

Portfolio-Exam – Standalone Task

This document is part of the the description of the portfolio-exam of the Data Science course MADS-EMDM (Advanced Topics of Data Mining). Read the document `portfolio.pdf` before you continue here. In this document, you will find the last task of the exam, Task 9.

Task 9 – Time Measurements

Measure and describe the runtime of the NetworkX implementations of the two community clustering algorithms “Girvan Newman algorithm” and “Louvain algorithm”.

For that purpose, sample multiple random subgraphs of different order from a given graph and evaluate the runtime of the two algorithms on them. In particular:

1. Use the graph dataset you chose for the other tasks of the portfolio exam.
2. For the experiments, choose ten different subgraph orders, distributed reasonably – given the order of your original graph.
3. For each order, sample a reasonable number of different induced subgraphs by choosing nodes uniformly at random from the node set.
4. Time the execution of each clustering algorithm on each subgraph.
5. Try to obtain reliable measurements.
6. Choose one or more helpful visualization(s) that shows the behavior of the runtime(s) over the order of the clustered graph.
7. Interpret the results and critically reflect on the experimental setup (including your choices and the choices in the task).

Hints:

- Make sure that the Girvan Newman algorithm is actually executed.
- If the runtime of Girvan Newman is too long, limit the number of actually computed levels in the clustering hierarchy.
- Use the same subgraphs for both algorithms.
- Choose the number of repetitions (different subgraphs per order) reasonably per algorithm.