Enabling Faster Convergence in Distributed Irregular Graph Processing

Abbas Mazloumi **Rajiv Gupta**

University of California, Riverside

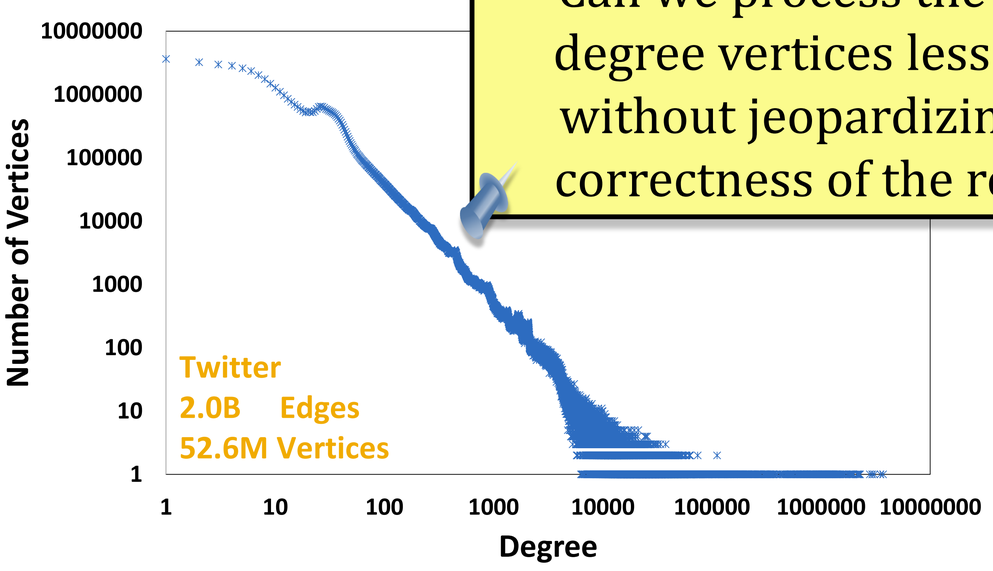
Objective

Degree distribution in real-world graphs is irregular and follows power-law distribution.

- The cost of processing high-degree vertices greatly exceeds the cost of processing low-degree vertices.
- This is observed in both computation and communication costs.

Major Challenge

Can we process the high-degree vertices less often without jeopardizing the correctness of the result?



Insight

- All incoming edges for each low-in-degree vertex are local to the machine on which the vertex resides.
- For balancing the computation, the incoming edges for each high-in-degree vertex are distributed across multiple machines.

DP: Differential Processing

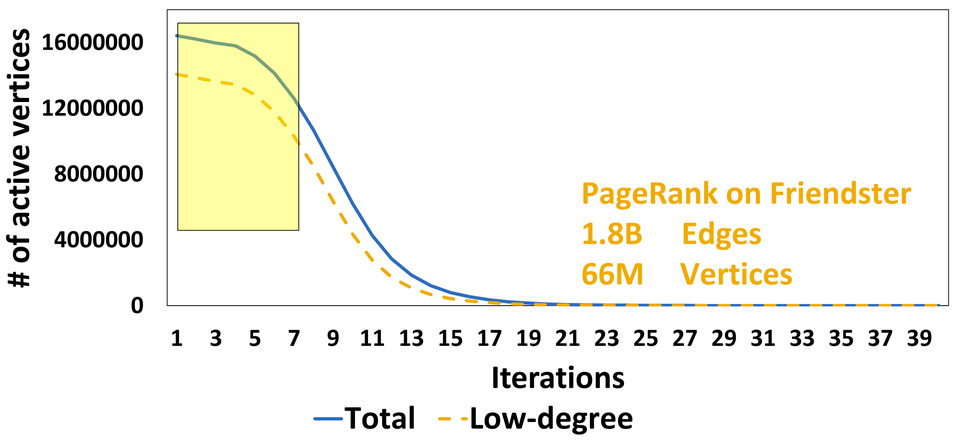
- Treats low- and high-degree vertices differently during each iteration in three steps.

1. low-degree vertices are processed multiple times as long as their values keep changing.
2. Multiple updates of low-degree vertices scatter latest values to high-degree vertices.
3. In the end, High-degree vertices are processed exactly once deploying all the changes.

- High-degree vertices are processed fewer times and also the iterative algorithm converges faster.

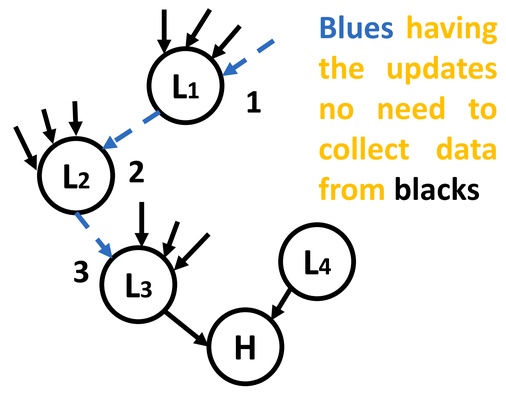
Turn off/on DP

- High number of low-degree vertices are active for participation in DP in order to deliver more changes to the high-degree vertices.
- The number of active high-degree vertices also need to be high enough to receive as many changes as possible.



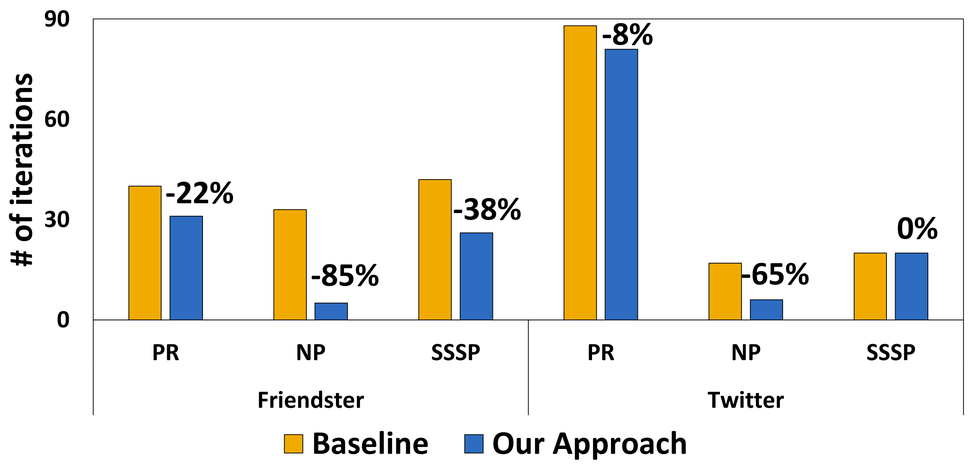
Delta Propagation

- Every vertex caches the gathered data when it is processed the first time and maintains this cached data for the rest of the execution.



- Caches are kept updated during the scatter.
- No need to collect the data from unchanged edges during the gather phase.

Number of Iterations



Execution Time (s)