

# Graph Mining: Exercises for the Final Exam

Mauro Sozio

`mauro.sozio@telecom-paristech.fr`

Below you can find a list of exercises related to our course on graph mining. The same or similar exercises will be asked at the final exam. Unless otherwise specified, you should provide a self-contained proof, i.e. you may not use any of the results we saw during our course.

## 1 PageRank

**Exercise 1** Consider the variant of the PageRank Algorithm presented on the slides “Community Detection with seed nodes”. a) Provide an example showing that the Markov Chain defined by the matrix  $A$  in slide 6 might not be ergodic. b) Show that there is a unique stationary distribution for such a Markov Chain. c) Show that the PageRank algorithm (slide 7) converges to the unique stationary distribution.

## 2 Dense Subgraphs

**Exercise 2** Let  $G = (V, E)$  be an undirected graph. Let  $H_1 = (V_1, E_1)$  and  $H_2 = (V_2, E_2)$  be two densest subgraphs in  $G$ , i.e., for any subgraph  $H = (V_H, E_H)$  of  $G$  it holds that  $\frac{|E(H)|}{|V(H)|} \leq \frac{|E_i|}{|V_i|}$ ,  $i = 1, 2$ . Let  $\hat{H} = (V_1 \cap V_2, E_1 \cap E_2)$  be the graph obtained by the intersection of  $H_1$  and  $H_2$ , while let  $\bar{H} = (V_1 \cup V_2, E_1 \cup E_2)$  be the graph obtained by the union of  $H_1$  and  $H_2$ . Give a lower bound (as tight as you can) on the density of  $\hat{H}$  assuming that  $E_1 \cap E_2$  is not empty. Give a lower bound (as tight as you can) on the density of  $\bar{H}$ .

**Exercise 3** Show that for any subgraphs  $H_1$  and  $H_2$  of an input graph  $G = (V, E)$  with  $\rho(H_1) \neq \rho(H_2)$ , it holds that  $|\rho(H_1) - \rho(H_2)| \geq \frac{1}{|V|^2}$ . Give an example for which  $|\rho(H_1) - \rho(H_2)|$  is  $\Theta\left(\frac{1}{|V|^2}\right)$ .

**Exercise 4** Let  $G = (V, E)$  be an undirected graph, the  $k$ -core of  $G$  is a subgraph of  $G$  with maximum number of nodes with every node having degree at least  $k$ . Let  $\bar{k}$  be largest such that there is a  $\bar{k}$ -core in  $G$ , which we denote with  $H$ . Show that  $\rho(H) \geq \frac{\rho_O}{2}$ , where  $\rho_O$  is the density of a densest subgraph in  $G$ .

**Exercise 5** Discuss how the algorithm for computing a 2-approximation for the densest subgraph problem, presented during our course, can be implemented in linear time in the size of the input. In particular, you should discuss which data structures should be used and how they are used.

**Exercise 6** Given an undirected graph  $G = (V, E)$ , consider the following linear program (LP) formulation for the densest subgraph problem:

$$\begin{aligned}
& \max && \sum_{ij \in E} x_{ij} \\
& \text{s.t.} && x_{ij} \leq y_i && \forall ij \in E \\
& && x_{ij} \leq y_j && \forall ij \in E \\
& && \sum_{i \in V} y_i \leq 1 \\
& && x_{ij}, y_i \geq 0 && \forall i, j.
\end{aligned}$$

Let  $\bar{x}, \bar{y}$  be an optimum solution for the above LP. Show that the graph induced by the nodes with maximum  $\bar{y}$  values is a densest subgraph.

**Exercise 7** Given an undirected graph  $G = (V, E)$ , show that for any  $\epsilon > 0$  there is a subgraph  $H$  of  $G$  such that : a)  $\rho(H) \geq \frac{\rho_O}{2(1+\epsilon)}$ , b) the diameter of  $H$  is  $O\left(\frac{\log |V|}{\epsilon}\right)$ , where  $\rho_O$  is the density of a densest subgraph in  $G$ . Develop an algorithm running in polynomial time which finds a subgraph  $H$  satisfying a) and b).

### 3 Influence Maximization

**Exercise 8** Let  $X = x_1, \dots, x_n$  be a set of rational numbers in  $[0, 1]$ . Let  $\sigma_X(A)$  be the number of active nodes at the end of the process in the linear threshold model (LT) (with initial set  $A$ ) where  $\theta_{v_i} = x_i$ ,  $i = 1, \dots, n$ . Show that there exists an instance of LT and  $X$  such that  $\sigma_X(\cdot)$  is not submodular.