DOCTORAL THESIS

Graph query autocompletion

Yi, Peipei

Date of Award:
2018

Abstract

The prevalence of graph-structured data in modern real-world applications has led to a rejuvenation of research on graph data management and analytics. Several database query languages have been proposed for textually querying graph databases. Unfortunately, formulating a graph query using any of these query languages often demands considerable cognitive effort and requires "programming" skill at least similar to programming in SQL. Yet, in a wide spectrum of graph applications consumers need to query graph data but are not proficient query writers. Hence, it is important to devise intuitive techniques that can alleviate the burden of query formulation and thus increase the usability of graph databases. In this dissertation, we take the first step to study the graph query autocompletion problem. We provide techniques that take a user's graph query as input and generate top-k query suggestions as output, to help to alleviate the verbose and error-prone graph query formulation process in a visual environment.

Firstly, we study *visual query autocompletion for graph databases*. Techniques for query autocompletion have been proposed for web search and XML search. However, a corresponding capability for graph query engine is in its infancy. We propose a novel framework for graph query autocompletion (called AUTOG). The novelties of AUTOG are as follows: First, we formalize query composition that specifies how query suggestions are formed. Second, we propose to increment a query with the logical units called *c-prime features*, that are (i) frequent subgraphs and (ii) constructed from smaller c-prime features in no more than c ways. Third, we propose algorithms to rank candidate suggestions. Fourth, we propose a novel index called *feature* DAG (FDAG) to further optimize the ranking.

Secondly, we propose *user focus-based graph query autocompletion*. AUTOG provides suggestions that are formed by adding subgraph increments to arbitrary places of an existing user query. However, humans can only interact with a small number of recent software artifacts in hand. Hence, many such suggestions could be irrelevant. We present the GFOCUS framework that exploits a novel notion of *user focus of graph query formulation*. Intuitively, the focus is the subgraph that a user is working on. We formulate *locality principles* to automatically identify and maintain the focus. We propose novel monotone submodular ranking functions for generating *popular* and *comprehensive* query suggestions only at the focus. We propose efficient algorithms and an index for ranking the suggestions.

Thirdly, we propose *graph query autocompletion for large graphs*. Graph features that have been exploited in AUTOG are either absent or rare in large graphs. To address this, we present *Flexible* graph query autocompletion for LArge Graphs, called FLAG. We propose *wildcard label* for query graph and query suggestions. In particular, FLAG allows augmenting users' queries using subgraph increments with wildcard labels, which summarize query suggestions that have similar increment structures but different labels. We propose an efficient ranking algorithm and a novel index, called *Suggestion Summarization* DAG (SSDAG), to optimize the online suggestion ranking.

Detailed problem analysis and extensive experimental studies consistently demonstrate the effectiveness and robustness of our proposed techniques in a broad range of settings.

Keywords: Subgraph query, query autocompletion, graphs, database usability.

Table of Contents