

Frequent subgraph mining and applications

Hoang Viet Dung

Abstract: Mining frequent subgraphs is a fundamental task in graph-based data mining, with critical applications in bioinformatics, cheminformatics, and social network analysis. This thesis addresses the computational challenges of discovering all subgraphs that occur frequently within a given graph database or a single large graph. We propose novel algorithms that leverage advanced pruning techniques and efficient canonical representation to reduce the exponential search space, thereby improving scalability for large-scale and high-density graph datasets. The core contribution includes a depth-first search framework integrated with a dynamic support counting mechanism, which significantly outperforms existing methods in both runtime and memory usage. Furthermore, this work explores practical applications, demonstrating how extracted frequent subgraphs serve as discriminative features for graph classification and as structural motifs for anomaly detection in complex networks. Experimental evaluations on real-world datasets validate the effectiveness of our approach, highlighting its potential for uncovering meaningful patterns in diverse domains. The findings provide a robust foundation for future research in graph pattern mining and its interdisciplinary applications.

Keywords: frequent subgraph, graph mining, applications, graph