

**A Two-Phase Algorithm for Differentially Private Frequent Sub graph Mining**

**Abstract:**

Mining frequent subgraphs from a collection of input graphs is an important task for exploratory data analysis on graph data. However, if the input graphs contain sensitive information, releasing discovered frequent subgraphs may pose considerable threats to individual privacy. In this paper, we study the problem of frequent subgraph mining (FSM) under the rigorous differential privacy model. We present a two-phase differentially private FSM algorithm, which is referred to as DFG. In DFG, frequent subgraphs are privately identified in the first phase, and the noisy support of each identified frequent subgraph is calculated in the second phase. In particular, to privately identity frequent subgraphs, we propose a frequent subgraph identification approach, which can improve the accuracy of discovered frequent subgraphs through candidate pruning. Moreover, to compute the noisy support of each identified frequent subgraph, we devise a lattice-based noisy support computation approach, which leverages the inclusion relations between the discovered frequent subgraphs to improve the accuracy of the noisy supports. Through formal privacy analysis, we prove that DFG satisfies ε-differential privacy. Extensive experimental results on real datasets show that DFG can privately find frequent subgraphs while achieving high data utility.