Linnaeus University / 4DV510

Project Requirements 3 and 4:

Networks

Introduction

Networks are datasets where the edges between nodes (or the explicit links between data points) dictate the structure of the data. The analysis of networks can involve many different tasks, such as the search for groups of similar nodes (or "communities"), the extraction of quantitative measures for nodes and edges, or the investigation of overall trends in the organization of the network as a whole.

Requirements

You should proceed with the implementation of your data mining framework by focusing on networks, community detection, and quantitative measures for nodes and edges (such as PageRank for nodes and Betweenness for edges).

1. **Create a component for a *Network*.** It should be able to read a network from a file (such as an edge list or an adjacency matrix), and should expose methods to get (and, by consequence, iterate over) the nodes and edges of the network.
2. **Create a component for *Community Detection*.** It should accept as input a *Network* and optional hyper-parameters. The hyper-parameters must not be hard-coded; every different technique will have different ones. The output can be a list of lists, for example, or a single list with the community label for each node.
3. **Create a component for a *Node Measure*.** It should take as input the network and generate as output a list of the values of each node.
4. **Create a component for an *Edge Measure*.** It should take as input the network and generate as output a list of the values of each edge.

Notes:

Remember to write also at least three different implementations for each component. In order to test your new requirements, make sure you can run a pipeline with them where you programmatically test different combinations of each component.