Mining for Tree-Query Associations in a Graph

Eveline Hoekx and Jan Van den Bussche
Hasselt University and transnational University of Limburg
Agoralaan D, 3590 Diepenbeek, Belgium
{eveline.hoekx, jan.vandenbussche}@uhasselt.be

Abstract

New applications of data mining, such as in biology, bioinformatics, or sociology, are faced with large datasets structured as graphs. We present an efficient algorithm for mining associations between tree queries in a large graph. Tree queries are powerful tree-shaped patterns featuring existential variables and data constants. Our algorithm applies the theory of conjunctive database queries to make the generation of association rules efficient. We propose a practical, database-oriented implementation in SQL, and show that the approach works in practice through experiments on data about food webs, protein interactions, and citation analysis.