Abstract

Since many real applications such as web connectivity, social networks, and so on, are emerging now-a-days, thus graph databases have been commonly used as significant tools to exemplify and query complex graph data wherein each vertex in a graph usually contains information, which can be modeled by a set of tokens or elements. The method for subgraphs extraction by considering set similarity query over a large graph database has already been proposed, which retrieves subgraphs that are structurally isomorphic to the query graph, and meanwhile satisfy the condition of vertex pair matching with the (dynamic/fixed) weighted set similarity in a centralized system. This paper explains the efficient implementation of subgraphs extraction in a large graph database in a distributed environment by considering both vertex set similarity and graph topology which offers a better price/performance ratio and increases availability using redundancy when parts of a system fail than centralized systems in case of a large dataset (i.e., a graph with millions/billions of nodes wherein each node contains some information) by performing parallel processing.
References

11. Liang Hong, Member, IEEE, Lei Zou, Member, IEEE, Xiang Lian, Member, IEEE, and Philip S. Yu, Fellow, IEEE, “Subgraph Matching with Set Similarity in a Large Graph Database”, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 9, SEPTEMBER 2015

Index Terms
Computer Science Distributed Systems

Keywords
apache spark, distributed systems, graph topology, set similarity