Graph Mining: Exercises for the Final Exam

Pierluigi Crescenzi
pierluigi.crescenzi@irif.fr

Below you can find a list of exercises related to our course on graph mining. In all exercises, apart from Exercise 3, you can assume that the graph is connected, undirected, and unweighted. The same or similar exercises will be asked at the final exam. Two of the following exercises are significantly easier than the others: Exercise 0 is to find these two exercises.

Exercise 1
Prove that, when applied to a tree, the 2-sweep algorithm always compute the correct diameter value.

Exercise 2
A graph $G$ is $\delta$-hyperbolic if for any four vertices $a$, $b$, $c$, and $d$ of $G$, the two largest of the three sums $S_1 = d(a, b) + d(c, d)$, $S_2 = d(a, c) + d(b, d)$, and $S_3 = d(a, d) + d(b, c)$ differ by at most $2\delta$. Prove that, if a graph is $\delta$-hyperbolic, then the value returned by the 2-sweep algorithm is no smaller then $D - 2\delta$, where $D$ is the diameter of the graph.

Exercise 3
Consider the graph $S_{k,p}$ made by a $k$-clique and a directed $p$-cycle (see figure below where $k = p = 8$).

A centrality measure satisfies the size axiom if for every $k$ there is a value $P_k$ such that for all $p \geq P_k$ in $S_{k,p}$ the centrality of a node of the $p$-cycle is strictly larger than the centrality of a node of the
$k$-clique, and if for every $p$ there is a value $K_p$ such that for all $k \geq K_p$ in $S_{k,p}$ the centrality of a node of the $k$-clique is strictly larger than the centrality of a node of the $p$-cycle. Determine which centrality measure (among the ones we have studied during the course) satisfies the size axiom (in the case of the degree centrality, you can refer to the in-degree value of a node).

**Exercise 4**

Adapt the sampling techniques we have used to compute an approximation of the distance distribution in order to compute an approximation of the closeness centrality of all nodes of a graph. In particular, describe the algorithm, bound the error probability, and analyse the time complexity.

**Exercise 5**

Combine the Brandes’ algorithm and the sampling techniques we have used to compute an approximation of the distance distribution in order to compute an approximation of the betweenness centrality of all nodes of a graph. In particular, describe the algorithm, bound the error probability, and analyse the time complexity.

**Exercise 6**

The load centrality measure is defined by referring to the following process. Each node sends a unit amount of some commodity to any other node. Starting from the respective source $s$, the commodity $c_{s,t}$ is always passed to the adjacent vertices closest to the target $t$, and in case there is more than one such vertex the commodity is divided equally among them. The load centrality of a vertex is the total amount of the commodity passing through it during all exchanges (this value has to be normalised similarly to what has been done with the betweenness centrality). (1) Show an example in which the load centrality is different from the betweenness centrality. (2) What is a sufficient condition for having the load centrality equal to the betweenness centrality? (3) Adapt the Brandes’ algorithm in order to compute the load centrality of all nodes.

**Exercise 7**

A fastest journey from $u$ to $v$ from time $t$ in a temporal graph is a journey whose departure time is no less than $t$ and whose duration is minimum. Design an algorithm to compute the duration of the fastest path from a source vertex $s$ to every other vertex of the temporal graph, and analyse its time complexity.

**Exercise 8**

Prove that any undirected temporal clique of four nodes with distinct labels admits a spanner with five edges.