

CS2223: Algorithms | Assignment 4

General Instructions

- Code vs. Pseudocode: Each question will state whether the deliverable is pseudocode or an executable program and code that the TA will run to give you a grade.
- You should make clear in your report:
 - How to compile your program
 - How to execute it, along with any arguments
- Submissions: Submit your work as a .Zip file through Canvas. All programs plus your written questions (which should be PDF files to ensure we can see them) should be zipped into a single archive when submitted. Please organize folders by question, e.g. Q1, Q2, etc.

Q1: Graph Explorations (100 pts)

In this project, your goal is to learn to load, process, and perform basic path- and search- based operations on multivariate graphs. The term “multivariate” may sound intimidating, but it’s just a fancy way of saying that the nodes and edges have multiple attributes associated with them. What’s cool about multivariate graph data is that you can use the algorithms we’re learning to create various subgraphs from such a dataset, to answer a range of interesting questions.

[Here’s the data](#) you’ll use. You may see some familiar names in it.

Task Set 1 (Constructing the Graph)

- 1) Load the JSON data in Java
 - a) JSON data isn’t necessarily something you’ve worked with yet, but it’s ****everywhere****, and there are tons of code examples and libraries for working with it.
- 2) Write functions to iterate and parse through the nodes and links, to build an Adjacency List data structure for your graph (~524) using the Symbol Graph technique (~548, 551)
 - a) You may treat mentions and retweets as the same.

Next, you are to perform some basic BFS and DFS operations on the data:

Task Set 2 (Basic Graph Functionality)

- 1) Implement functionality (use the book) to find neighbors, given a source node.
 - a) Use it to answer the following question. “Find all of giCentre’s North American neighbors.”
 - b) Note that you will have to find the neighbors, then process the results further to find those that are North American.
- 2) Implement functionality (again, use the book) to find shortest paths, given a source node.
 - a) Use it to answer the question, “What is the institution on a shortest path between Lane and Rob?”
 - b) Hint: use Lane as a source in your path search code¹.

Task 3 (Exploratory Analysis)

Finally, given the many things that could be found in this dataset, perform an exploratory analysis with your existing code.

Implement new functions if needed, and produce a brief writeup (PDF) of what you found and what code (if any) you had to write to do it.

¹ Note: it’s weird using my own name in a homework question, but I can assure you that I did not write this question originally!

For anyone looking for a challenge, consider the large range of possibilities in this dataset, like identifying the degrees of separation between a popular user, e.g. jeffrey_heer or tamaramunzner, and other users in the dataset (~552), or deriving directed / weighted graphs as part of a more complex analysis question.

Output:

The output of this should include Java code TAs can inspect and run, as well as a PDF report that describes your exploratory analysis.