0.1 Papers read by Ning Dong

The first paper was the GBASE paper by U Kang [**?**]

- *Main idea*:
  This paper proposes a graph processing system called GBASE. It aims to solve 3 problems in graph analysis using their framework. First, represent and store graphs in distributed settings efficiently. Second, define a set of core algorithms. Third, optimize user queries on the graph.
  For storage, this paper represents graphs in homogeneous blocks(either very sparse or very dense) and compresses them using zip compression or Gap Elias-γ encoding. Using grid placement, their system guarantees both in-neighbor and out-neighbor queries are handled efficiently, compared to vertical and horizontal placement which has O(K) file accesses where K is number of files.
  For defining core algorithms, this paper gives matrix-vector version and SQL version for graph mining operations including k-step neighbors, induced subgraph and single-source shortest distance. In practice, SQL queries could be highly optimized in both single machine and distributed clusters. Users do not need to care about lower level operations when system scales up.
  For query optimization, this paper introduces their grid selection strategy which selects only relevant grids for targeted queries. And they answer incidence matrix queries with adjacency matrix in hadoop mapreduce scheme.
  This paper presents experiment results in different settings and run experiments on graphs of different scale. GBASE is space and time efficient in indexing and storage. It also gives better performance in both global and targeted query time.
- *Use for our project*:
  When building GBASE, a set of core algorithms in graph analysis is implemented in SQL, which would be extreamely useful resource to reference, when we implement D-CUBE in SQL.
- *Shortcomings*:
  Hadoop mapreduce is a heavy I/O task where a lot of read/write operations to disks are involved. In tasks where user wants to use induced graph repetitively, I/O cost would be huge. Today computing frameworks like Spark may offer better in memory solution to this problem and could be plugged into the system.

The second paper was the Graph Analytics paper by Alekh Jindal [**?**]

- *Main idea*:
- *Use for our project*:
- *Shortcomings*:

The third paper was by . . .

0.2 Papers read by Mary Thompson

. . .

0.3 Papers read by Michael Miller

. . .