

# Preface

Due to the ability of graphs to represent properties of entities and binary relations at the same time, a growing interest in graph based object representation can be observed in science and engineering. Yet, graphs are still not the common data structure in pattern recognition and related fields. The reason for this is twofold. First, working with graphs is unequally more challenging than working with feature vectors, as even basic mathematic operations cannot be defined in a standard way for graphs. Second, we observe a significant increase of the complexity of many algorithms when graphs rather than feature vectors are employed. In conclusion, almost none of the standard methods for pattern recognition can be applied to graphs without significant modifications and thus we observe a severe lack of graph based pattern recognition tools.

This thesis is concerned with a fundamentally novel approach to graph based pattern recognition based on vector space embeddings of graphs. We aim at condensing the high representational power of graphs into a computationally efficient and mathematically convenient feature vector. Based on the explicit embedding of graphs, the considered pattern recognition task is eventually carried out. Hence, the whole arsenal of algorithmic tools readily available for vectorial data can be applied to graphs. The key idea of our embedding framework is to regard dissimilarities of an input graph to some prototypical graphs as vectorial description of the graph. Obviously, by means of such an embedding we obtain a vector space where each axis is associated with a prototype graph and the coordinate values of an embedded graph are the distances of this graph to the individual prototypes.

Our graph embedding framework crucially relies on the computation of graph dissimilarities. Despite adverse mathematical and computational conditions in the graph domain, various procedures for evaluating the dis-

similarity of graphs have been proposed in the literature. In the present thesis the concept of graph edit distance is actually used for this task. Basically, the edit distance of graphs aims at deriving a dissimilarity measure from the number as well as the strength of distortions one has to apply to one graph in order to transform it into another graph. As it turns out, graph edit distance meets the requirements of applicability to a wide range of graphs as well as the adaptiveness to various problem domains. Due to this flexibility, the proposed embedding procedure can be applied to virtually any kind of graphs.

In our embedding framework, the selection of the prototypes is a critical issue since not only the prototypes themselves but also their number affect the resulting graph mapping, and thus the performance of the corresponding pattern recognition algorithm. In the present thesis an adequate selection of prototypes is addressed by various procedures, such as prototype selection methods, feature selection algorithms, ensemble methods, and several other approaches.

In an experimental evaluation the power and applicability of the proposed graph embedding framework is empirically verified on ten graph data sets with quite different characteristics. There are graphs that represent line drawings, gray scale images, molecular compounds, proteins, and HTML webpages. The main finding of the experimental evaluation is that the embedding procedure using dissimilarities with subsequent classification or clustering has great potential to outperform traditional approaches in graph based pattern recognition. In fact, regardless of the subsystems actually used for embedding, on most of the data sets the reference systems directly working on graph dissimilarity information are outperformed, and in the majority of the cases these improvements are statistically significant.

*Kaspar Riesen and Horst Bunke*