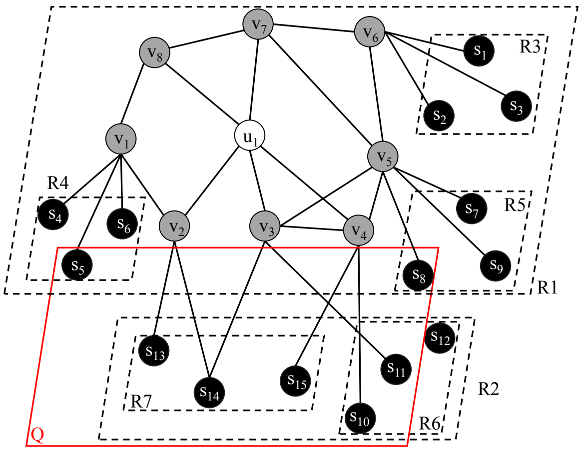


Figure 2: Example Graph

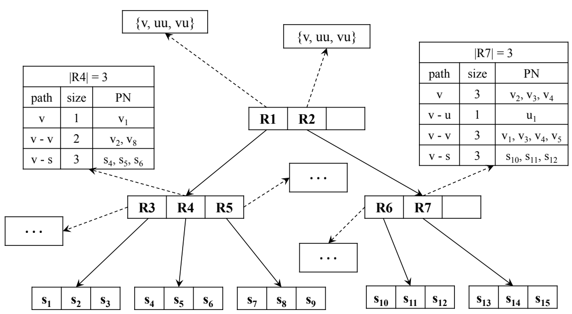
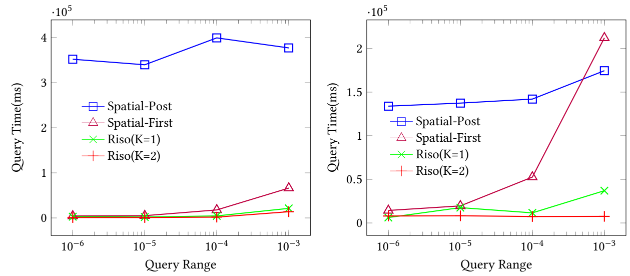


Figure 3: Riso-Tree Structure

2-hops are considered. For the lowest-level MBRs, for instance $R4$, there are three types of paths, v, vv and vs. The vertices that can be reachable from $R4$ through path vv are v_2, v_8, denoted by PN^p_{R4}. For MBRs in higher-level, for instance $R1$, Riso-Tree only stores all the existing paths that start from spatial objects within $R1$, which are $\{v, vv, vs\}$.

3 QUERY PROCESSING

Existing graph database has its own optimizer for determining strategy of executing a SGPM based on graph statistical information, such as cardinality of labels and the average degree. However, it ignores the influence of spatial predicate in SGPM. The influence of spatial predicate has two aspects. One is that the spatial vertex in the query can be highly-selective when area of the query rectangle is small. Then it is intuitive to start the search from these spatial objects because the search space is smaller. Another aspect is that spatial predicates not only change cardinality of the spatial vertex in the query, but also that of other vertices connected to the spatial vertex. Use the query $u - v - s$ where s should be within Q as an example. The algorithm will compute all the influential paths in the query, which are v and vu. Then it searches Riso-Tree from the root node. Even though both $R1$ and $R2$ can satisfy the spatial predicate, existing paths of $R1$ does not contain one influential path vu. So the tree branch of $R1$ can be safely pruned. Then the algorithm



(a) Foursquare **(b) Gowalla**

Figure 4: Query response time (spatial selectivity ranging from 0.0001 to 0.1)

goes to the children of $R2$. The mapping set for v can be computed by performing the union operation on PN^v_{R6} and PN^v_{R7}. So the mapping set of v is reduced to $\{v_2, v_3, v_4\}$. Then the search will go to the leaf level of Riso-Tree to calculate the candidate set of spatial query vertex s. For this case, it will be $\{s_{10}, s_{11}, s_{13}, s_{14}, s_{15}\}$. As a result, the size of matching candidates for both spatial and non-spatial vertices in the query can be reduced by using Riso-Tree.

4 EXPERIMENT

Figure 4 shows query time of using Spatial-Post, Spatial-First and Riso-Tree with K being set to 1 and 2 respectively. Two real datasets, Foursquare and Gowalla are used in our evaluation. 100 non-spatial labels are randomly assigned to their original non-spatial vertices. The selectivity of the spatial predicate varies between 10^{-6} and 10^{-3}. The Spatial-Post approach has the worse performance because it cannot take advantage of the high-selective spatial predicate. Spatial-First can perform well when spatial selectivity is high. But it can become slow when spatial selectivity increases and it is even worse than Spatial-Post approach. Both Riso-Tree approaches can outperform the baseline approaches. They can also keep a stable performance in different spatial selectivities. It can be observed that Riso-Tree($K = 2$) can achieve better performance than Riso-Tree($K = 1$) because it has more information of reachable subgraphs which incurs more pruning power. To sum up, Riso-Tree can achieve 100x better performance than the baseline incur in that it can reduce the search space for both spatial and non-spatial vertex in the query.

REFERENCES

[1] Xiangnan Kong, Philip S Yu, Ying Ding, and David J Wild. 2012. Meta path-based collective classification in heterogeneous information networks. In *Proceedings of the 21st ACM international conference on Information and knowledge management.* ACM, 1567–1571.

[2] Chuan Shi, Xiangnan Kong, Philip S Yu, Sihong Xie, and Bin Wu. 2012. Relevance search in heterogeneous networks. In *Proceedings of the 15th International Conference on Extending Database Technology.* ACM, 180–191.

[3] Chuan Shi, Yitong Li, Jiawei Zhang, Yizhou Sun, and S Yu Philip. 2017. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering* 29, 1 (2017), 17–37.

[4] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment* 4, 11 (2011), 992–1003.

[5] Yizhou Sun, Brandon Norick, Jiawei Han, Xifeng Yan, Philip S Yu, and Xiao Yu. 2013. Pathselclus: Integrating meta-path selection with user-guided object clustering in heterogeneous information networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 7, 3 (2013), 11.