Final Project Overview W24 - SI 507

Milestones

There are 3 milestones that need to be turned in.

- Project Proposal
- Data Checkpoint & Interactive Presentation Design
- Final Project Demo and Repository Link Submission

Project Overview

The goal of the final project is for you to showcase the different things you've learned in 507 in a project you can talk about during an interview or internship. You will need to pick an interesting topic, find data related to that topic, use a graph data structure to analyze that data, and build a program that allows the user to interrogate that dataset/examine your analysis.

In the second half of this course we will introduce several ways to do basic analysis of graph data structures. We will develop methods to determine the most highly connected points in a graph, identify shortest distance between different data points, and identify which data points are the most related.

You can use any or all of these methods to develop an interesting program that tells you something interesting about the topic. You don't have to 'discover' anything novel, simply highlighting obvious relationships is fine.

Your project should be of the appropriate complexity for a final project. You will get feedback on this based on your proposal. You can chose to add complexity to a topic either through connecting multiple datasets or through adding modes of interactives.

Here are a couple of examples that would be reasonable final projects:

- A program, modeled after Project 3 Kevin Bacon, that examined teammate relationships in the NBA over time. So it accessed all the rosters from NBA teams for every year on basketball-reference.com, and then it allowed the user to enter 2 players (for example, Kobe Bryant and Micheal Jordan) and then identified the

chain of teammates that would connect them. It also allowed the user to identify the most connected players in the nba in each season, and then allowed the user to launch in a browser the basketball-reference page for each player. This project used .csv files directly from a website, so the data access was not difficult, bt this program compensates by having multiple layers of interactivity.
- A program that built a network based on which artists existed in spotify playlists together most often. The user was able to enter an artist, and get recommendations for similar artists based on which other artists appeared with them in playlists more often. The user was also able to request a playlist based on certain artists or themes.
- A program that used yelp data for a given city to identify the most broadly popular restaurants, and the 'cult classic' restaurants (i.e. highly popular with a narrow set of users) within a given cuisine. This project has less interactivity options than the other examples, but the data access was harder (an API we didnt teach in class), so the complexity was still high.

The best projects will integrate two related but distinct dataset and therefore be able to tell us something interesting that couldn't be found with just one

Upcoming Milestones, in Brief

- Proposal
  - ~ 0.5 to 1 page, describing your data sources and how you plan to use them (processing, interaction, and presentation).
  - You will receive feedback on this proposal until it is approved.
- Data Checkpoint & Interactive Presentation Design
  - Data Checkpoint: Demonstrate that you are successfully collecting, caching, and storing in a json or html 'cache' all relevant data from your sources.
  - Interactive Presentation Design: ~ 1 page, describing your plans for implementing interactive presentation capabilities, including user options supported and presentation types
  - You will only receive feedback on this checkpoint if something is wrong!
- Demo and final submission
  - A 2-4 minute video demonstrating your project
  - a 1 page report on your programs data and features, and any interesting highlights.
  - the code of your program and underlying data.