Exam

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

December 16th, 2024

This exam subject consists of 5 exercises and has 3 pages.

You must **select 4** of the 5 exercises and only answer those 4 exercises
(this means 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 (5 points)

We recall the formula for modularity:

$$M = \sum_c \left(\frac{L_c}{m} - \left(\frac{k_c}{2m} \right)^2 \right)$$

where L_c is the number of edges within the community, k_c the sum of the (total[1]) degrees of nodes in the community, and m the number of edges in the graph.

Question 1. Give the hypotheses on which the community identification approaches considered in the lecture rest.

Question 2. What are the minimal and maximal values of modularity?

Question 3. Based on the formula above, show how the modularity evolves when merging 2 communities in general, which are connected by x edges. And show how it evolves when the 2 communities merged contain respectively 4 nodes of degree 2 and 3 nodes of degree 3, and are connected by a single edge, in a graph of 24 edges.

[1] we naturally count edges within the community as well as edges leaving the community

Exercise 2 – Width measures: separators (5 points)

We define a vertex-weighted graph as an (undirected) graph $G = (V, E)$ together with a function w mapping each vertex v to a weight $w(v) \in \mathbb{N}_{>0}$. A (plain) graph ican be viewed as a particular case of vertex-weighted graphs, where the weight function maps each vertex to 1. A $(k, 1/2)$ separator for a vertex-weighted graph G is a set S of at most k vertices such that each connected component in $G - S$ has weight at most $1/2 \cdot \sum_{v \in V} w(v)$.

We recall that, given a graph $G_0 = (V, E)$, there is an algorithm with running time $2^{O(k)} \cdot |V|$ which takes as input G_0 and k and returns a tree decomposition of width at most k of G_0 if there is one, or reports that the tree-width is larger than k otherwise. Similarly, for every constant k, there is an algorithm with running time $O(|V|)$ which returns a path decomposition of width k if there is one, or reports that the pathwidth is larger than k otherwise.

Question 1. Give a $(2, 1/2)$ separator for the weighted cycle whose nodes are labeled $1, \ldots, 10$ and whose edges are $\{(i, i+1) \mid 1 \leq i \leq 9\} \cup \{(10, 1)\}$, such that for all $i \leq 5$, $w(i) = 1$ and for all $i \geq 6$, $w(i) = 100$.

Question 2. Give an algorithm that takes as input a (weighted) graph G of pathwidth at most k (for some constant k, ex. $k = 20$), and returns a $(k + 1, 1/2)$ separator of G. Can your algorithm return a better guarantee?

Exercise 3 – Width measures (5 points)

Question 1. Give the treewidth of the graph G from Figure 1, and give a tree decomposition of minimal width of this graph.

Question 2. Give the diameter of the graph G from Figure 1 and its degree distribution.

Exercise 4 – Epidemics, network models (5 points)

We recall that in the SIS model, nodes get infected by their infected contacts at rate β and can heal at rate μ (healing does not grant them immunity), so the equation is :

$$\frac{\partial i}{\partial t} = \beta \langle k \rangle i(1 - i) - \mu i$$

and the characteristic time for the SIS model is given by:

$$\tau = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \mu \langle k \rangle)}$$

Question 1. Give the possible final outcomes of an epidemics in the SIS model.

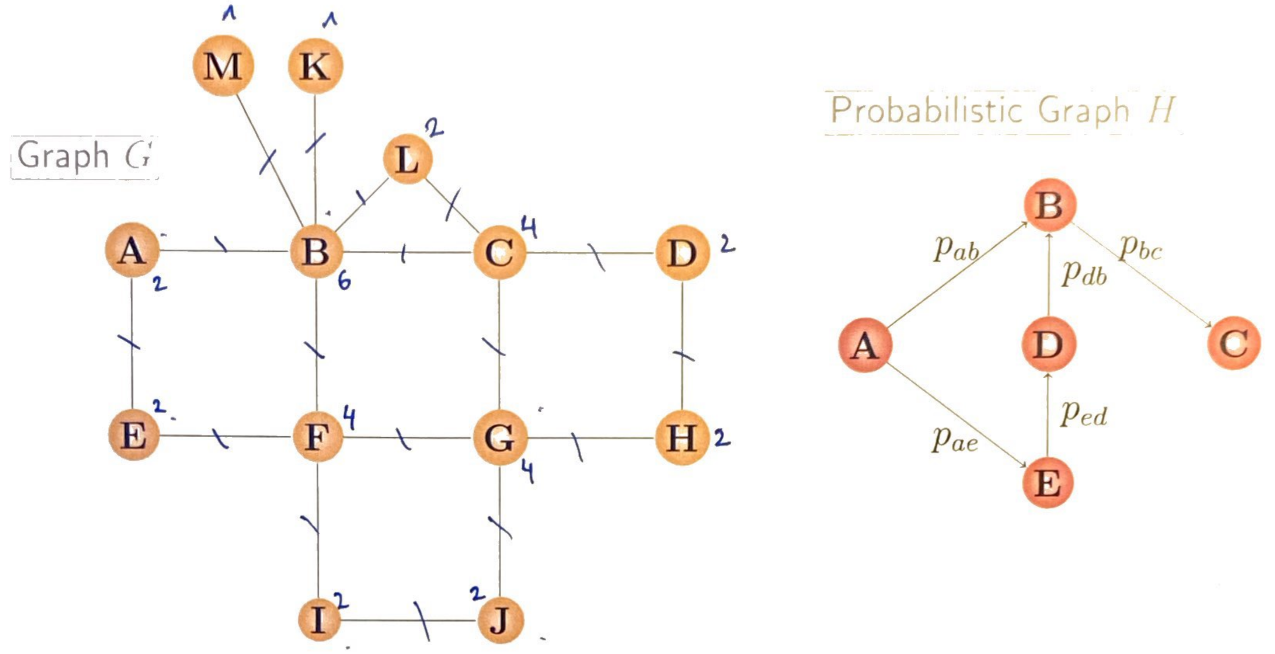

Figure 1: Graph G and probabilistic graph H

Question 2. Derive from the formula the numeric value of the characteristic time for the SIS model on a scale free network (exponent $2 \leq \gamma < 3$) of extremely large size. Justify it with intuition on how the "epidemics" spreads on scale-free networks.

Question 3. Give the average distance for the regimes of power-law degree distributions with respective exponent $2 \leq \gamma < 3$ and $\gamma > 3$. Why don't we generally consider networks with power-law degree distribution of exponent less than 2? ✓ ~

Exercise 5 – Probabilistic graphs, random graphs and network models (5 points)

Question 1. In the probabilistic graph H from Figure 1, the probability of each edge from x to y is denoted by p_{xy}. Give a possible world H_0 for this graph in which C is reachable from E, and the probability of H_0. ✓

Question 2. In the probabilistic graph H, give the probability that C is reachable from A. ✓

Question 3. Give (up to a multiplicative constant) the degree distribution for Erdos-Renyi random graphs and scale-free networks. ✓ ~

Binomial Distribution

$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$ → prob of not having the remaining N-1-k edges

subíndice — n° of ways to choose k edges from N-1 possible connections

exponente → prob of having k edges

$p_k \sim k^{-\gamma}$ $\quad p_k = \frac{c}{k^\gamma}$ where $C = \frac{1}{\zeta(\gamma)} = \left(\sum_{k=1}^{\infty} k^{-\gamma}\right)^{-1}$

↳ Riemann Zeta Function used to normalize so that degrees probabilities sum to one