Filtering: A methods for Solving Graph Problems in MapReduce – First Draft

He Xinyu

May 7, 2024

Abstract

The graph data from both industry and academia can be too large to compute in one machine's RAM. To be able to process this data, the MapReduce framework was created to split the graph data into multiple small data and then process them in parallel. In the original paper, the author presents a general algorithmic design technique in the MapReduce framework called *filtering*. Filtering reduces the size of the input into distributed forms and uses distributed machines to solve the targeted problems. The first draft of this paper only showcases the understanding of the problem and the research plan. The full understanding of the given paper is not accomplished in this phase, and errors are likely to be present in this draft.

1 Introduction

How important the problem is? What is the problem? How do we solve it in the past?

What is MapReduce in a nutshell? What is filtering? in mathematic ways.

Conclusions [1]

2 Overview

The overview of the MapReduce, and the problems discussed in the paper: Connected components and minimum spanning trees; unweighted maximal matchings; maximum weighted matching; minimum edge cover and minimum cut.

3 Outline of the future work

It is very difficult to understand the paper without prior knowledge of the related field. It is first priority to fully understand the problems mentioned in the paper and gain more insight into the MapReduce framework.

1. more surveys into the problems discussed in the framework. 2. combining the findings in the problems itself, gain more knowledge and understanding in the MapReduce framework.

Timeline written here.

References

[1] Silvio Lattanzi, Benjamin Moseley, Siddharth Suri, and Sergei Vassilvitskii. Filtering: a method for solving graph problems in mapreduce. In *SPAA 2011: Proceedings of the 23rd Annual ACM Symposium on Parallelism in Algorithms and Architectures*, pages 85–94, 2011.