

## ABSTRACT

Enumerating maximal bicliques has several applications in the field of data mining like online social networks, protein interactions in a biological field and telecommunication network users data analysis. Finding a subgroup, who have complete interaction in bipartite context from an online social network is an interesting problem which is commercially more useful for business communities. Two protein groups which have a complete interaction can be identified by finding the maximal bicliques from the protein interaction graphs that is useful for a variety of purposes. This idea can be extended to the analysis of user communities, to find interacting telecommunication network user communities that perhaps are commercially useful.

The graph can be transformed into a 2-dimensional matrix representation. The interaction between a pair of users is represented by '1', and '0', if there is no interaction. Maximal biclique is one set of row elements is having complete interaction with another set of column elements and this 1-rectangle cannot be expanded, i.e., no superset of this 1-rectangle has complete interaction. For the same set of users, interaction matrix can be generated at different period of time, which results in 3-dimensional symmetric matrix, where the third dimension is time period. In 3D context, it will be useful in analyzing the density of interaction over a period of time.

When the closed pattern mining algorithms if applied for symmetric data sets with diagonal elements zero, obviously it generates the maximal biclique patterns twice. One among the two is a duplicate pattern. The major objective of this thesis is to design efficient algorithms to generate the maximal biclique patterns only once.

Enumerating maximal biclique subgraphs from a graph is a computationally challenging problem, as the number of patterns can be exponentially large with respect to the increase in number of vertices of a graph. The closed pattern algorithms generally fit into any one of the following two strategies namely 1) grouping of ‘1’ contained rows and columns and 2) eliminating ‘0’ contained rows and columns. This thesis work addresses the following problems: 2D Maximal Bicliques Mining in ‘1’ Grouping Context and 2D Maximal Bicliques Mining in ‘0’ Removal Context from a 2-dimensional Boolean symmetric adjacency matrix, 3D Maximal Bicliques Mining in ‘0’ Removal Context and 3D Maximal Bicliques Mining in ‘1’ Grouping Context from a 3-dimensional Boolean symmetric adjacency matrix.

An algorithm named TWINBLADE is proposed for efficient enumeration of maximal bicliques mining in ‘1’ grouping context from 2-dimensional boolean symmetric adjacency matrix. The proposed algorithm exploits the symmetric property of input to reduce the search space by larger extent, and prune the duplicate patterns completely. Here, grouping ‘1’ contained row-columns strategy is adopted to generate the patterns.

Two algorithms named MD-Miner and MD-Miner+ are proposed for efficient enumeration of maximal bicliques mining from 2-dimensional boolean symmetric adjacency matrix in ‘0’ removal context. Both the algorithms operate in ‘0’ removal strategy. This strategy is adopted with the help of cutter generation. A new enumeration strategy and a novel right cutter constraint technique are introduced. MD-Miner algorithm applies the right cutter constraint technique on left generated node, whereas MD-Miner+ applied on all the leaf nodes. Experimental results show the better performance of these algorithms for highly dense datasets. The pros and cons of each of the algorithms are clearly described.

An algorithm named Cubeminer-MBC\* is proposed for efficient enumeration of maximal bicliques mining from 3-dimensional boolean symmetric adjacency matrix in ‘0’ removal context. For eliminating ‘0’ containers, cutter generation strategy is introduced. By eliminating ‘0’ contained row-columns with user specified minimum size constraints for all the dimensions, the proposed algorithm introduces row-column equal column removal strategy and closed column set checks to eliminate the duplicate patterns completely. Experimental results show that the proposed algorithm reduces the running time to a larger extent.

An algorithm named S-Datapeeler is proposed for efficient enumeration of maximal bicliques mining in ‘1’ grouping context from 3-dimensional boolean symmetric adjacency matrix. The major challenge lies in the elimination of all duplicate patterns and it is accomplished with the help of a novel element selection technique and subtree pruning strategy. The experiments involving several synthetic datasets show the effectiveness of the novel element selection technique in terms of reduction in search space by 50% and massive reduction in running time of the algorithm.