Networks Analysis and Mining

Sorbonne Université
Master d'Informatique spécialité Réseaux

Final Exam

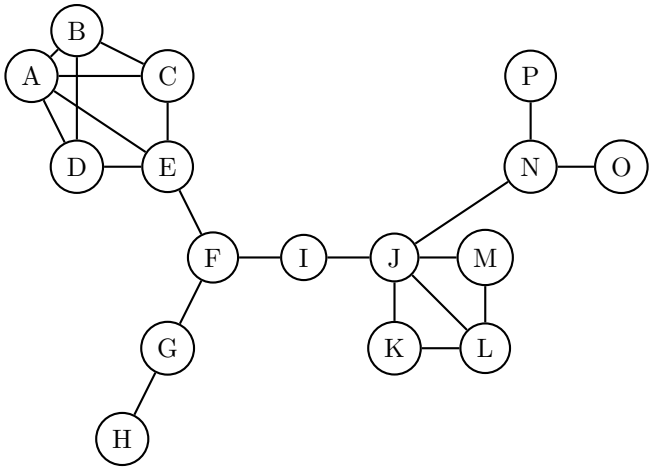
February 16th 2021

Maximilien Danisch and Lionel Tabourier

The exam is 2hrs long. All documents are authorized. Points are only indicative. Thank you to answer on a separate paper. Good luck!

Exercice 1 — *Measures on a small graph (6pts)*

Consider the following graph:



Q1. The degrees of the nodes are the following:
1: P, O, H
2: G, I, K, M
3: B, D, C, F, L, N
4: A, E
5: J
Hence the coordinates of the points of the ICDD: (1, 16), (2,13), (3,9), (4,3), (5,1), (6,0).

Q2. there are 6 triangles in the graph: ABC, ABD, ACE, ADE, JKL, JLM.

Q3. $cc(C) = 2/3$, $cc(J) = 1/5$

Q4. composition of each core:
core 1: G, H, O, P, N
core 2: F, I, J, K, L, M
core 3: A, B, C, D, E

Q5. For example, high degree, low betweenness: A ; low degree, high betweenness: I

Exercice 2 — *Graph models (4pts)*

We give below measures realized on graphs. Tell in each case if the model proposed is consistent with the measures observed and if not, explain why.

Q1. No: with an ER model, the clustering would be of the same order as the density.

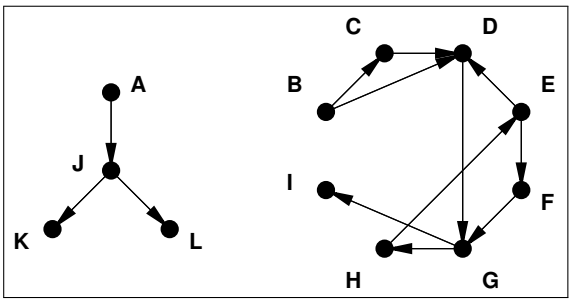
Q2. No: with a BA model, all nodes are part of the largest connected component.

Q3. Yes, the properties are consistent with a configuration model.

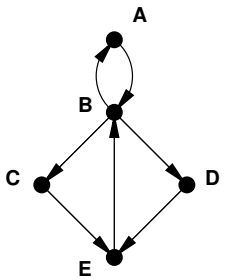
Q4. No, with a SW model, the degree distribution would be homogeneous.

Exercice 3 — *About PageRank (4pts)*

Q1. a) nodes in the largest strongly connected component: D,E,F,G,H
b) nodes upstream of the largest strongly connected component: B,C
c) nodes downstream of the largest strongly connected component: I



Q2. Considering



Suppose that x is the PageRank of B, then we must have
PR(A) = x/3 , PR(C) = x/3 , PR(D) = x/3 , PR(E) = 2x/3
so using the nomalization (sum of all PR is 1): x=3/8

Exercice 4 — *Course understanding (4pts)*

In this exercise, you are asked to answer in one or two simple sentences.

Q1. In complex networks analysis, why do we nearly always favor using the table of adjacency lists format over the adjacency matrix format when representing graphs?

Because an adjacency matrix takes too much place in RAM.

Q2. Why do we say that weak links have an essential role in a social network?

Because they are bridges between communities.

Q3. What is the initial state of every agglomerative community detection algorithm? Give an example of such an algorithm.

In the initial state, all nodes are in their own community with an agglomerative algorithm. This is the case for the Louvain algorithm (but not for Label Propagation).

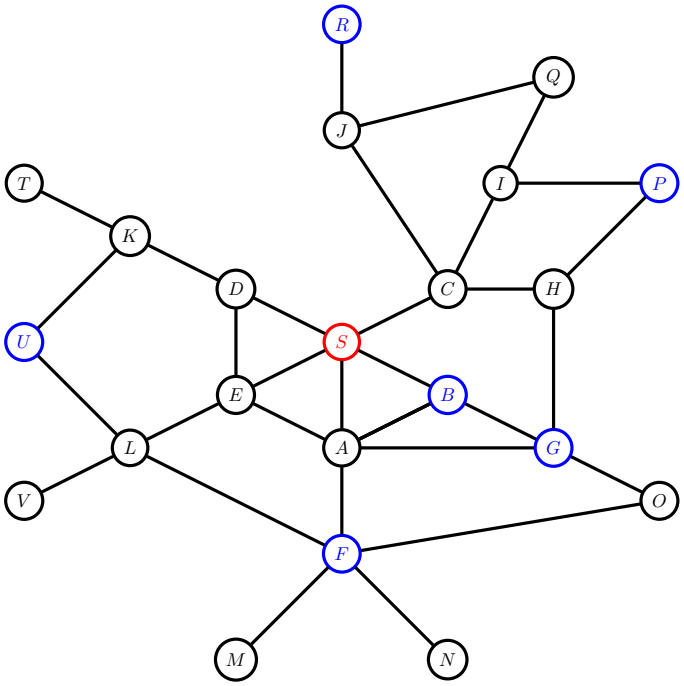
Q4. Give an advantage and a disadvantage of a recommendation system based on collaborative filtering in comparison with a content-based recommendation system.

For instance, collaborative filtering does not demand to have an expert knowledge of the items, but it does not allow to recommend an item which has never been evaluated.

Exercice 5 — *Simulating traceroute (8pts)*

In this exercise, we simulate a traceroute measurement from S to several destinations using a restricted BFS algorithm, following the principle described in the *Internet topology metrology* course.

The structure of the graph G under examination is given below:



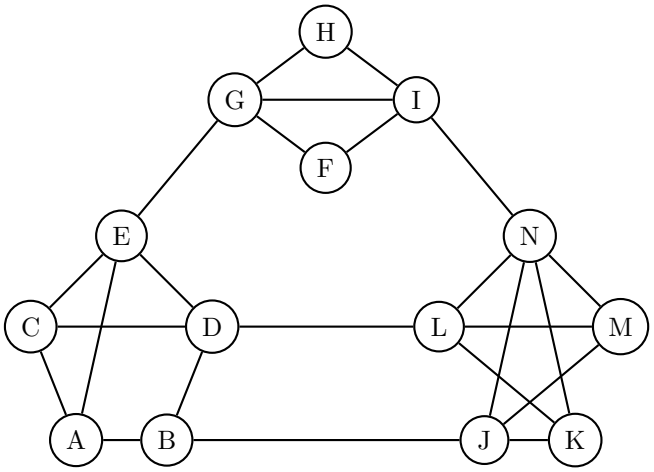
Q1. b) Possible restricted BFS yields tree with branches:
$S-D-K-U$
$S-C-J-R$
$S-C-I-P$
$S-A-F$
$S-B-G$

Q2. a) $k=1$: 5, $k=2$: 4, $k=3$: 6, $k=4$: 4, $k=5$: 3.
b) With our restricted BFS: $k=1$: 5, $k=2$: 6, $k=3$: 1, $k=4$: 1.
c) Underestimate the degree of higher degree nodes.

Q3. a) Links at distance 1: $S-D$, $S-E$, $S-A$, $S-B$, $S-C$
Links at distance 2: $D-E$, $D-K$, $E-A$, $E-L$, $A-B$, $A-F$, $A-G$, $B-G$, $C-H$, $C-I$, $C-J$
Links at distance 3: $K-T$, $K-U$, $L-U$, $L-V$, $L-F$, $F-M$, $F-N$, $G-O$, $G-H$, $H-P$, $I-P$, $I-Q$, $J-Q$, $J-R$, $O-F$
b) fraction of distance 1 links observed: 4/5
fraction of distance 2 links observed: 5/11
fraction of distance 3 links observed: 1/5
As in the lectures, we observe that the further we are from the source, the worst is the link sampling.

Exercice 6 — *Community detection using conductance (8pts)*

In this exercise, we consider the following graph $G = (V, E)$



Q1. $\phi(\{H\}) = 1$

Q2. $\phi(\{G, H, I, F\}) = 2/12 = 1/6$

Q3. $\phi(\Pi_{iso}) = 1$

Q4. $\phi(\Pi_{all}) = \infty$

Q5. $\phi(A, B, C, D, E) = 3/17$; $\phi(F, G, H, I) = 2/12$; $\phi(L, M, N, J, K) = 3/19$
so $\phi(\Pi_{good}) = max(0.167, 0.177, 0.158) = 0.177$

Q6. $\phi(A, B, C, E) = 2/7$; $\phi(F, G, H, I) = 2/12$; $\phi(B, L, M, N, J, K) = 4/22$
so $\phi(\Pi_{bad}) = max(0.167, 0.286, 0.182) = 0.286$

Q7. If we use choice a) and a set of the partition as a very high ϕ, we simply ignore the optimization of the rest of the partition which is not appropriate, thus choice b) seems more appropriate.