FACHHOCHSCHULE KIEL
University of Applied Sciences

Advanced Topics of Data Mining **Prof. Dr. Stephan Doerfel**

MADS-EMDM WiSe 2023

Portfolio-Exam – Graph 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. Here, you will find Tasks 1–7 of the portfolio exam.
The following tasks together form a set of experiments to explore a graph dataset.

Read the following 3 pages carefully before you start working on the experiments and before your final decision about the dataset!

Task 1 – Story

In the first part, explain the story and circumstances of the experiments.

- Imagine and describe a fictitious situation or describe a real scenario in which you are a data scientist working for some organization.

- Explain the context – project or task – for your experiment.

- Explain the plan of the experiment and the purpose – the value you expect to create for the organization.

Task 2 – The Data

In the second part, load and present a dataset in which the data is represented in a graph structure.

- Explain the dataset itself (e.g., what do the nodes and edges represent?).

- Explain how the dataset is suitable for the project from Task 1.

- Condition 1: Use real-life data. Do NOT use artificial datasets NOR datasets from the lecture!

- Condition 2: There are two possible availability scenarios for the data: A) It is available online with proper license. In that case indicate in your notebook where the data can be downloaded. B) You (legally!) obtain a dataset and share it with me via Moodle. Such data should not come with any form of NDA or other obligations. If you are not sure about these criteria regarding the dataset you consider, please contact me before you invest too much time in the experiments. If you write your own code to acquire data (e.g. querying an API), create a **separate notebook** for that purpose only and submit everything in a zip file.

Task 3 – IDA

Conduct an initial data analysis.

- Present some distributions and statistical properties that inform the reader about the dataset or that are relevant for your project.

- Find one or more helpful visualizations of the data.

Task 4 – Graph Properties

Bring the dataset into the form that you need for the experiments.

- Compute and explore various graph properties and interpret the results.

Task 5 – Central Nodes

Determine central nodes.

- Compute several useful centrality measures.

- Compare nodes according to different measures.

- Draw conclusions about their relevance/importance/position/role in the network (depending on what fits to the problem described in your story.

Task 6 – Community Discovery

Compute community clusterings of the graph.

- Run at least two community discovery algorithms.

- Compare and interpret results.

- Investigate the community membership of the central nodes (Task 5).

Task 7 – Conclusions and Future Work

- Summarize and interpret the achieved results.

- Compare to the original plan (Task 1).

- Explain the generated value.

- Recommend a course of action for the organization in your story based on the results.

- Propose ideas for future work (a short sketch or enumeration of ideas is sufficient, no further experiments).

Note: One further task of the portfolio will be published at the end of the term. This will, however, be a standalone task, unconnected to the experiments here.