PREGEL: A SYSTEM FOR LARGE SCALE GRAPH PROCESSING

Connor Loughlin

24 November 2013

Grzegorz Malewicz, Matthew H. Austern, Aart J. C. Bik, James C. Dehnert, Ilan Horn, Naty Leiser, and Grzegorz Czajkowski

MAIN IDEA OF THIS PAPER

- Currently there is no good way to deal with large scale graphs and the data of those graphs
- Pregel is the best way to deal with large scale graphs with millions or billions of vertices that represent data sets and their connections
- Pregel's design and implementation makes it efficient

HOW THAT IDEA IS IMPLEMENTED

- One master machine, with many worker machines to work along with it
- All worker machines run in parallel, using supersteps to send, execute, and receive messages from other machines working on other groups of vertices, this allows effecieny
- Designed for google cluster architecture

MY ANALYSIS OF THAT IDEA AND IMPLEMENTATION

- Pregel is a very efficient large-scale graph system, the fact that it runs processes in parallel gets the job done faster.
- It does not run for any longer than it needs too, as soon as all vertices are inactive Pregel stops running
- Assigning certain groups of vertices to each worker machine distributes the workload more evenly, making Pregel more efficient

ADVANTAGES AND DISADVANTAGES

- ADVANTAGES:
- Processes graphs with billions of vertices efficiently
- There are already many applications that have been written for it
- uses parallel processing.
- DISADVANTAGES:
- No point in using Pregel on smaller scale graphs
- Limited memory space

REAL WORLD USE CASES

- Can be used to find shortest path from “A” to “B” on a large map very efficiently
- Works well with social networks seeing who is friends with who, and connections between people