

Many problems in areas as diverse as recommendation systems, social network analysis, semantic search, and distributed root cause analysis can be modeled as pattern search on labeled graphs which is also called “heterogeneous information networks” (HINs) [2]. Given a large graph and a query pattern with node and edge label constraints, a fundamental challenge is to find the top-k matches according to a ranking function over edge and node weights.

Since it is difficult to select value k for users, we decide to investigate different aspect of the field. In this project, we implement an any-k ranking algorithm [1]: for a given time budget, return as many of the top-ranked results as possible. Then, given additional time, produce the next lower-ranked results quickly as well. It can be stopped anytime, but may have to continue until all results are returned. We focus on acyclic patterns over arbitrary labeled graphs.

We are interested in practical algorithms that effectively exploit

- (1) properties of heterogeneous networks, in particular selective constraints on labels.
- (2) the users often explore only a fraction of the top-ranked results.

We will implement Kernelization And Rapid Pruning-based Exploration for Tree patterns (KARPET) which carefully integrates aggressive pruning that leverages the acyclic nature of the query, and incremental guided search. We want to validate the strong non-trivial time and space guarantees in the original paper, which is generally considered very hard for this type of graph search problem.

REFERENCES

- [1] Any-k: Anytime Top-k Tree Pattern Retrieval in Labeled Graphs
- [2] Jiawei Han, Yizhou Sun, Xifeng Yan, and Philip S Yu. 2010. Mining knowledge from databases: an information network analysis approach. In Proc. ACM SIGMOD. 1251–1252.