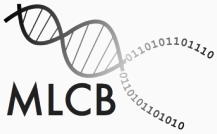
Kernels for proteins and MMD.

Philip Hartout

February 28, 2022



D BSSE

ETH zürich

Introduction

- Here we want to go through the maths behind the kernels to make sure our implementation is efficient.
- Linear, Weissfeiler-Lehmann kernel

Linear kernel

- Not much to optimize there. Just check: this link.

W-L kernel

- Acutally, W-L K. is a generalized version of the bag of node degrees since node degrees are one-hop neighborhood node enrichment while W-L iteratively enrich node voc. by doing n-hops.
- Steps: take input graph, (1) iteratively compute hash value of each node based on their neighborhood, (2) create feature map with counts of nodes with given color at each iteration of the hash value computation
- Notes: complexity is linear in # edges. only the colors appearing in two graphs needs to be tracked. Counting colors can be done in linear time in # nodes. In total, $\mathcal{O}(e)$

References i