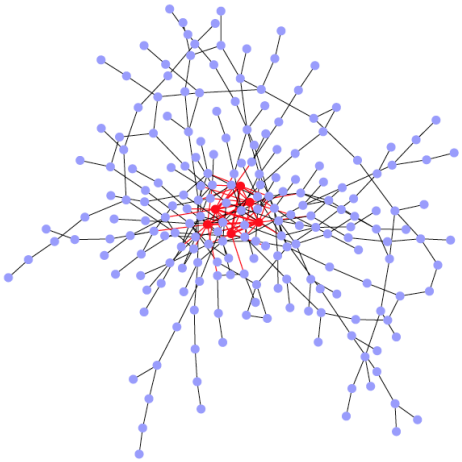
Question 1. Give an intuition why it is extremely unlikely that there are 2 giant components in a randomly wired network.

Question 2. Explain the hypotheses used for community detection in social networks.

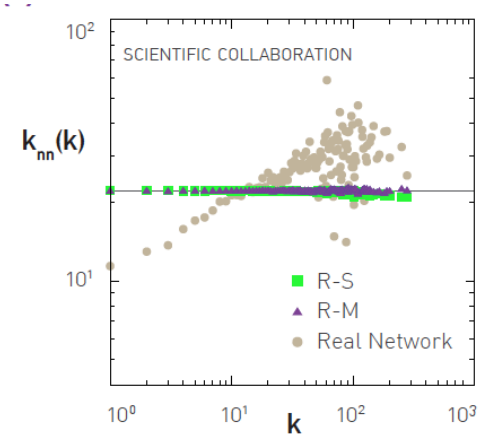
Question 3. For each of the following 4 networks, indicate if they are assortative, disassortative or neutral, and justify ($k_{nn}(k)$ is the degree correlation function, see on the last page a reminder for the definition). For networks c and d (scientific collaboration and metabolic network), you should discuss the assortativity of the "real network" but also what additional information is provided by the green square plots (network perturbed randomly while preserving the degree distribution and without multilinks).



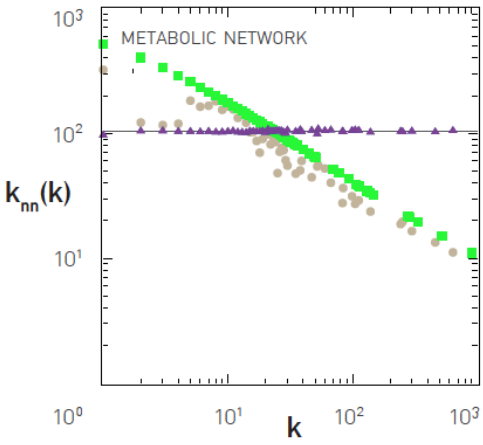
network a



network b



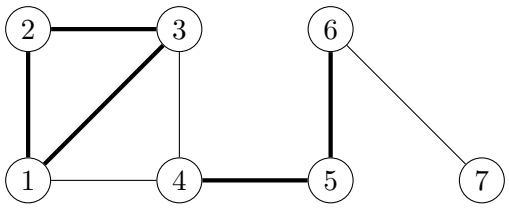
network c



network d

Exercise 2 – Graph Measures (10 points)

Consider the graph G in the following figure:



Question 1 Represent the graph as an *adjacency matrix*.

Question 2 Write down the degree distribution of G, and the average degree $\langle k \rangle$.

Question 3 Compute the *diameter* $d_{\max}$ of G, and show a path of length $d_{\max}$ in G.

Question 4 Assume that the graph was computed using an Erdős–Rényi *random network* model with parameter p. What is the value of p? Explain how you found it.

Question 5 If we assume that links represented with thick edges are strong links, and the other links are weak, does the graph G satisfy the strong triadic closure property ?

Question 6 Give at least 2 measures for which social networks generally behave differently from the Erdős–Rényi model. Explain succinctly how they differ.

Question 7 Recall the phase transitions observed for connected components when a random graph is generated using Erdős–Rényi model.

Exercise 3 – Power laws (10 points)

Question 1 Depending on the value γ, which moments exist for power-law distributions with exponent γ? Prove it (you may consider directly the continuous version of the distribution).

Question 2 Explain why networks following a power-law distribution are called scale-free (provide two explanations for the term).

Question 3. Explain the differences between how robust the social networks are as compared to random network under targeted attacks that remove the optimal nodes to disconnect the network.

Reminder

- degree correlation function: $k_{nn}(k) = \sum_{k'} k' P(k', k)$ where $P(k', k)$ is the conditional probability that by following a link of a degree-k node, we reach a degree-k' node.