

Keynote Abstract

Subspace Search for Community Detection and Community Outlier Mining in Attributed Graphs

by Emmanuel Müller, Head of Knowledge Discovery and Data Mining Group, Hasso-Plattner-Institute, Germany

Attributed graphs are widely used for the representation of social networks, gene and protein interactions, communication networks, or product co-purchase in web stores. Each object is represented by its relationships to other objects (edge structure) and its individual properties (node attributes). For instance, social networks store friendship relations as edges and age, income, and other properties as attributes. These relationships and properties seem to be dependent on each other and exploiting these dependencies is beneficial, e.g. for community detection and community outlier mining. However, state-of-the-art techniques highly rely on this dependency assumption. In particular, community outlier mining is able to detect an outlier node if and only if connected nodes have similar values in all attributes. Such assumptions are generally known as homophily and are widely used. However, looking at multivariate spaces, one can observe that not all given attributes have high dependencies with the graph structure. For example, social properties such as income or age have strong dependencies with the graph structure of social networks. In contrast, properties such as gender are rather independent from it. Consequently, recent graph mining algorithms degenerate for multivariate attribute spaces that lack dependency with the graph structure in some of the attributes.

This talk introduces novel methods that select relevant subspaces, i.e. subsets of the attributes, showing dependencies with the graph. All methods are designed as general pre-processing steps to traditional graph mining and widely applicable. As first method, we propose a statistical selection of congruent subspaces, i.e. subsets of attributes showing a dependency with the graph structure. A core challenge in selecting these subspaces lies in the modeling of dependence between graph structure and attribute values. As second method, we incorporate the user preference into the selection of relevant subspaces in attributed graphs. We consider communities and community outliers based on user preference. We show that the selection of congruent subspaces clearly enhances outlier detection by measuring outlierness scores in selected subspaces only. Furthermore, focused attributes enable a more user-oriented mining of community structures. Experiments show that both approaches outperform traditional full space approaches and as general pre-processing steps they enhance the later data mining steps on attributed graphs.