

Abstract: Subgraph Covers-An information theoretic approach to motif analysis in networks.

A large number of complex systems can be modeled as networks of interacting units. From a mathematical point of view the topology of such systems can be represented as graphs of which the nodes represent individual elements of the system and the edges interactions or relations between them. In recent years networks have become a principal tool for analyzing complex systems in many different fields.

This thesis introduces an information theoretic approach for finding characteristic connectivity patterns of networks, also called network motifs. Network motifs are sometimes also referred to as basic building blocks of complex networks. Many real world networks contain a statistically surprising number of certain subgraph patterns called network motifs. In biological and technological networks motifs are thought to contribute to the overall function of the network by performing modular tasks such as information processing. Therefore, methods for identifying network motifs are of great scientific interest.

In the prevalent approach to motif analysis network motifs are defined to be subgraphs that occur significantly more often in a network when compared to a null model that preserves certain features of the network. However, defining appropriate null models and sampling these has proven to be challenging. This thesis introduces an alternative approach to motif analysis which looks at motifs as regularities of a network that can be exploited to obtain a more efficient representation of the network. The approach is based on finding a subgraph cover that represents the network using minimal total information. Here, a subgraph cover is a set of subgraphs such that every edge of the graph is contained in at least one subgraph in the cover while the total information of a subgraph cover is the information required to specify the connectivity patterns occurring in the cover together with their position in the graph.

The thesis also studies the connection between motif analysis and random graph models for networks. Developing random graph models that incorporate high densities of triangles and other motifs has long been a goal of network research. In recent years, two such model have been proposed. However, their applications have remained limited because of the lack of a method for fitting such models to networks. In this thesis, we address this problem by showing that these models can be formulated as ensembles of subgraph covers and that the total information optimal subgraph covers can be used to match networks with such models. Moreover, these models can be solved analytically for many of their properties allowing for more accurate modeling of networks in general.

Finally, the thesis also analyzes the problem of finding a total information optimal subgraph cover with respect to its computational complexity. The problem turns out to be NP-hard hence, we propose a greedy heuristic for it. Empirical results for several real world networks from different fields are presented. In order to test the presented algorithm we also consider some synthetic networks with predetermined motif structure.