

STATISTICAL ANALYSIS AND DATA MINING

Original Article

FS³: A sampling based method for top - *k* frequent subgraph mining*Tanay Kumar Saha, Mohammad Al Hasan 

First published: 27 July 2015

<https://doi.org/10.1002/sam.11277>

Cited by: 4

* This research is supported by Mohammad Hasan's NSF CAREER Award (IIS - 1149851).



About



Access

ols



ire

Abstract

Mining labeled subgraph is a popular research task in data mining because of its potential application in many different scientific domains. All the existing methods for this task explicitly or implicitly solve the subgraph isomorphism task, which is computationally expensive, and thus they suffer from the lack of scalability problem when the graphs in the input database are large. In this work, we propose FS³, which is a sampling - based method. It mines a small collection of subgraphs that are most frequent in the probabilistic sense. FS³ performs a Markov chain Monte Carlo (MCMC) sampling over the space of a fixed - size subgraphs such that the potentially frequent subgraphs are sampled more often. Besides, FS³ is equipped with an innovative queue manager. It stores the sampled subgraph in a finite queue over the course of mining in such a manner that the top - *k* positions in the queue contain the most frequent subgraphs. Our experiments on the database of large graphs show that FS³ is efficient and obtains subgraphs that are the most frequent among the subgraphs of a given size.

Citing Literature Supporting Information 

Help & Support



Opportunities



Connect with Wiley



Copyright © 1999-2018 John Wiley & Sons, Inc. All rights reserved

WILEY