

Weekplan: Massively Parallel Computation

02807 Computational Tools for Data Science

References and Reading

- [1] Chap. 3 of Mining of Massive Data Sets, Jure Leskovec, Anand Rajaraman, and Jeff Ullman. Chapter 3.
- [2] Documentation for mrjob. See mrjob.readthedocs.io/en/latest/

Exercises

- 1 [w] **Install mrjob** Search relevant documentation and install mrjob on your system.
- 2 [w] **Word Frequency** Implement the word frequency example discussed in class. Test your solution on a small example.
- 3 **Inverted Index** Implement the inverted index example discussed in class. Test your solution on a small example.
- 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 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 $x = 1, 2, 3$.
- 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 $x : y_1, y_2, \dots, y_k$ and encodes that vertex x is adjacent to vertices y_1, y_2, \dots, y_k . The output should be a lines with pairs and their common friends, i.e., $x, y : c_1, c_2, \dots, c_j$ if x and y have common friends c_1, \dots, c_j . Test your solution on the graph in the file `friends.txt`.
- 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.