Measuring Graphs with Shortest Distances

by

WENTAO LI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Australian Artificial Intelligence Institute (AAII)
Faculty of Engineering and Information Technology (FEIT)
University of Technology Sydney (UTS)

January, 2021

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Wentao Li declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: 12/01/2021

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my principal supervisor A/Prof. Lu Qin. Lu is a responsible teacher, a kind friend, and a supportive brother. As a teacher, he guides me to acquire the necessary research skills. I have benefited greatly from discussions with him. As a friend, he is always ready to help me. Whenever I feel frustrated, he is ready to offer advice and practical support. As a brother, he teaches me how to balance research and life. His encouragement has helped me to have a positive attitude towards life.

Secondly, I would like to thank my co-supervisor Prof. Ying Zhang. He always provides enlightening advice on my research topics. In addition, he often encourages and guides me to optimize my career development plan.

Thirdly, I would like to thank my co-supervisor Dr. Miao Qiao. She always encourages me to explore the essence of a problem, from which I experience the enjoyment of conducting research. She patiently guides me in acquiring research skills, such as writing papers and preparing presentations. I learn a lot from her way of thinking about problems and her rigorous attitude to research.

Fourthly, I would like to acknowledge Prof. Xuemin Lin and Dr. Lijun Chang for their support — most of the works presented in this thesis are conducted together with them. I would like to thank Prof. Lin for serving as an example of an excellent researcher for me. I would like to thank Dr. Chang for his insightful

suggestions to improve the quality of the works.

I would also like to appreciate Prof. Wenjie Zhang, Dr. Xin Cao, Dr. Zengfeng Huang, Dr. Shiyu Yang, Dr. Yixiang Fang, Dr. Weiwei Liu, for sharing brilliant thoughts and experiences. Thanks to Dr. Dong Wen, Dr. Xubo Wang, Dr. Dian Ouyang, Dr. Fan Zhang, Dr. Longbin Lai, Dr. Long Yuan, Dr. Xing Feng, Dr. Haida Zhang, Dr. Wei Li, Dr. Chen Zhang, Dr. Yang Yang, Dr. Wanqi Liu, Dr. Kai Wang, Dr. You Peng, Dr. Boge Liu, Mr. Bohua Yang, Ms. Conggai Li, Mr. Mingjie Li, Mr. Junhua Zhang, Mr. Yuxuan Qiu, Mr. Hanchen Wang, Mr. Yilun Huang, Mr. Yuanhang Yu, Mr. Peilun Yang, Ms. Xiaoshuang Chen, Mr. Xuefeng Chen, Mr. Yuren Mao, Mr. Yixing Yang, Mr. Zhengyi Yang, Mr. Qingyuan Linghu, Mr. Michael Ruisi Yu, Mr. Chenji Huang, Mr. Yu Hao, Mr. Peng Zhang, Mr. Tao Shen, Mr. Xunxiang Yao, Mr. Xiaolin Zhang, Ms. Muwen Yang, Ms. Qian Li, Mr. Chengzhang Zhu. The days we spent together bear unforgettable memories.

I am also very grateful to my supervisors at Chongqing University: A/Prof. Hua Li and A/Prof. Min Gao, for stimulating my research interest and providing useful advice for my study. Furthermore, I would like to express my gratitude to Prof. Qingyu Xiong, Prof. Junhao Wen, A/Prof. Jun Zeng and A/Prof. Wei Zhou of Chongqing University for their kind help.

Finally, I would also like to acknowledge my father Mr. Dongyun Li, and my mother Ms. Linren Wang, who provide me with selfless love and endless encouragement. I would like to thank my grandparents, my sister, and other relatives and friends for their love and support.

ABSTRACT

Shortest distances characterize the pair-wise relationships among nodes in a graph. Given a graph with a node set and an edge set, the shortest distance between two nodes is defined as the minimum path length between them. Computing the shortest distance between two nodes is a fundamental operation of graphs, which can be used both as a primary function and as a building block for applications. Given the important roles of shortest distances, this thesis focuses on the efficient computation of shortest-distance-based measures on graphs.

Firstly, we investigate how to accelerate the index time of 2-hop labeling. 2-hop labeling approaches are widely adopted to speed up the online performance of shortest distance queries. The construction of the 2-hop labeling, however, can be exhaustive especially on big graphs. For a major category of large graphs, small-world networks, the state-of-the-art approach is Pruned Landmark Labeling (PLL). However, PLL's strong sequential nature hinders it from being parallelized. It becomes an urgent issue on massive small-world networks whose index can hardly be constructed by a single thread within a reasonable time. Based on the dependency analysis of PLL, we propose a Parallelized Shortest-distance Labeling (PSL) scheme to exploit parallelism to shorten the index time.

Secondly, we study how to reduce the index size of 2-hop labeling. While the index time can be shortened by parallelized labeling, the index size becomes the bottleneck for a massive real graph with a relatively large treewidth — 2-hop labeling can hardly be constructed due to the oversized index. We disclose

the theoretical relationships between the graph treewidth and 2-hop labeling's index size and query time. To scale up distance labeling, we propose Core-Tree (CT) Index to dramatically reduce the index size, thereby enabling CT-Index to handle massive graphs that no existing approaches can process.

Thirdly, we compute and maintain the eccentricities of all nodes. Given a graph, eccentricity measures the shortest distance from each node to its farthest node. Existing eccentricity computation algorithms are not scalable enough to handle real large networks. Our solution optimizes existing eccentricity computation algorithms on their bottlenecks — one node eccentricity computation and the upper/lower bounds update — based on a line of original insights; it also provides the first algorithm on maintaining the eccentricities of a dynamic graph without recomputing the eccentricity distribution upon each edge update.

Extensive empirical studies validate the efficiency of our techniques.

PUBLICATIONS

- ***Wentao Li***, *Miao Qiao, Lu Qin, Ying Zhang, Lijun Chang, and Xuemin Lin. "Scaling Distance Labeling on Small-World Networks." In Proceedings of the 2019 ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1060-1077. 2019. (**Chapter 3**)*

- ***Wentao Li***, *Miao Qiao, Lu Qin, Ying Zhang, Lijun Chang, and Xuemin Lin. "Scaling Up Distance Labeling on Graphs with Core-Periphery Properties." In Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1367-1381. 2020. (**Chapter 4**)*

- ***Wentao Li***, *Miao Qiao, Lu Qin, Ying Zhang, Lijun Chang, and Xuemin Lin. "Exacting Eccentricity for Small-World Networks." In 2018 IEEE 34th International Conference on Data Engineering (ICDE), pp. 785-796. IEEE, 2018. (**Chapter 5**)*

- ***Wentao Li***, *Miao Qiao, Lu Qin, Ying Zhang, Lijun Chang, and Xuemin Lin. "Eccentricities on Small-World Networks." The VLDB Journal 28, no. 5 (2019): 765-792. (**Chapter 5**)*

- ***Wentao Li***, *Min Gao, Fan Wu, Wenge Rong, Junhao Wen, Lu Qin. "Manipulating Black-Box Networks for Centrality Promotion." In 2021 IEEE 37th International Conference on Data Engineering (ICDE). IEEE, 20121.*

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES