

How good is CSR really?

Mahammad Valiyev

November 30, 2017

Abstract

In the last decade, the data size is growing exponentially and processing those data is becoming a difficult problem. Nowadays researchers and industry have been interested in the analysis of the graph to get deep understanding of social networks. Many of these graphs used in industry have become very large, containing hundreds of millions of nodes and edges. In this paper, I propose Compressed Sparse Row(CSR) which is a fast graph container. My CSR implementation supports updates on graph. On a static graph, CSR gives 6-10x better performance over Adjacency List Implementation(AL) on Depth-First-Search(DFS) and Breadth-First-Search(BFS) and 2-3x better performance on Dijkstra algorithm.

Keywords: Graph databases, Compressed Sparse Row

1 Introduction

Graph analysis have become more popular in the last few years. Graphs are usually used to represent social networks in which vertices indicate users and edges indicate friendship between users. Facebook has 1.39 billion active users as of 12/2014 with more than 400 billion edges.[1].

2 Related Works

3 Design choices

4 Implementation

few details about implementation

5 Evaluation

...

5.1 Experimental platform

5.2 Algorithm comparisons

5.3 Scaling

5.4 Complex Scenario

5.5 Memory consumption

5.6 Cache friendliness

References

- [1] Ching, A., Edunov, S., Kabiljo, M., Logothetis, D., Muthukrishnan, S. (2015). One trillion edges: Graph processing at facebook-scale. Proceedings of the VLDB Endowment, 8(12), 1804-1815