Karl M. Heuer / David E. Roberson

Autumn 2024
Exercise class:
in 302 on September 17th

Exercises for Computational Tools for Data Science (02807)

Week 3: MapReduce

References and Reading

1. Chapter 2 of Mining of Massive Data Sets, Jure Leskovec, Anand Rajaraman, and Jeff Ullman.

Exercise 1: Learn how to use map and reduce functions

Take a look at the python tutorial for `map` and `reduce` functions here, and import the `functools` module so you can use the `reduce` function.

Exercise 2: Word Frequency

Implement the word frequency example discussed in class, i.e., the input is a document of words and the output is the frequency of each word. Test your solution on a small example.

Exercise 3: Inverted index

Implement the inverted index example discussed in class, i.e., the input is a collection of documents and the output is a set of <key, value> pairs where each key is a word appearing in at least one document and the value is the list of documents it appears in. Test your solution on a small example.

Exercise 4: Euler Tour

Determine if a graph has an Euler tour. To do so count and output the number of vertices of even and odd degree. The input is a file representing a graph G, where each line consists of two numbers x and y representing an edge (x, y) in G. The output should be a count of the number of nodes with even degree and odd degree. Test your solution on the graphs given in the files `eulerGraphx.txt`, where $x = 1, 2, 3$.

Exercise 5: Common Friends

Implement the common friends example discussed in class. The input is a file representing a graph in an adjacency list style-format. Each line in the file is of the form $x : y_1, y_2, \ldots, y_k$ and encodes that vertex x is adjacent to vertices $y_1, y_2, \ldots, y_k$. The output should be pairs of ADJACENT vertices and their common neighbors, i.e., $x, y : c_1, c_2, \ldots, c_j$ if x and y have common neighbors $c_1, \ldots, c_j$. Test your solution on the graph in the file `friends.txt`.

Exercise 6: Triangle Counting

Compute the number of triangles in a graph. The input is in the same format as the Euler Tour exercise. Test your solution on the graph in the file `roadnet.txt`. *Hint:* The solution to the common friends exercise may be useful here.

Exercise 7: Install and explore NetworkX

With any remaining time after the above exercises, install the `NetwowkX` package and explore its capabilities (see `https://networkx.org`). Can you use it to do any of the above exercises? Try making and/or drawing some interesting graphs.