Graph Algorithms
Final Project

Master in Data Analytics, 2022

1 Assignment

Find and download a network from the web that you wish to analyze. Two possible sources are networks.skewed.de/ or networkrepository.com, but feel free to select other datasets of your preference. Provide a brief explanation of why you chose to analyze that network and of what you want to achieve through the analysis.

Describe the network and its basic properties through the instruments introduced during the course. A set of possible steps for the analysis follows, but you might focus on just some of these steps or consider aspects on the network that are not listed below:

- provide a graphical representation of the network;
- classify the network based on the discussed taxonomy (e.g., as directed/undirected, weighted/unweighted, etc.);
- summarize the basic topological properties of the network, such as numbers of nodes and edges, network diameter, and average path length;
- identify the connected components and determine their size, establish the existence of a giant component, and plot the component size distribution;
- find the maximum and average node degrees, characterize and plot the degree distribution;
- use centrality measures (degree, betweenness, pagerank, etc.) to find the most important vertices of the network, and try to explain in which sense each of them is important;
- measure the level of transitivity of the network;
- partition the network in communities and characterize the community structure and their intra- and inter-connection patterns;

- compute the degree assortativity of the network and, if the vertices have scalar attributes, compute the assortativity of the network with respect to those attributes;
- use some random graph null model to measure the "randomness" of the network and to identify features of the network that are especially noteworthy.

Please, elaborate on the results of the analysis instead of just reporting numbers, figures and tables. If possible, hand-in your work as a jupyter notebook, using markdown cells for normal text. If you do not want to use jupyter, please present all the results in a single file, and provide the pieces of code used for the analysis.

2 How to choose the network

Choose networks that are not too small (not interesting to analyze) nor too large (computationally hard to analyze). A good choice might be a network having a few hundred thousand edges – but even a few tens of thousands might work. Here are some suggestions:

- a network of import/export relations between FAO countries;
- a network representing the Internet at the level of Autonomous Systems (ASs) in 2004;
- a network of regularly occurring flights among airports worldwide;
- a transportation network of Ile-de-France in 2016, jointly representing the road network, train network, and subway network;
- a network of collaborations between characters of the Marvel Universe;
- one of the these networks of collaborations between scientists, for different categories and time windows;
- a network of HIV transmissions between people through sexual, needle-sharing, or social connections;
- one of two networks of user friendships on Foursquare across 2012 and 2013.

Beware that some of these are multiplex networks, you might want to flatten/simplify them before the analysis.

3 Deadline

The deadline for submitting your work is September 30, 2022.

If you have any doubt or issue, do not hesitate to contact me at s.guarino[at]iac.cnr.it.