Abstract

A common task in natural sciences is to describe, characterize, and infer relations between discrete objects. A set of relations E on a set of objects V can naturally be expressed as a graph $G = (V, E)$. It is therefore often convenient to formalize problems in natural sciences as graph theoretical problems. In this thesis we will examine a number of problems found in life sciences in particular, and show how to use graph theoretical concepts to formalize and solve the presented problems. The content of the thesis is a collection of papers all solving separate problems that are relevant to biology or biochemistry.

The first paper examines problems found in self-assembling protein design. Designing polypeptides, composed of concatenated coiled coil units, to fold into polyhedra turns out to be intimately related to the concept of 1-face embeddings in graph topology. We show that 1-face embeddings can be canonicalized in linear time and present algorithms to enumerate pairwise non-isomorphic 1-face embeddings in orientable surfaces.

The second and third paper examine problems found in evolutionary biology. In particular, they focus on inferring gene and species trees directly from sequence data without any a priori knowledge of the trees topology. The second paper characterize when gene trees can be inferred from estimates of orthology, paralogy and xenology relations when only partial information is available. Using this characterization an algorithm is presented that constructs a gene tree consistent with the estimates in polynomial time, if one exists. The shown algorithm is used to experimentally show that gene trees can be accurately inferred even in the case that only 20% of the relations are known. The third paper explores how to reconcile a gene tree with a species tree in a biologically feasible way, when the events of the gene tree are known. Biologically feasible reconciliations are characterized using only the topology of the gene and species tree. Using this characterization an algorithm is shown that constructs a biologically feasible reconciliation in polynomial time, if one exists.

The fourth and fifth paper are concerned with with the analysis of automatically generated reaction networks. The fourth paper introduces an algorithm to predict thermodynamic properties of compounds in a chemistry. The algorithm is based on the well known group contribution methods and will automatically infer functional groups based on common structural motifs found in a set of sampled compounds. It is shown experimentally that the algorithm can be used to accurately predict a variety of molecular properties such as normal boiling point, Gibbs free energy, and the minimum free energy of RNA secondary structures. The fifth and final paper presents a framework to track atoms through reaction networks generated by a graph grammar. Using concepts found in semigroup theory, the paper defines the characteristic monoid

of a reaction network. It goes on to show how natural subsystems of a reaction network organically emerge from the right Cayley graph of said monoid. The applicability of the framework is proven by applying it to the design of isotopic labeling experiments as well as to the analysis of the TCA cycle.