Quality-Aware Sub graph Matching over Inconsistent Probabilistic Graph Databases

Resource Description Framework (RDF) has been widely used in the Semantic Web to describe resources and their relationships. The RDF graph is one of the most commonly used representations for RDF data. However, in many real applications such as the data extraction/integration, RDF graphs integrated from different data sources may often contain uncertain and inconsistent information (e.g., uncertain labels or that violate facts/rules), due to the unreliability of data sources. In this paper, we formalize the RDF data by inconsistent probabilistic RDF graphs, which contain both inconsistencies and uncertainty. With such a probabilistic graph model, we focus on an important problem, quality-aware sub graph matching over inconsistent probabilistic RDF graphs (QA-gMatch), which retrieves sub graphs from inconsistent probabilistic RDF graphs that are isomorphic to a given query graph and with high quality scores (considering both consistency and uncertainty). In order to efficiently answer QA-gMatch queries, we provide two effective pruning methods, namely adaptive label pruning and quality score pruning, which can greatly filter out false alarms of sub graphs. We also design an effective index to facilitate our proposed pruning methods, and propose an efficient approach for processing QA-gMatch queries. Finally, we demonstrate the efficiency and effectiveness of our proposed approaches through extensive experiments.