

Complex Networks - Project

Remy Cazabet

2022/2023

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 Specificity by Master specialty

2.1 Students from the Complex Systems option

Since you have 8 hours dedicated to it, I request (Unless your project is based only on synthetic networks) that you pick an interpretable network, and that you include in your report a discussion on the modeling of the original data as a network: what exactly do node represents? What exactly do edge represents? 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 Students from Computer Science/Informatique Fondamentale

Since you have very little time for this project, consider it as a simple application of the computer practicals we do in class, on another network. Pick any network you want, and analyze it with the tools seen in class.

3 Project details

3.1 Format

We expect at least a jupyter notebook, containing code, results and discussions. If there is a need to develop some ideas or include external resources, a notebook might not be enough. Feel free to join a PDF document, and any another useful resource.

3.2 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. For Complex System students, you should either create your own network, or take a rich network in order to be able to extract interpretation from it. For Computer Science, you can pick directly a dataset from a repository and analyze it as seen in exercises.

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 out of it.

3.2.1 Create your own 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/>, that 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.

3.3 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. If you have problems using Gephi on your computer, you can also learn to use the pyvis library <https://pyvis.readthedocs.io/en/latest/>

3.4 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...)