

Complex Networks - Project

Remy Cazabet

1 In short

1. Choose a real network
2. Use what you learned in the lectures to analyse it
3. Write a report illustrating your experiments and insights

2 Project details

2.1 Creating your own network

I request that you create your own network(s) to analyze. This means

- Collecting the network from external sources, e.g., using an API
- Starting from an existing dataset which is not in graph format, and create a graph version of it (e.g., transport network from data of taxi trips including source and destination, network of music similarity from a dataset of User/music listening, etc.)
- Start from an existing rich and large network dataset and clean it/enrich it to get a new view of it.

In all cases, your network should be interpretable, and your report must start by introducing it: What exactly do nodes represent? What exactly do edges represent? Describe precisely if some choices have been made when creating the graph from original data, and what do you think about those choices, what impact they might have on the network structure, etc.

2.2 Format

You can choose either to send a Jupyter Notebook or a PDF file. If a Notebook, it can make call to functions in external files, and must be easy to read. If a PDF, you should also join the code used in another document.

2.3 Choosing a network

You can choose any network dataset. On the website of the class, you can find references to repositories having collections of networks. You must either create your own network, or take a rich network in order to be able to extract interpretation from it.

Be careful that if your data does not contain node attributes (e.g., age, location, country...), you won't be able to interpret much. If the network is on a topic you have no idea about, it will also be hard to make sense of it.

2.3.1 How to build your network

If you're interested in a topic in particular, you might be able to easily create your own network. Basically, what you need is to end up with an edge list (that can be easily imported in networkx and any other tool). For instance, if you're interested in movies, you could download the MovieLens dataset <https://grouplens.org/datasets/movielens/latest/>, which collects rankings given by users to movies. From it, you can build a bipartite, weighted network (user-movie), and/or derive a movie similarity network (counting the number of similar votes by the same people ? Or something more clever). In previous years, students have also analyzed biological networks, road networks, and many more, coming either from freely available datasets or collected by them. Feel free to ask for help and advice.

2.4 Visualization

A nice graph drawing is worth 1000 words. You can include graph drawings to support the discussion. If the graph drawing comes from Gephi, you can include them in the notebook too, but do not forget to include the original image when you send your report. You can also learn to use the pyvis library <https://pyvis.readthedocs.io/en/latest/>

2.5 What should I do ?

Everything is fine. Try to tell a story. Show us what you have learned. If you have a real question you would like to answer on real data, it's the right occasion.

If you're particularly interested in a method rather than a dataset, you can also investigate more precisely this method (or set of methods) on several datasets. For instance, you might want to compare several methods for community detection on several datasets, or try to design your own network generator and show how good its properties are... The only limit is your imagination! (And time, unfortunately...)