MASTERS
FINAL ORAL EXAMINATION

Monday, May 9th
2:00p.m. @ 213 Atanasoff

Yuanyuan Tang
Faculty Advisor: Professor Wallapak Tavanapong
Iowa State University

Recent years have seen advances in graph databases and graph database management systems (GDBMS). In a typical graph data model, a node represents an entity and an edge represents a relationship between two nodes. Nodes and edges typically have associated properties (attributes) and are assigned types for fast query response times. In the application domains where relationships are of importance, GDBMS increasingly gains popularity since the relationships can be explicitly modeled and easily visualized in a graph data model. To measure performance of GDBMS, a number of graph database benchmarks have been proposed. Nonetheless, these benchmarks are not yet as rigorous as those of relational database manage ment systems (RDBMS). Inspired by Wisconsin Benchmark, we propose Cyclone Benchmark for measuring performance of graph databases in several aspects of which some have not been investigated in the literature. Our benchmark comes with (1) two data graph models: a simple model with all nodes of the same node type and a complex model with multiple node types, (2) data graph generation programs, and (3) Create, Read, Update, and Delete (CRUD) operations. The data graph generation programs create a graph structure and annotate values of node and edge attributes in the graph to allow for a study of the impact of attribute selectivity factors as well as a correlation between attributes. The programs generate a predened graph structure, a random graph, or a Kronecker graph that has been shown to model real-world networks well. The read operations include several graph structure queries.

We measured the average execution times of the proposed CRUD operations on several synthetic graphs generated by the benchmark with a varying number of nodes from 1,000 to 1,000,000 nodes. The graphs were stored as graphs in Neo4j, a popular native GDBMS, and as relations in MySQL, a popular RDBMS. For most CRUD operations including the graph structure queries in our benchmark, MySQL was signicantly faster than Neo4j.

IOWA STATE UNIVERSITY
Department of Computer Science