PDSW-DISCS 2019: 4th Joint International Workshop on Parallel Data Storage & Data Intensive Scalable Computing Systems

Understanding Performance Bottleneck to Improve Parallel Efficiency of Louvain Algorithm

Naw Safrin Sattar

Shaikh Arifuzzaman

Big Data and Scalable Computing Research Group

THE UNIVERSITY of NEW ORLEANS

New Orleans, LA 70148 USA

Louvain Method for Community Detection

THE UNIVERSITY of NEW ORLEANS

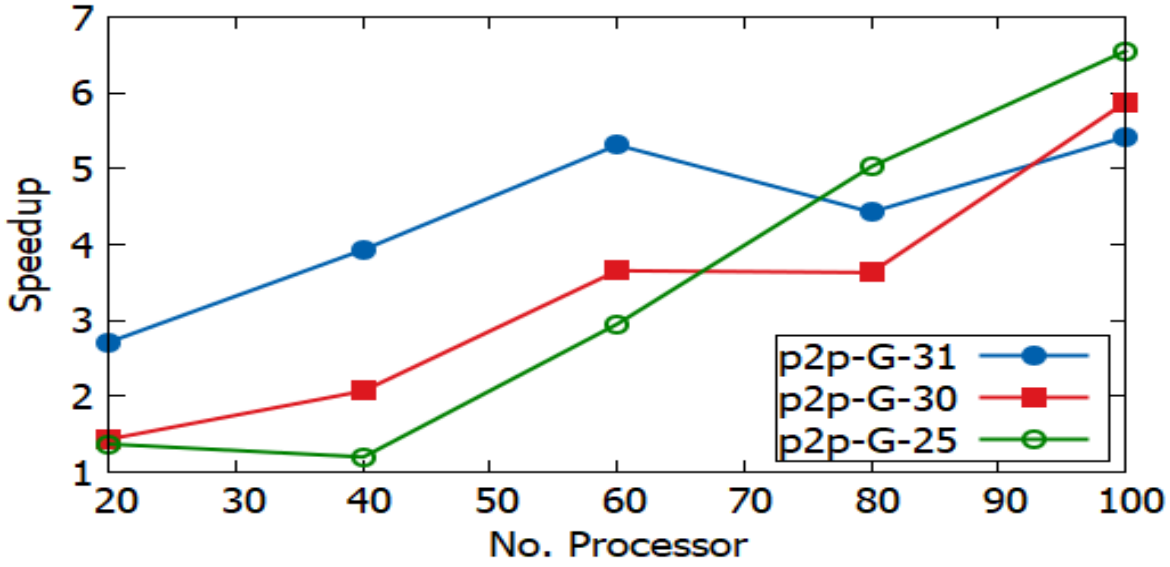
- Detects community based on modularity optimization
- One of the best methods in literature
  - Computation time and
  - Quality of the detected communities
- Reveals a hierarchy of communities at different scales
- Helps understanding the global functioning of a network

THE UNIVERSITY of NEW ORLEANS

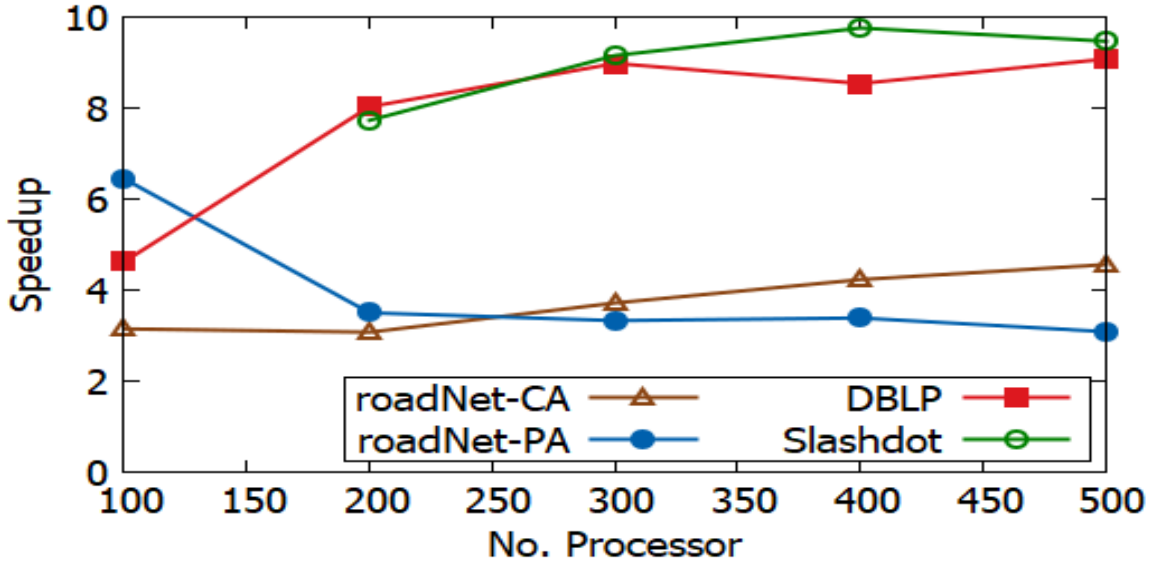
Motivation

- Existing scalable shared memory parallel Louvain
- Analyze the performance bottlenecks in distributed environment
- Scope of improvements in a hybrid parallel implementation

Speedup of our DPLAL (Distributed Parallel Louvain Algorithm with Load-balancing)

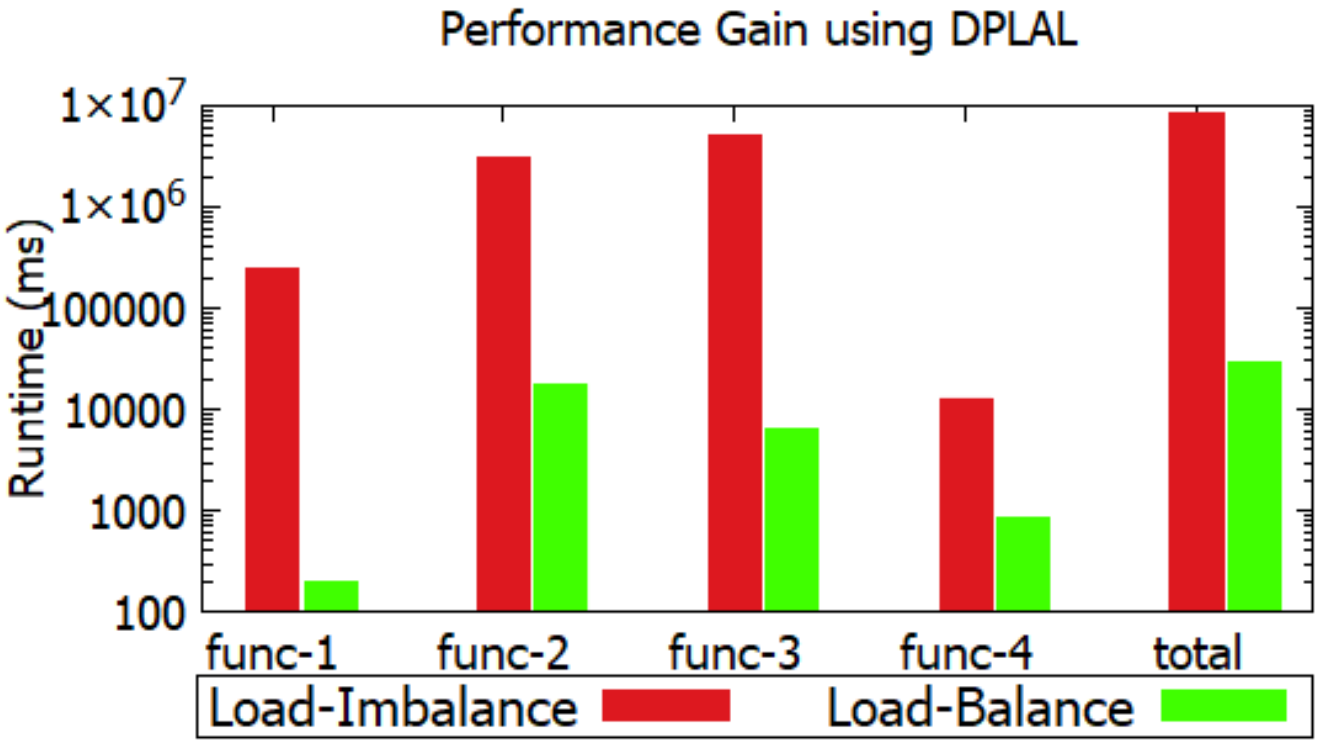


(a) relatively small graphs



(b) large graphs

Load-balancing with graph partitioner METIS

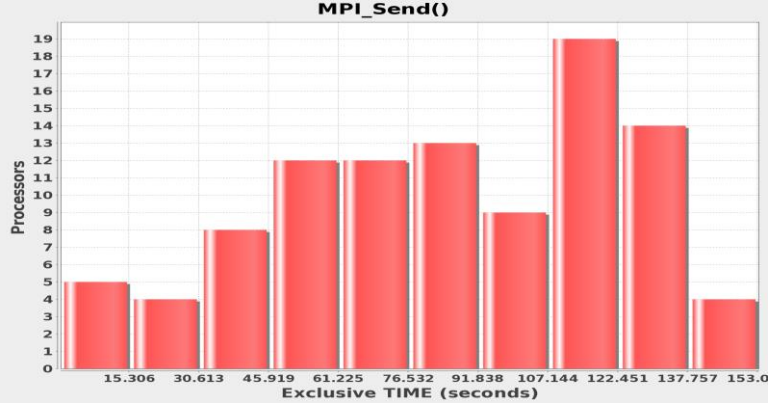


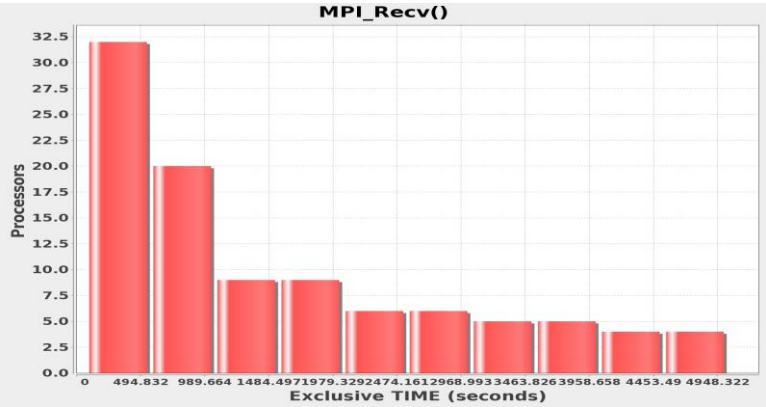
func-1:
gathering neighbour info
func-2:
exchanging updated community
func-3:
exchanging duality resolved community
func-4:
gathering updated communities

MPI profiling with TAU: Runtime of MPI Functions

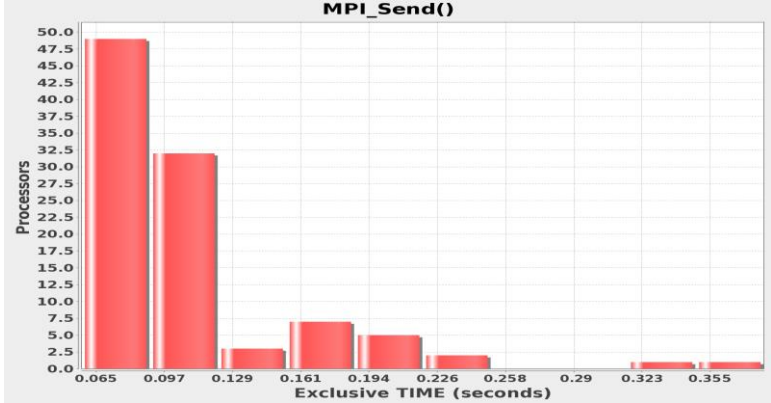
THE UNIVERSITY of NEW ORLEANS

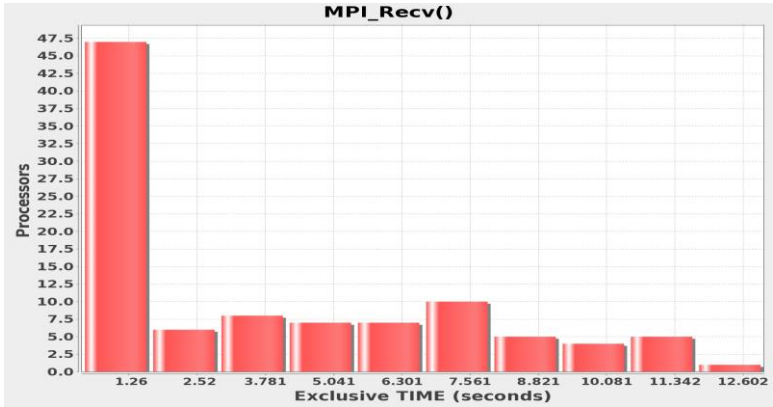
Load-Imbalanced





Load-balanced



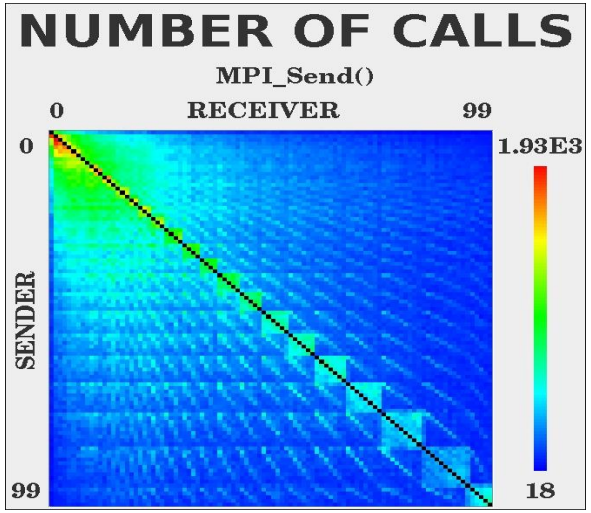


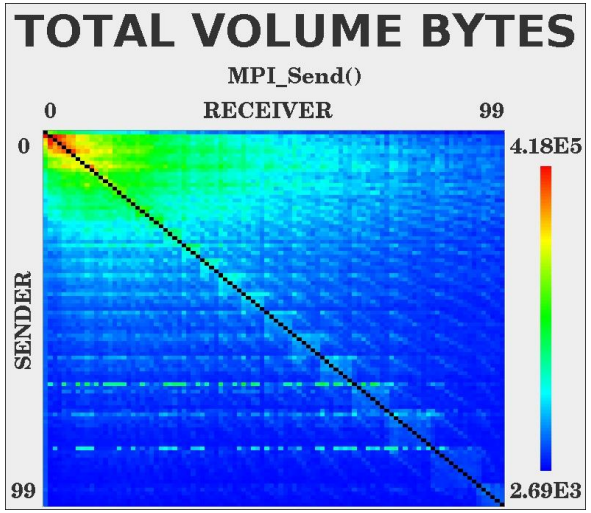
- 65% and 69% of the processors respectively, takes less than average time
- Load-Imbalanced MPI_Send, MPI_Recv, functions are 430.1x, 392.6x, slower than the balanced approach

MPI profiling with TAU: MPI Communications

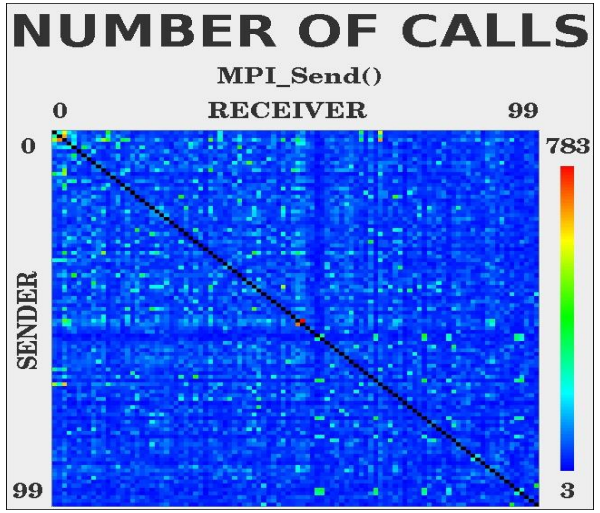
THE UNIVERSITY of NEW ORLEANS

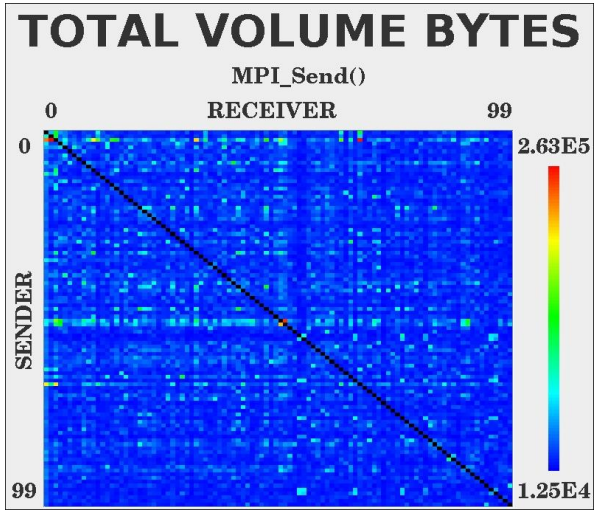
Load-Imbalanced





Load-balanced





THE UNIVERSITY of NEW ORLEANS

Future Works

- Profiling memory consumption
  - branching and cache access patterns,
  - time stalled waiting for resources (such as in memory reads), etc.
- Communication time at different phases of the algorithm
  - identify whether communication time overweighs computation time
  - change algorithm design accordingly
- Different graph-partitioning techniques for improved load-balancing and higher parallel efficiency
  - Hypergraph partitioning for social networks
- Parallel Louvain with data parallel computations in GPUs

THE UNIVERSITY of NEW ORLEANS

THANK YOU

Any Suggestions Appreciated

Contact: nsattar@uno.edu, smarifuz@uno.edu