**Computational methods for small molecules**

**Type of talk**

PhD seminar

**Speaker**

Markus Heinonen, IBISC, Arob@s Team, Evry, France

**Abstract**

Metabolism is the system of chemical reactions sustaining life in the cells of living organisms. It is responsible for cellular processes that break down nutrients for energy and produce building blocks for necessary molecules. The study of metabolism is vital to many disciplines in medicine and pharmacy. Chemical reactions operate on small molecules called metabolites, which form the core of metabolism. In this thesis we propose efficient computational methods for small molecules in metabolic applications. In this thesis we discuss four distinctive studies covering two major themes: the atom-level description of biochemical reactions, and analysis of tandem mass spectrometric measurements of metabolites.

In the first part we study atom-level descriptions of organic reactions. We begin by proposing an optimal algorithm for determining the atom-to-atom correspondences between the reactant and product metabolites of organic reactions. In addition, we introduce a graph edit distance based cost as the mathematical formalism to determine optimality of atom mappings. We continue by proposing a compact single-graph representation of reactions using the atom mappings. We investigate the utility of the new representation in a reaction function classification task, where a descriptive category of the reaction's function is predicted. To facilitate the prediction, we introduce the first feasible path-based graph kernel, which describes the reactions as path sequences to high classification accuracy.

In the second part we turn our focus on analysing tandem mass spectrometric measurements of metabolites. In a tandem mass spectrometer, an input molecule structure is fragmented into substructures or fragments, whose masses are observed. We begin by studying the fragment identification problem. A combinatorial algorithm is presented to enumerate candidate substructures based on the given masses. We also demonstrate the usefulness of utilising approximated bond energies as a cost function to rank the candidate structures according to their chemical feasibility. We propose fragmentation tree models to describe the dependencies between fragments for higher identification accuracy.

We continue by studying a closely related problem where an unknown metabolite is elucidated based on its tandem mass spectrometric fragment signals. This metabolite identification task is an important problem in metabolomics, underpinning the subsequent modelling and analysis efforts. We propose an automatic machine learning framework to predict a set of structural properties of the unknown metabolite. The properties are turned into candidate structures by a

novel statistical model. We introduce the first mass spectral kernels and explore three feature classes to facilitate the prediction. The kernels introduce support for high-accuracy mass spectrometric measurements for enhanced predictive accuracy.

**Reference**

Ph.D. Thesis: Computational methods for small molecules, University of Helsinki, Department of Computer Science, 2012