



# Social Distance Computation

Hoi Ki Leung, Wai Fun Wong  
Advised by Dimitris PAPANIKOLAOU

# Abstract

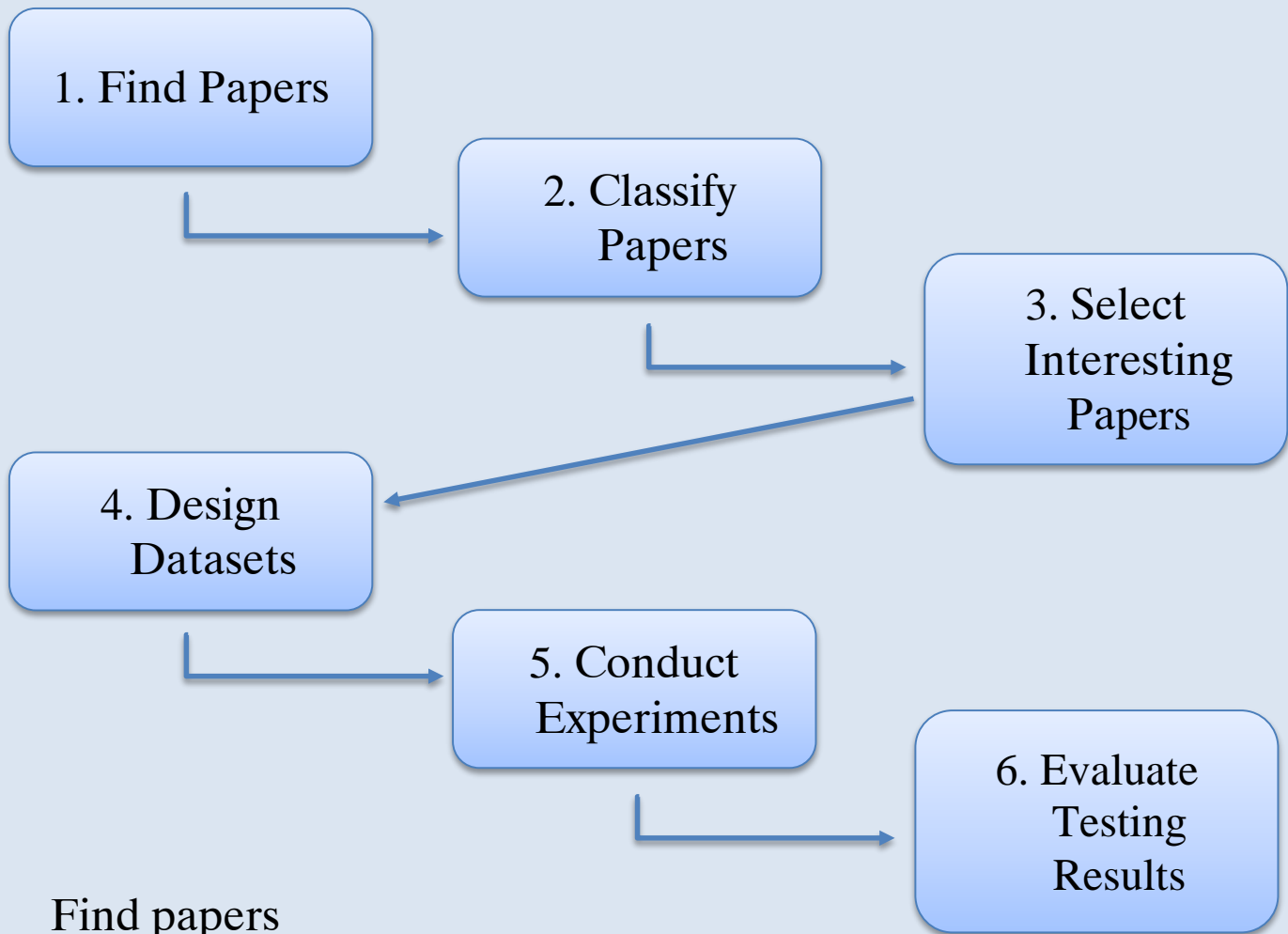
This report discusses the existing social distance computations and tests them for the evaluations. First, we find the papers of the existing works. Next, we categorize the papers we found with some specific categories. Moreover, we focus on two papers, which are from different ways on finding the shortest distance and under different settings. The different ways include solving problems in exact or approximate method while different settings imply different kinds of networks such as social networks and road networks. Then, we design two approaches for the experiment which are using random queries and our designed queries. Finally, we test the codes and evaluate them based on the testing results. We find that both the approximate and exact methods have pros and cons on different kinds of data.

# Objectives

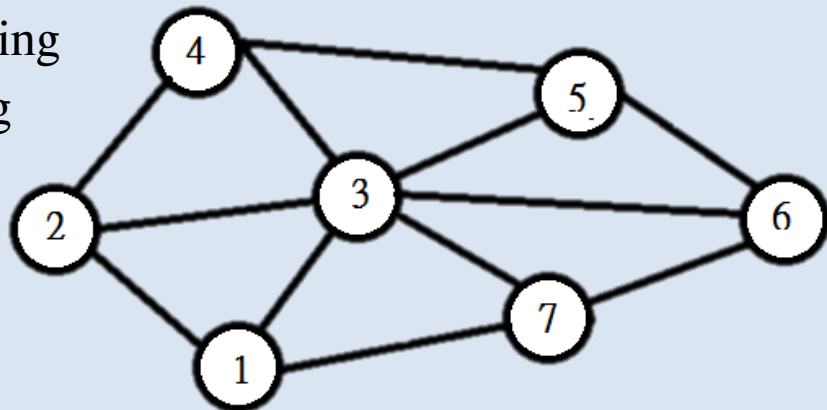
The goal of this final year project was to find tools to evaluate these state-of-the-art researches on social distance computation, resulting in analyzing the highlights and drawbacks, such as accuracy and time required for querying, of each research via a series of fair experiments for different types of graphs (or data). In this project, the following objectives were focused on:

1. To survey the state-of-the-art research papers on social distance computation.
2. To build a table for categorizing the papers.
3. To study deeply the earliest papers about exact and approximate distance
4. To build the data sets to store different kinds of data to test algorithms and store the testing results.
5. To implement the existing algorithms to address the (1) accuracy; (2) execution time including loading time, indexing time and querying time.

# Methodology

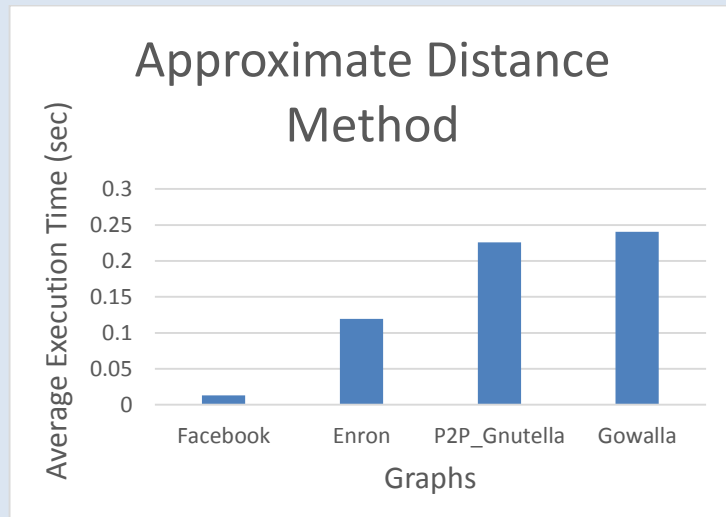
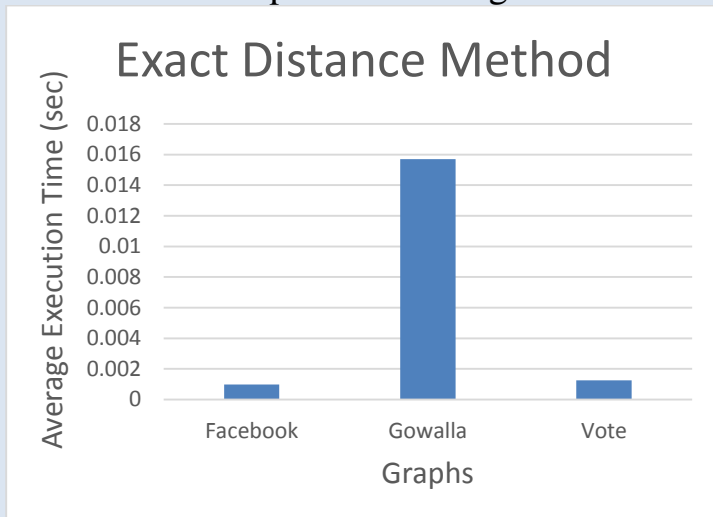


1. Find papers
  - Published in well-known scientific journals
2. Classify papers
  - Using categories
3. Select interesting papers
  - Hop Doubling Label Indexing
  - Pruned Landmark Labeling
4. Design datasets
  - Storing the test cases
  - Saving the results
5. Conduct experiments
  - 100 random queries
  - 50 own designed queries for Facebook dataset
6. Evaluate testing results
  - Execution Time: Loading time, Indexing time, Querying time
  - Shortest Distance



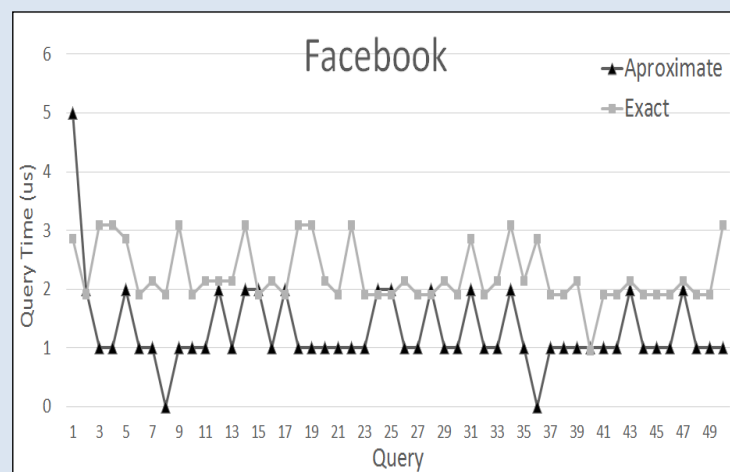
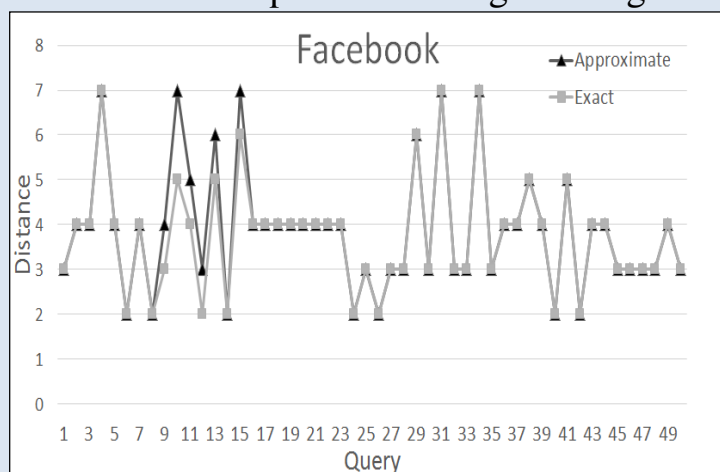
# Results

The results of experiment using 100 random queries shown as below:



- The **larger** the input data file is, the **longer** the average execution time used.
- The average execution time in EXACT method is **shorter** than that in APPROXIMATE one.

The results of experiment using 50 designed queries for Facebook dataset shown as below:



- The answers of the queries are almost the same in BOTH methods.
- The query time of APPROXIMATE approach is shorter than EXACT one.

# Conclusions

- Social distance computations can be categorized into different ways on computing the distance and under different settings like types of graphs.
- APPROXIMATE and EXACT methods are both accurate enough.
- APPROXIMATE computations have faster query time.
- EXACT computations have simple indexing and loading.
- BOTH approaches have highlights and limitations.