Fragmentation Trees from MS^n Data

Franziska Hufsky[1,2], Kerstin Scheubert[1], Florian Rasche[1], and Sebastian Böcker[1]

[1] Lehrstuhl für Bioinformatik, Friedrich-Schiller-Universität Jena, Ernst-Abbe-Platz 2, Jena, Germany,
{franziska.hufsky,kerstin.scheubert,florian.rasche,sebastian.boecker}@uni-jena.de
[2] Max Planck Institute for Chemical Ecology, Beutenberg Campus, Jena, Germany

Since metabolites cannot be predicted from the genome sequence, high-throughput *de-novo* identification of metabolites is highly sought. Mass spectrometry (MS) in combination with a fragmentation technique is commonly used for this task. Unfortunately, automated analysis of such data is in its infancy. Recently, fragmentation trees have been proposed as an analysis tool for such data. Additional fragmentation steps (MS^n) reveal more information about the molecule.

We propose to use MS^n data for the computation of fragmentation trees and adjust the fragmentation model for MS^2 to reflect the succession of fragmentation reactions. We present the Colorful Subtree Closure problem to formalize this task: There, we search for a colorful subtree inside a vertex-colored graph, such that the weight of the transitive closure of the subtree is maximal. We give several negative results regarding the tractability and approximability of this problem and present an exact dynamic programming algorithm, which is parameterized by the number of colors in the graph and is swift in practice.

Evaluation of our method on a dataset of 45 reference compounds showed, that almost one quarter of all fragments are changed due to the information from MS^n data. As our scoring scheme is "chemically reasonable", we argue that the trees are actually improved.