Evaluating Multi-Way Joins over Discounted Hitting Time

Mr. Zhang Wangda

Abstract:

The prevalence of graphs in emerging applications has recently raised a lot of research interests. To acquire interesting information hidden in large graphs, tasks including link prediction, collaborative recommendation, and reputation ranking, all make use of proximities between graph nodes. The discounted hitting time (DHT), which is a random-walk similarity measure for graph node pairs, has shown to be useful in various applications. In this seminar, we examine a novel query, called the multi-way join (or n-way join), over DHT scores. Given a graph and n sets of nodes, the n-way join retrieves a ranked list of n-tuples with the k highest scores, according to some aggregation function of DHT values. By extracting such top-k results, this query enables the analysis and prediction of various complex relationships among n sets of nodes on a large graph.

Since an n-way join is expensive to evaluate, we develop the Partial Join algorithm (or PJ), which decomposes an n-way join into a number of top-m 2-way joins, and combines their results to construct the answer of the n-way join. As the process of PJ may necessitate the computation of top-(m + 1) 2-way joins, we study an incremental solution, which saves the trouble of recomputation and allows the results of top-(m+1) 2-way join to be derived quickly from the top-m 2-way join results earlier computed. For better performance, we further examine efficient processing algorithms and pruning techniques for 2-way joins. Through extensive experiments on three real graph datasets, we show that the proposed PJ algorithm accurately evaluates n-way joins, and is significantly faster than basic solutions.

About the Speaker:

Mr. Zhang Wangda is a full-time MPhil student supervised by Prof. Ben Kao and Dr. Reynold Cheng. His research interest is database and data mining.