Advanced Data Management

Medha Atre

Office: KD-219
atrem@cse.iitk.ac.in

Sept 22, 2016

Announcements

- Assignment-3 will be posted in the next couple of days. It will be a programming assignment requiring use of Hadoop.

How it is all tied together

- Data Management
  - Single compute node.
  - Cluster.
- For single compute node – query optimization, data compression, indexing.
- For a cluster – all the aspects above plus cost of data shipping, data distribution strategies etc.
- Special topics in graph data – reachability queries, regular path queries, top-k queries, keyword search.
- Theoretical aspects of query optimization and data distribution.

Reachability defined

Given a graph $G(V, E)$ with V as the set of nodes and E as the set of edges, a reachability query asks – does there exists *any* path between nodes x and y.
In case of directed graph, the path is directed and other for undirected graphs the path is undirected.
The graphs may have cycles too, giving rise to "strongly connected components".

Query Processing Challenges

- Identification of strongly connected components and merging them.
- Efficient traversal over large graphs, e.g., number of vertices and edges being several millions.
- High computational complexity $O(V^3)$ for standard algorithms like all-pair shortest path, which also gives us *transitive clousure* for answering reachability.
- High storage space $O(V^2)$ if the entire transitive clousure is meant to be stored.

Some specific approaches

- 2-hop labeling.
- Interval labeling.
- Compressed bit-vectors.
- Chain cover.

Some trivial brute-force approaches

- Repeated adjancency matrix multiplication. $R = A + A^2 + A^3 ... + A^n$ Computational complexity is $O(n^4)$, note that at a time we do only one matrix multiplication and we do this n times.
- Do a DFS walk over the DAG of the graph (after coalescing SCCs) and construct compressed bit-vectors [SIGMOD 2010].
- Do repeated walk over the DAG of the graph, and construct hyper-dimensional label covers [Grail, VLDB Conf. 2010].