A Demonstration of Query-Oriented Distribution and Replication Techniques for Dynamic Graph Data

Analysis of Large, Dynamic Networks
- Transportation
- Social and Political Studies / Marketing / National Security
  - How do communities or the centrality of an entity change over time?
  - Who are rising stars?

The G* System (1/3)
- distribution of graph snapshots

Impact of Snapshot Distribution (Example)
- 100 similarly-sized graph snapshots
- 100 G* workers
- PageRank on one snapshot for all snapshots
  - query: 1 worker/snapshot
  - all snapshots: 2.000 seconds

Segment Exchange (Example)
- poor balancing: low locality
- good balancing: high locality

Segment Exchange (Algorithms)

Graph Snapshot Replication
- \( r \) copies of each snapshot to mask up to \( r-1 \) simultaneous worker failures
- queries classified into \( r \) categories
- \( j \)-th replica optimized for the \( j \)-th query category (e.g., one replica distributed over many workers, another replica distributed over a few workers)

This work has been supported by the National Science Foundation under CAREER award IIS-1149372