Data Mining

Winter Term 2020/2021

Overview – Programming Assignment

Lecture: Lukas Miklautz, Nils Kriege, Sebastian Tschiatschek, Katerina Schindlerova
Tutor: Johannes Stangl

Overview

Two connected subprojects – 3 deadlines

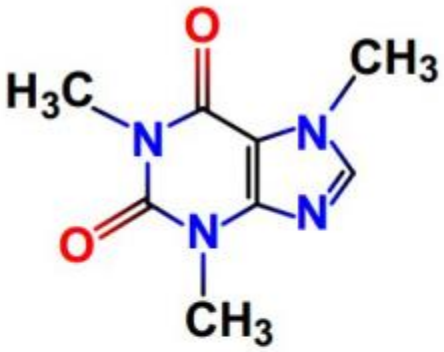
1. Part 1: Exploratory Data Analysis – 29.11.2020, 11:59 p.m.
2. Part 2: Graph Property Prediciton – 24.01.2021, 11:59 p.m.
3. Peer Review Phase – 07.02.2021, 11:59 p.m.

Groups of 4 – 5 students

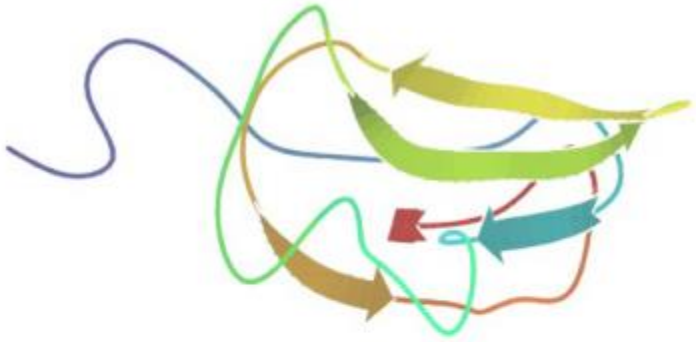
- Form a team of 4 – 5 students until 29.10.2020, 11:59 p.m.
- Hand in zip archive with naming scheme **team_X.zip**
  - Code
  - PDF or HTML report
  - 10 minute video for each subproject

Review/Q&A sessions for your questions regarding the Assignment

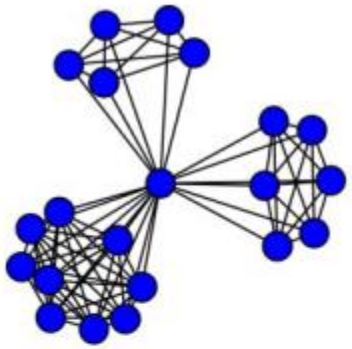
Data Sets



(a) Small molecule



(b) Enzyme



(c) IMDB ego-network

(d) Semantic image processing

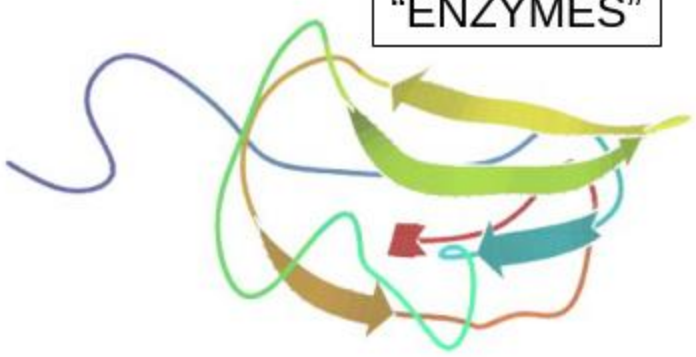
Figure 1: Examples of the graph from different domains.[2]

Data Sets

(a) Small molecule

"NCI1"

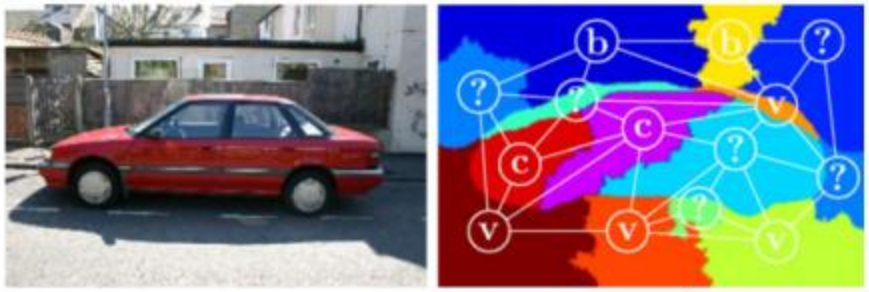
"ENZYMES"



(b) Enzyme

(c) IMDB ego-network

"IMDB-BINARY"

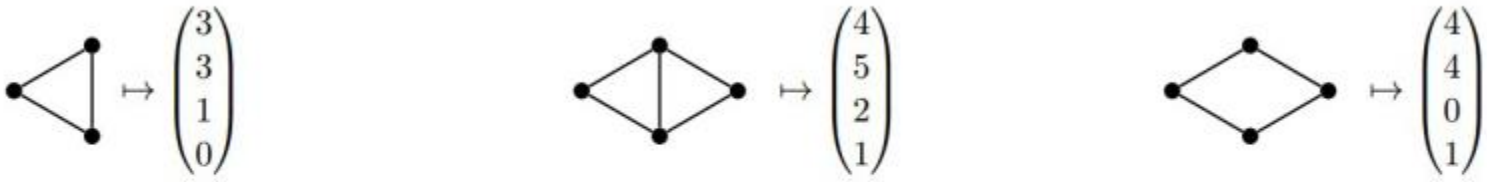


(d) Semantic image processing

"MSRC_21"

Figure 1: Examples of the graph from different domains.[2]

Representing Graphs as Vectors



(a) F (b) G (c) H

Figure 2: Examples of a simple vector representation for the graphs F, G and H, where the components of the vector are the number of vertices, edges, triangles and cycles of length four, respectively.

- We will use:
  - Graphlet
  - Weisfeiler-Lehman
  - Shortest Path

Task – 1 Exploratory Data Analysis

You choose the dataset, we give you the graph kernels. Then:

- Choose an algorithm from the provided list (or beyond)

  Deadline: 03.11.2020, 11:59 p.m.
- Implement and validate the algorithm by comparing with ELKI. If you chose an algorithm from beyond the list, compare with results of the original paper.
- Perform an exploratory data analysis of your chosen data set
- Apply your algorithm to the dataset you have chosen and evaluate your results

Task – 2 Graph Property Prediction

You will use the same data set as in the 1st task. Then:

- Make yourself familiar with evaluation methods for classification and, in particular, the cross validation based approach.
- Perform classification experiments with the baseline graph kernels including those used for clustering in Task 1. Try to relate your findings to your results for the clustering.
- Design your own graph kernel or graph neural network. Kernels can be easily combined and all kinds of graph and node properties can be used.
- Experimentally evaluate your graph kernel. Try to improve your graph kernel based on your findings

Questions?

Main Contacts:
Johannes Stangl: johannes.stangl@univie.ac.at
Lukas Miklautz: lukas.miklautz@univie.ac.at