INFO-F438 - Algorithms in computational biology Project Proposal
Heuristic Algorithms for the Maximum Colorful Subtree Problem

Prateeba Ruggoo

May 5, 2019

1 The subject

Tandem MS is a technique used to break down selected ions (precursor ions) into fragments (product ions). The fragments then reveal aspects of the chemical structure of the precursor ion. This computational task can be formulated as the Maximum Colorful Subtree problem.

2 Project

The project will be done in the form of an explanatory page with jupyter notebook.

The main parts will be the following :

1. The first part will consist of an introduction to the problem.

2. The second part will contain the proof of $\mathcal{NP}-$ hardness of the Maximum Colorful Subtree Problem, to explain the reason for implementing heuristics and not exact methods.

3. The third part will consist of the pseudo codes and implementation of the different heuristics used to compute the Maximum Colorful Subtree problem. The following heuristics will be implemented :

   (a) Kruskal-style Heuristics
   (b) Prim-style Heuristics
   (c) Insertion technique Heuristics
   (d) Top-down approach Heuristics
   (e) Critical Path approach Heuristics

4. The complexity of the different heuristics will be analysed in the fourth part.

5. The last section will be dedicated for the results, comparisons and conclusion.

References

[1] Imran Rauf, Florian Rasche, Francois Nicolas, and Sebastian Bocker. Finding Maximum Colorful Subtrees in practice. page 13.

[2] S. Bocker and F. Rasche. Towards de novo identification of metabolites by analyzing tandem mass spectra. *Bioinformatics*, 24(16):i49–i55, August 2008.

[3] Kai Dührkop, Marie A. Lataretu, W. Timothy J. White, and Sebastian Böcker. Heuristic Algorithms for the Maximum Colorful Subtree Problem. page 14 pages, 2018.