Description for general public

Entropy is commonly used in science to measure the degree of disorder. The higher the entropy of a given system, the more disorganized the system is. In other words, high entropy represents a high degree of uncertainty about the state of the system. Take the example of a box full of white and black balls. If there are an equal number of white and black balls, the uncertainty about the color of the randomly chosen ball is the highest, which means that the entropy of the box is maximal. On the other hand, if the box contains balls of only one color, there is no uncertainty involved and the entropy of the box equals 0. This simple analogy can be applied to any complex system, within which we want to apply a learning process.

In the world of complex networks algorithms oftentimes are faced with the need to estimate various properties of vertices. Strongly regular networks have low entropy, which does not necessarily imply the uniform distribution of vertex properties, such as degree, betweenness or closeness. It is possible that the network contains a few groups of vertices, and within each group the variance of properties is low, whereas there are significant intra-group differences. Knowing the entropy of centrality measures of vertices can be very useful for network processing algorithms, because it allows to direct the algorithm along the gradients of increasing or decreasing entropy, depending on the task at hand. It is worth mentioning however, that there are many conflicting definitions of network entropy proposed in the scientific literature, and these definitions are often incompatible, while trying to grasp different aspects of network topology. In addition, many of these definitions require computations which are simply infeasible with regard to sizes of contemporary complex networks (e.g. social networks) and render these definitions practically unusable.

The goal of this project is to examine the properties of entropies in the world of complex networks, with a special emphasis on signed networks, multiplex networks, and multimodal networks. In signed networks each edge has a valence (most often positive vs. negative), which represents the semantics of the relationship depicted as the edge, such as preference vs. dislike. Multiplex network contain edges that belong to different classes, for instance one type of edges represents the flow of e-mail messages while the other type of edges represents organizational hierarchy relations). Finally, multimodal networks contain vertices that belong to disjoint classes, such as companies and persons with their relationships). The main goal of the project is to develop intuitive, computationally feasible entropy definitions and extending these definitions to new classes of complex networks, as well as proving the utility of the proposed definitions by incorporating them in selected network processing algorithms.

In the first project phase we will adapt different entropy definitions to the world of complex networks and we will simplify the definitions to make them computationally feasible with respect to sizes of contemporary complex networks. In the second project phase we will operationalize definitions introduced in the first phase by introducing entropy-aware network processing algorithms. We have selected three classes of algorithms: network classification (assigning a given network to a general network topology class), network modularization (finding groups of vertices which form coherent and dense structures within the network), and network compression (by merging vertices and edges, or by sampling vertices and edges). This way we hope to prove the universal usefulness of entropy in complex network processing.