Nathalie VIALANEIX
Année 2019/2020

M2 in Statistics & Econometrics
Graph mining
Exam February 5th, 2020 - 2 hours

This exam uses the files available in the zip file http://www.nathalievialaneix.eu/teaching/m2se/star_wars.zip. These files were obtained from the github repository https://github.com/evelinag/StarWars-social-network and contain co-appearance information from star wars movies (episodes 1-6). More precisely, the zip file contains:

- two text files, respectively containing the characters (nodes) and their relations (edges) in the first episode of Star Wars;
- an rda file, that contains 5 igraph objects that are the networks of episodes 2-6 respectively.

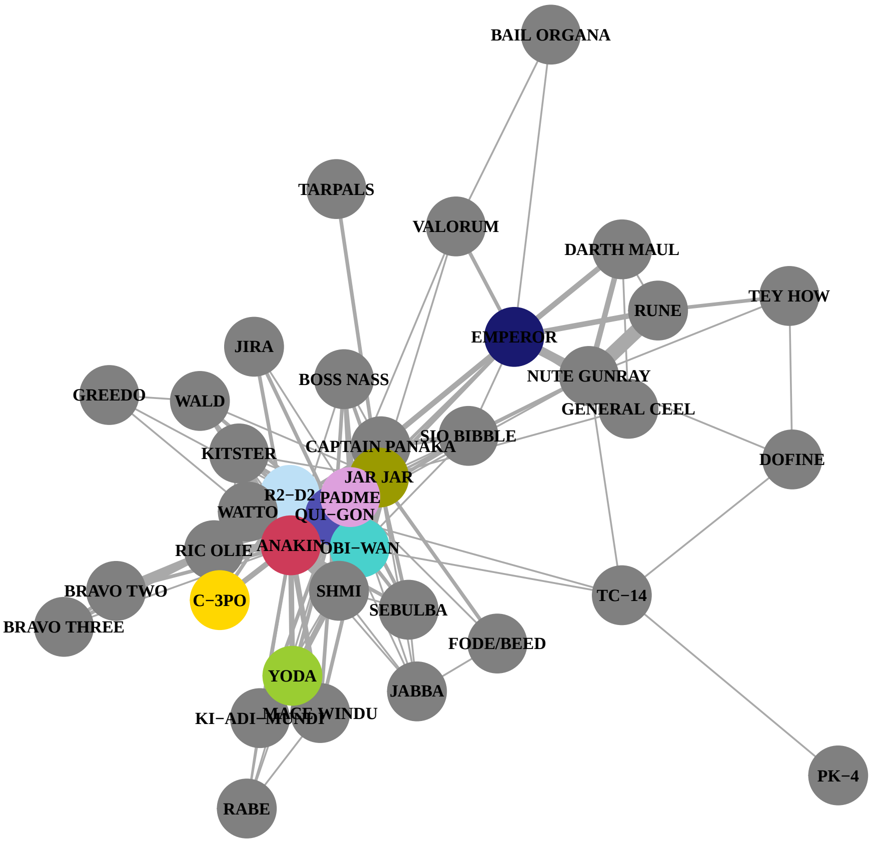
An analysis of the dataset is provided at http://evelinag.com/blog/2016/01-25-social-network-force-awakens/index.html and can help you check your results but be aware that any answer that would not be fully justified by a script will not be evaluated as correct.

Answer the questions below. The answers must include comments (if requested), R script and output of the script. Most questions are independant so do not stay too long on one question if you don't know how to answer it. You are strongly advised to use an RMarkdown file. Answers must be sent by email at nathalie@nathalievialaneix.eu. You are responsible to check that I have received your email properly before leaving the exam room.

MTFBWY

Exercice 1 Creating the graph

1. Import the edge list (`net1-edges.csv`) into your R session and create an igraph object, from the first two columns of the edge list. The network has to be an undirected and unweighted graph and named `net1`.
2. Create an edge attribute `weight` from the third column of the edge list.
3. How many vertices and edges does the network have?
4. Import the node description (`net1-nodes.csv`) into your R session and create two vertex attributes: `name` and `color` (the vertices in `net1-nodes.csv` are sorted in ascending number of the node identifier used in `net1-edges.csv` so that node "1" is in the first row, node "2", in the second, ...).
5. Print the network (copy/paste the result in your answers if you don't use an RMarkdown file). What is the meaning of each of the 4 characters that are given in the first row of the output, just after `IGRAPH` and the graph identifier number and before the number of vertices?
6. Create the graph attribute `layout` with a relevant layout for the graph and make a plot of the graph, using the weight attribute to display the edge width.



Exercice 2 Main facts about the graph

Load all the graphs included in the rda file with:

In your R session, you have loaded 5 igraph objects named `net`X with $X \in \{2, 3, 4, 5, 6\}$ that correspond to the co-appearance networks of episode X.

1. Give the number of edges and nodes of the co-appearance networks for episodes 2 to 6.

2. Among these 6 networks (the 5 loaded networks and the one that you have created in the first exercice), which one(s) is (are) not connected? For this (these) network(s), give the number of nodes in the largest connected component.

3. Explain what is the density of a graph and its transitivity. Provide these two values (unweighted versions) for the 6 networks and comment on the difference between the networks and between the two quantities.

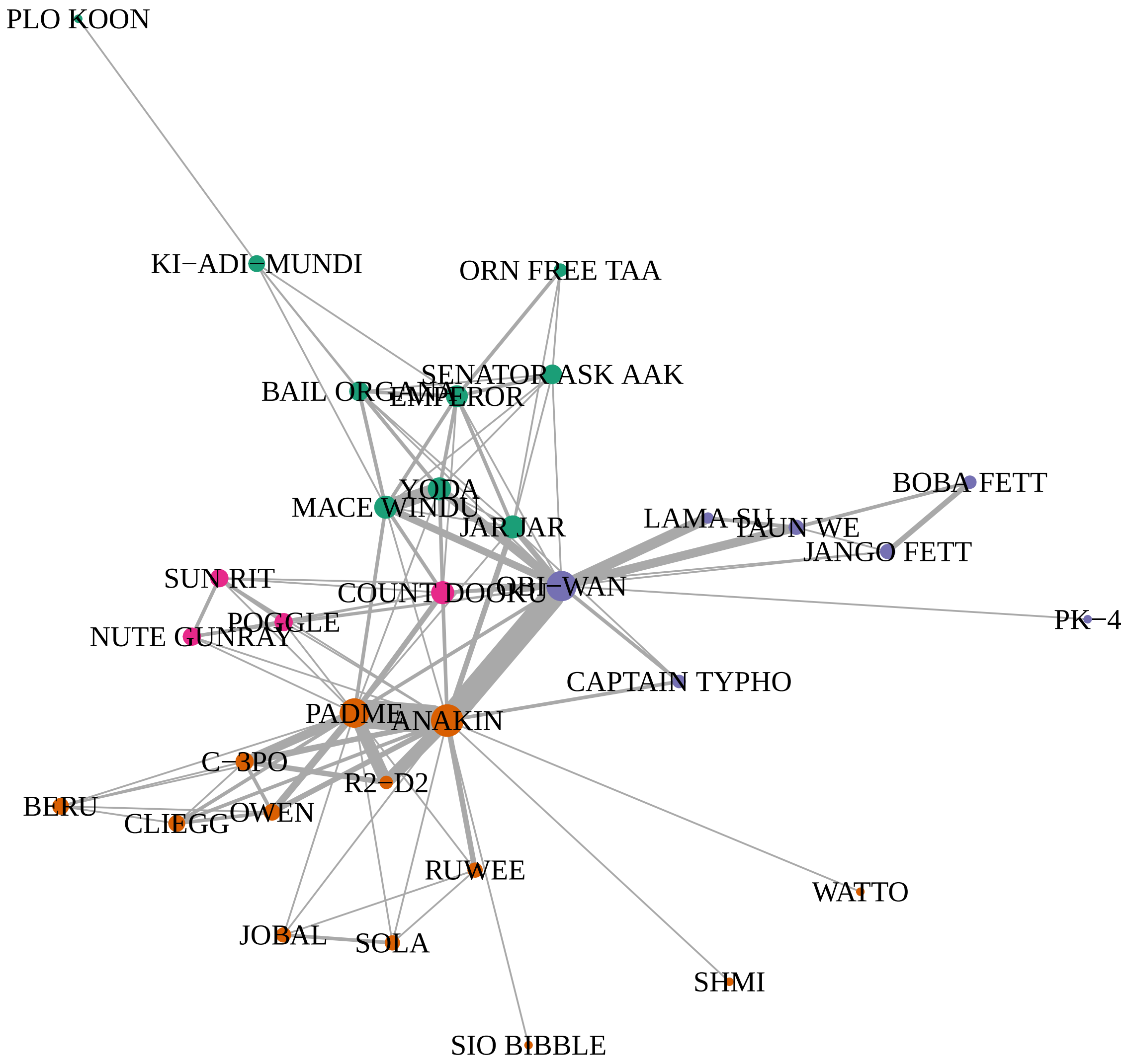
4. In each of the 6 networks, find the character with the largest strength and the one with the largest (unweighted) betweenness. In which episodes, the character with the highest degrees is not the one with the highest betweenness. For episode 3, what is the rank, in terms of degree (by decreasing value), of the character with the highest betweenness. Is the result of your analysis in accordance with Lucas's declaration about these episodes:

> It really is the story of the tragedy of Darth Vader, and it starts when he's nine, and it ends when he's dead.

Exercice 3 Clustering

1. Use the Louvain to compute a clustering of the nodes of the `net2`. What is the modularity of the solution and how many clusters does it have?

2. Display the network with the vertex colors representing the clusters and the vertex area proportional to their degrees. Your graph should look like the one below (colors come from the `RColorBrewer` palette "Dark2").


PLO KOON
KI−ADI−MUNDI
ORN FREE TAA
SENATOR ASK AAK
BAIL ORGANA
EMPEROR
YODA
MACE WINDU
JAR JAR
LAMA SU
TAUN WE
BOBA FETT
JANGO FETT
SUN RIT
COUNT DOOKU
OBI−WAN
POGGLE
NUTE GUNRAY
PK−4
CAPTAIN TYPHO
PADME
ANAKIN
C−3PO
R2−D2
BERU
CLIEGG
OWEN
RUWEE
WATTO
JOBAL
SOLA
SHMI
SIO BIBBLE