Pattern Recognition
Molecular Exercise

Tobias Verheijen, Mattéo Bonvin, Zhihui Wang

20 May, 2024

1 Molecular Exercise

For this exercise, we chose to complete the Molecular/graph-based problem. We used a relatively straight forward set of cost functions: node insersion/deletion = 1, node substitution = 2 if not equal, 0 otherwise, edge insersion/delection = 1, edge substitution = —difference in valence—.

We then used the GED as the basis for our KNN classifier.

With a base-line of $k = 3$, we achieve an accuracy of **0.972**. With $k = 5$, we achieve an accuracy of **0.956**.

2 Competition

For this competition, we tried multiple K but decided to use a prediction with $K = 5$ to create the .tsv file.

We took the $read_all_graph.py$ and copied it into a updated version fot this part. As we do not need validation no more, instead of keeping it away, we added it to the training set to make it bigger. The results are in the $test_predictions_K5.tsv$ file