Towards Hybrid Methods for Graph Pattern Queries

Foteini Katsarou
School of Computing Science
University of Glasgow, UK
f.katsarou.1@
research.gla.ac.uk

Nikos Ntarmos
School of Computing Science
University of Glasgow, UK
nikos.ntarmos@
glasgow.ac.uk

Peter Triantafillou
School of Computing Science
University of Glasgow, UK
peter.triantafillou@
glasgow.ac.uk

ABSTRACT

In the subgraph querying problem, given a query graph and a dataset of graphs, we want to locate which graphs in the dataset contain the query and/or find all its occurrences. Over the years, numerous methods, fragmented in 2 categories, were proposed to tackle the problem. In the first category, methods follow the filter-then-verify (FTV) paradigm where an index is used to filter out graphs that definitely do not contain the query as an answer. On the remaining set of graphs, a subgraph isomorphism algorithm is applied to verify the graphs that contain the query graph. A second category, so called no-filter, verify (NFV), invested in optimizing the subgraph isomorphism process, by employing a lightweight index primarily to locate candidate vertices on the graphs in the dataset. The current trend is to totally dismiss the FTV methods and employ NFV methods instead. With this work we wish to point out that a hybrid solution that is, a combination of the filtering of some FTV method with the subgraph isomorphism test of top-performing NFV methods, can be highly competitive in terms of performance vis-a-vis methods from either the FTV or the NFV category. We will present some preliminary results, based on experiments, where such a combination proves to be beneficial by primarily avoiding the initiation of a large number of redundant subgraph isomorphism tests.

Keywords

Graph databases, graph query processing, subgraph isomorphism

1. INTRODUCTION

Graphs are ideal for representing complex entities as they physically incorporate relationships/interactions in the form of edges. Revealing the existence of a pattern graph in various graphs/parts of the same graph is essential to graph analytics. In *subgraph querying*, given a pattern graph (query) and a dataset of graphs, we want to know whether the query graph is contained in each graph and/or find all its occurrences within a larger stored graph. Subgraph querying entails the subgraph isomorphism problem (abbreviated as sub-iso), which is known to be NP-complete. Over the years, subgraph querying has received a lot of attention which is evident by the numerous proposed methods. Related work is categorized in 2 major categories: the *filter-then-verify (FTV)* methods and the *no-filter, verify (NFV)* methods. Various recent experimental analysis papers (e.g. [2, 3, 1]) compare and stress-test the proposed methods and provide interesting insights about the performance of both FTV and NFV methods.

2. APPROACH, RESULT OVERVIEW, AND CONCLUSIONS

The latest works, e.g. [3], tend to dismiss the FTV methods with the claim that the fast sub-iso test of the NFV methods significantly outperforms the index-based scan of the FTV methods. In the current work, we put forth our concern about the possibility that lots of current state of the art and on-going efforts miss a potential for deriving highly performant algorithms by investigating hybrid solutions which maintain the filtering of the FTV and the much faster sub-iso test algorithms of the NFV methods. Based on our initial experiments, this design option can introduce very significant performance gains in the subgraph matching problem. Essentially, the benefits stem from using the FTV-style indexes in order to avoid the need to perform the sub-iso test on a large majority of the graphs in the graph DB. And the benefits are large both against FTV and NFV methods alone.

3. REFERENCES

[1] F. Katsarou, N. Ntarmos, and P. Triantafillou. Subgraph Querying with Parallel Use of Query Rewritings and Alternative Algorithms. In *EDBT*, '17.

[2] F. Katsarou, N. Ntarmos, and P. Triantafillou. Performance and scalability of indexed subgraph query processing methods. *PVLDB*, 8(12):1566–1577, 2015.

[3] J. Lee, W.-S. Han, R. Kasperovics, and J.-H. Lee. An in-depth comparison of subgraph isomorphism algorithms in graph databases. *PVLDB*, 6(2), 2012.