Nathalie VIALANEIX
Année 2018/2019

M2 in Statistics & Econometrics
Graph mining
Exam - February 13, 2019 - 2 hours

This exam uses the material available at http://www.nathalievialaneix/teaching/m2se/marvel.zip. This file is a compressed (ZIP) file that contains two other files:

- `marvel_edgelist.txt` is a list of interactions inside the Marvel Universe. Two Marvel characters are considered linked if they jointly appear in the same Marvel comic book. Characters are identified by integers;
- `marvel_vertices.txt` is a two column data table that gives the correspondance between the integer used in the previous file and the character's name (some names are duplicated, which we will ignore in this basic study).

The data are taken for the study `alberich_etal_p2002.pdf`, which is an unpublished article:

> Alberich R., Miro-Julia J. and Rossello F. (2002) Marvel Universe looks almost like a real social network. Preprint arXiv 0202174?

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 RMarkdown file. Answers must be sent by email at mailto:nathalie@nathalievialaneix.eu. You are responsible to check that I have received your email properly before leaving the exam room.

Exercice 1 Creating the graph

1. Import the two files `edgelist.txt` into your R session and create an igraph object, `marvel_net` from the edge list. The network has to be an undirected and unweighted graph.
2. Set a node attribute called `name` using the second text file.

   If you have not been able to do this question, its result (the network `marvel_net`) can be loaded with `load("marvelNet.rda")`.
3. How many vertices and edges does the network have?
4. Is the network connected? If not, extract its largest connected component in `marvel_lcc`.

   If you have not been able to do this question, its result (the network `marvel_lcc`) can be loaded with `load("marvelLCC.rda")`. The rest of the exam will be done using this network.

Exercice 2 Node mining

1. Compute the node degrees. Who are the characters which the five biggest degrees?
2. Compute the node betweennesses. Who are the characters which the five biggest betweennesses (if you are using RMarkdown, I strongly advice you to cache this result that can take long to run)?

Exercice 3 Clustering

1. Perform a node clustering with a modularity based approach (the Louvain algorithm is advised). How many clusters do you find and what is the modularity of the clustering?
2. In which cluster is classified my favorite character (Wolverine / Logan, id 6306) and what is the size of this cluster?
3. Extract Logan's cluster and plot it with the default layout.