L&S 88-4 Social networks
Final project

Overview

Over the course of the semester, we have used the labs to (1) explore some important concepts in the field of social networks; and (2) learn how to work with social network data using python. The homeworks have been intended to reinforce and extend the material we learned in the labs.

The final project gives you the opportunity to investigate a particular social network topic in greater depth. This will be helpful because you will gain experience in several important aspects of scientific data analysis that are hard to practice in the lab / homework format. In a lab, you are typically guided step by step through a data analysis. For this project, you will explore a social networks dataset on your own. This project also gives you the opportunity to develop an extended written argument that you support with your data analysis.

Important dates

- Wed, Nov 16th - hand in a **one-page proposal** (a description of your plan for the project). You should hand this in on Bcourses; if you are working with a partner, put both names on the proposal and each of you should hand a copy in
- Thursday, Nov 17 - Friday, Nov 18: meet with me to discuss your proposal
- Friday, Dec 2nd - Project is due; you must hand in
    - a **4-5 page writeup as a pdf**
    - **the .ipynb file you used**
    - a **pdf printout of the ipynb file** (like we did for the labs)

Datasets

I've put together a few datasets that you can use for this project. All three of these datsets are about networks and US politics. It's completely fine to just pick one of these datasets; I want there to be a few options so that we are all working on the same topic area, but not the exact same thing.

There are three networks datasets:

* information about high-dollar donors to PACs (political action committees)
* bill co-sponsorship network among members of congress
* who-follows-whom network on Twitter among members of congress

Additionally, I've included a few datasets that have some information on individual members of congress, which should be helpful for the second and third datasets. If you want to use this extra information, you'll have to do a bit of work to merge the datasets together. If you figure out

how to merge interesting stuff together, please share it here on Piazza so that everyone can all benefit.

You might also find other sources of information; for example, the legislators come from states / congressional districts, and you might be able to find information about the districts and states online.

The nature of this project is that it is less structured than a lab or homework. I hope that is exciting, but should make doing this project a bit more challenging. Piazza will be a valuable resource, both for asking questions and for sharing things that you and your partner have learned.

Here is the link to the repository that has the datasets and some ipynb files that demonstrate how to open them up:

https://data8.berkeley.edu/hub/interact?repo=social-networks-connector&path=project

UPDATE: new version of the Twitter network (please see Piazza post for more info):

https://data8.berkeley.edu/hub/interact?repo=social-networks-connector&path=project_update

You should create a different directory to hold your project datafiles and analysis just in case this repository needs to get updated at some point. In other words, look at this repository and the example notebooks it contains to get a feel for the different datasets. Then copy the actual datasets you will use into a separate directory.

Goals

Here are some of the objectives I would like you to think about while you work on your project. (Some of these categories overlap with one another.)

Social networks objectives

- Demonstrate an understanding of what the network you are studying is (nodes, edges, directed/undirected, et cetera)
- Demonstrate a familiarity with some of the key concepts we have discussed this semester (e.g. homophily, connectivity metrics, and so forth)
- Demonstrate an understanding of where the data about the social network came from, and how that might affect your analysis

Communications objectives

- Write about your analysis in a way that could be understood by an intelligent reader who had not necessarily taken Data 8

- Briefly summarise general features of the dataset (e.g. for a survey dataset, you would want to summarise who the people who responded to the survey are)
- Use informative plots with sensible axes, labels, et cetera
- Explain the limitations of your analysis
- Describe next steps that could be taken in the future to expand on your analysis (you might describe another analysis, more or different data collection, or other things)

Data analysis and programming objectives

- Open a not-completely-clean dataset and prepare it for analysis
- Explore a dataset by describing it using plots and quantitative summaries (e.g. tables, means, etc)
- Generate a prediction from your exploratory analysis
- Test the prediction using a techniques you learned in Data 8 and in our class
- Describe how to interpret the results of your test (including the assumptions it makes)