2019 ECC Conference
June 9-11 at Marist College

Shared Memory Multicore PageRank Estimation

Piyush Garewal
Sindhu Kotapati
Tekenah Opuba
ADVISOR:
DR. APARICIO CARRANZA

Outline


Objective


Background Information


PageRank Algorithm


Parallel PageRank Algorithm


Result and Observation


Applications and future work


References

Objective

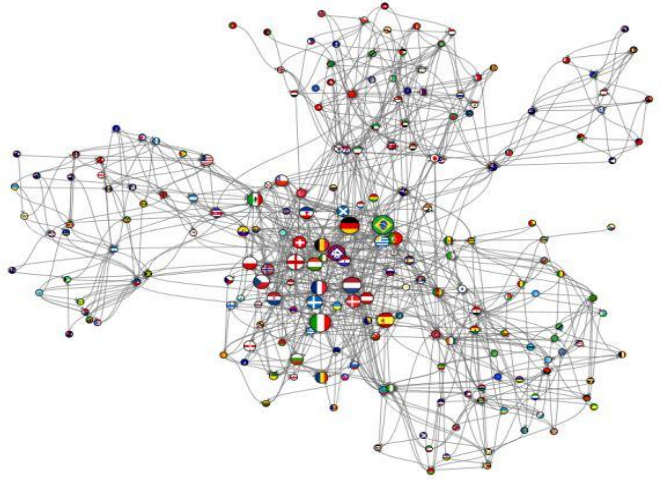
- To design, implement and test a multiprocessing parallel computation.
- PageRank Algorithm using OpenMP
- A random walk methodology using dataset from Stanford network
- Management of load balancing, synchronization
- Handling deadlocks
- Comparison with parallel algorithmic concepts using speedup, runtime and efficiency

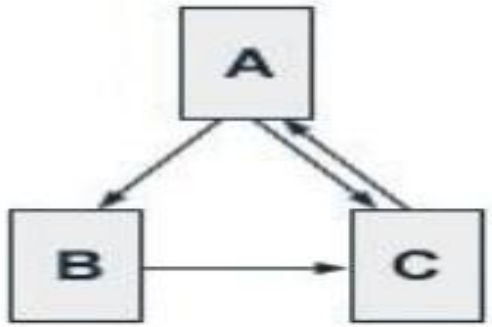
Background

The co-founders of Google, Larry Page and Sergey Brin innovated the algorithm the PageRank that took in the vast amount of data that is available on the Internet, organized it and ranked the sites in accordance to the number of links it receives, higher the rank the greater the importance

This algorithm is unparalleled in its design and accuracy in delivering accurate results to queries, and continues to be the leading innovation that sets Google apart from its competitors

THE CONCEPT BEHIND PAGERANK





Consider an imaginary web of 3 web pages.
And the inbound and outbound link structure is as shown in the figure. The calculations can be done by following method :

PR(A) = 0.5 + 0.5 PR(C)
= 0.5 + (0.5*1)
= 1

PR(B) = 0.5 + 0.5 (PR(A) / 2)
= 0.5 + 0.5 (1/2)
= 0.5 + (0.5 * 0.5)
= 0.5 + 0.25
= 0.75

PR(C) = 0.5 + 0.5 ((PR(A) / 2)+ PR (B))
= 0.5 + 0.5 (1/2 + 0.75)
= 0.5 + 0.5 (1.25)
= 0.5 + 0.625
= 1.125

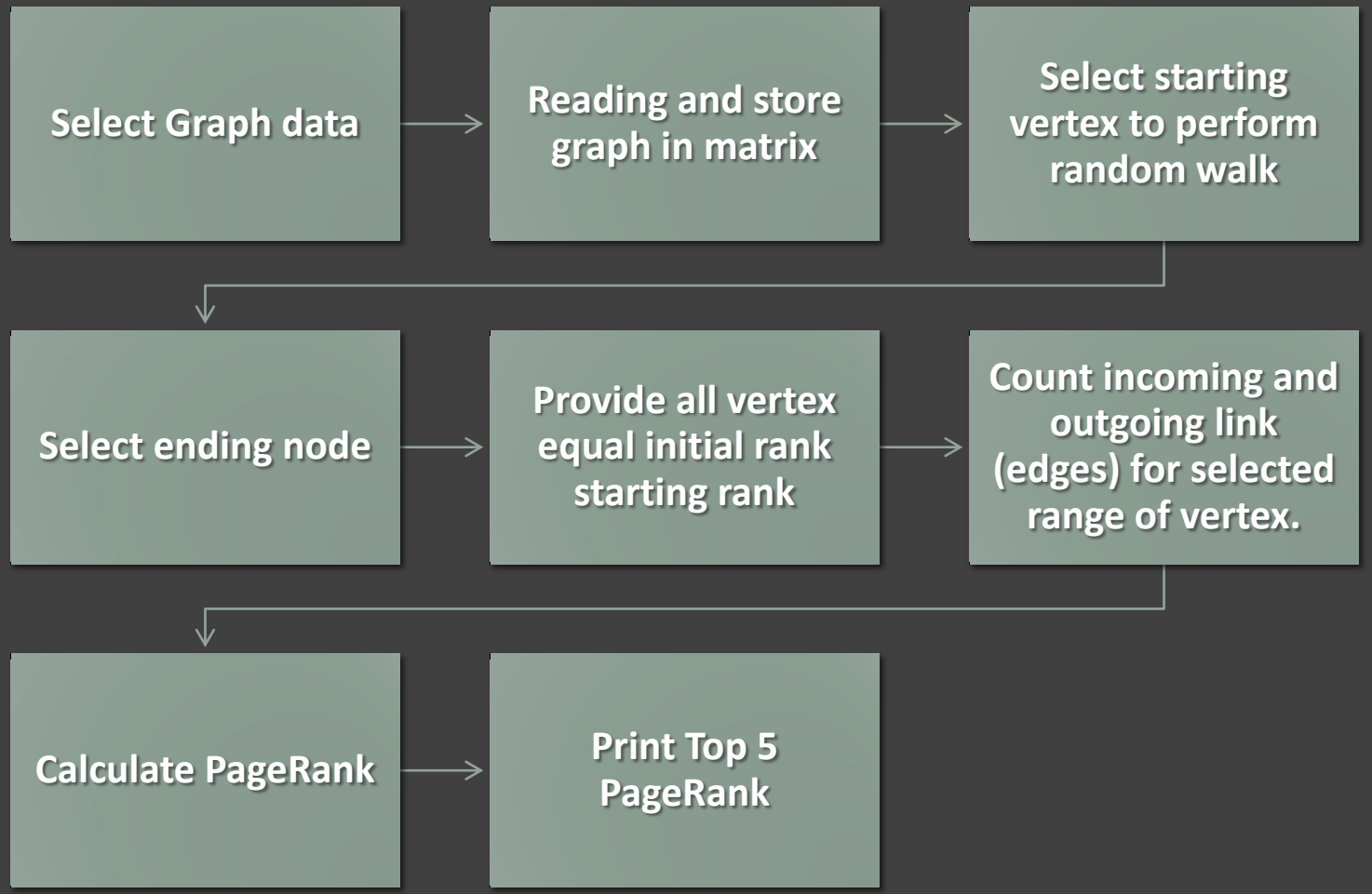
Detailed Approach

To design, implement and test a parallel estimator for PAGERANK of all nodes in a graph.

The PageRank of a node (v) in a graph is an indicator of the node's importance. For instance, a web graph, where nodes are web pages and edges are crosslinks (directional; incoming or outgoing) between two webpages.

In this context, the PageRank of a node is the relative importance of that webpage on the internet (higher the better).

The original Pagerank paper was written by Larry Page, Sergey Brin and others (@Stanford, and later Google)

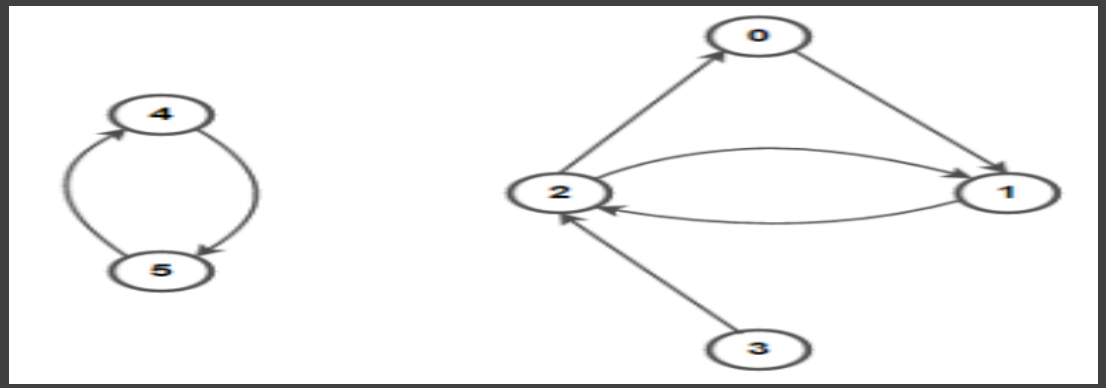


Parallel PageRank

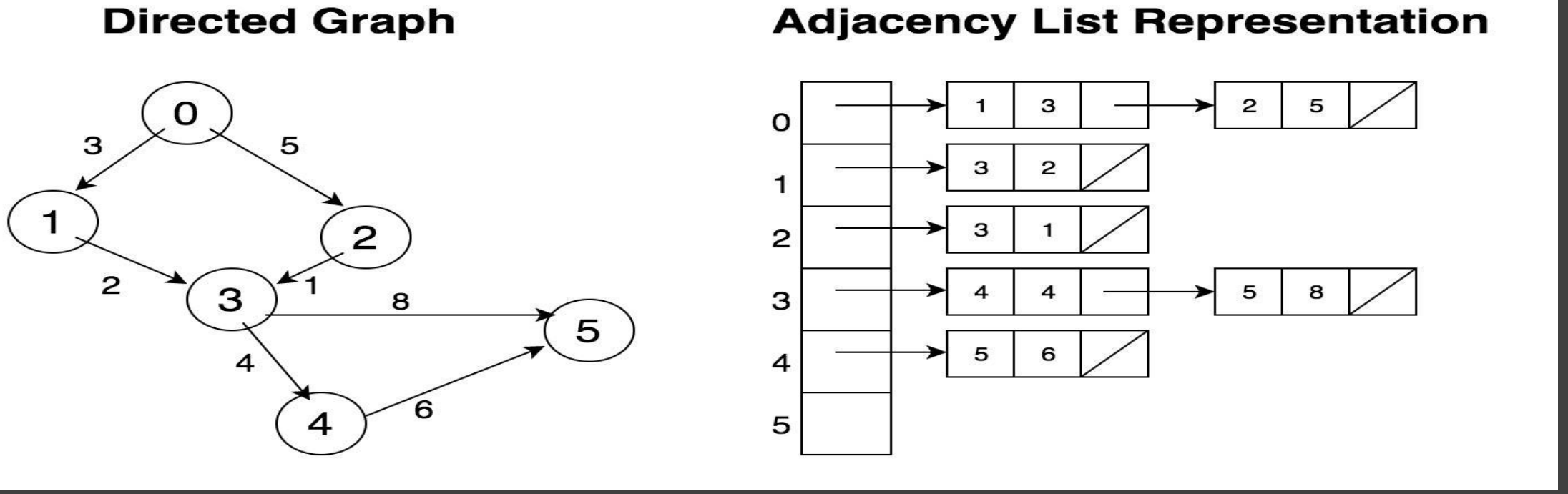
Selecting data (Graph)

Web graphs

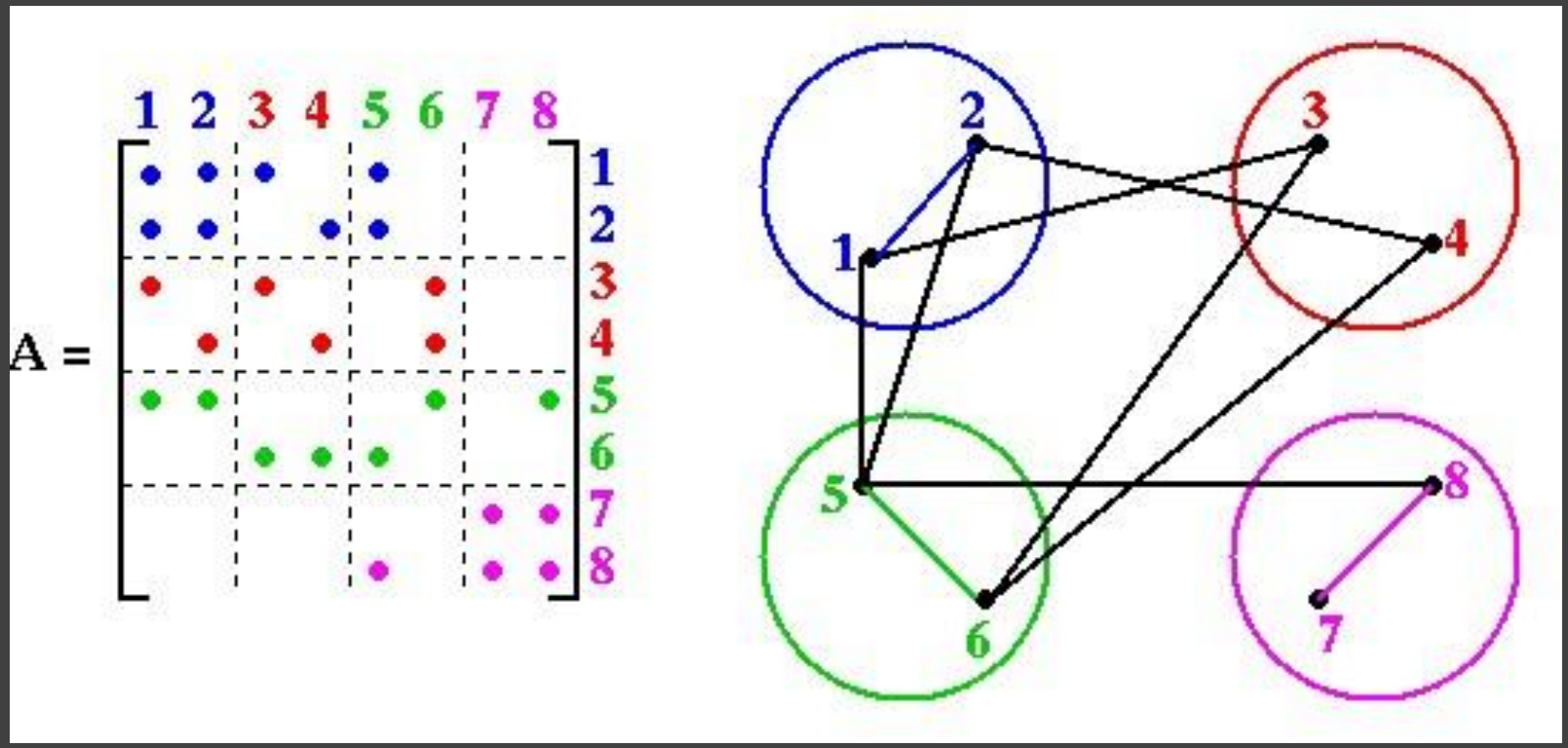
Name	Type	Nodes	Edges	Description
web-BerkStan	Directed	685,230	7,600,595	Web graph of Berkeley and Stanford
web-Google	Directed	875,713	5,105,039	Web graph from Google
web-NotreDame	Directed	325,729	1,497,134	Web graph of Notre Dame
web-Stanford	Directed	281,903	2,312,497	Web graph of Stanford.edu



Source: https://snap.stanford.edu/data/



Storing Data

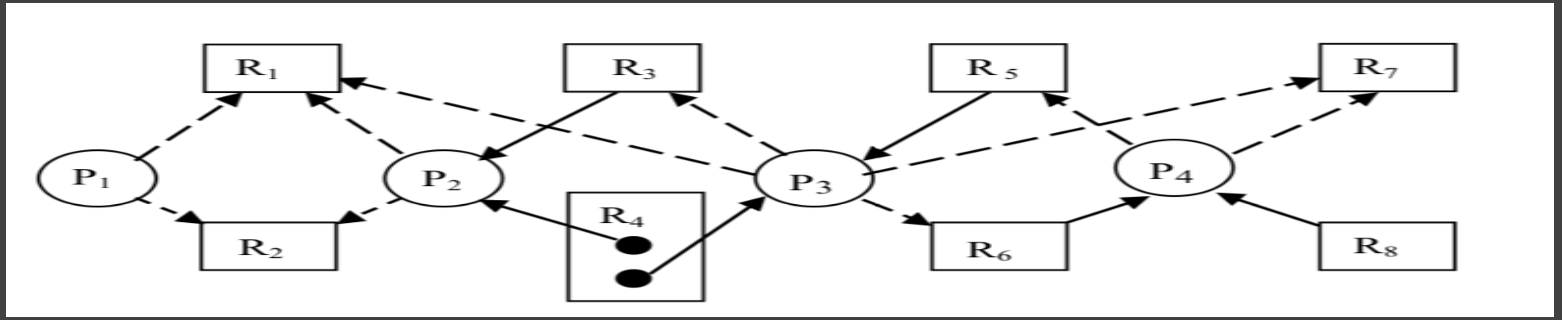


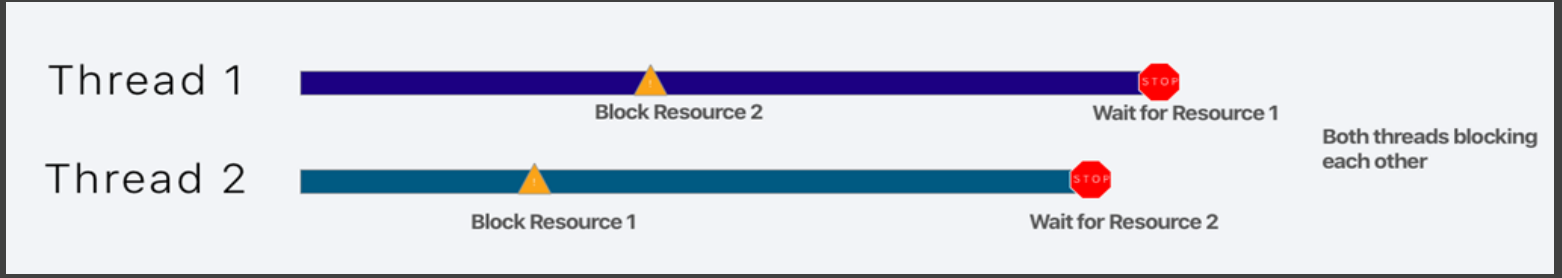
Partitioning the store graph (Load balance)

Key challenge

Avoiding deadlock (Race Case)

A race condition occurs when two or more threads can access shared data and they try to change it at the same time.



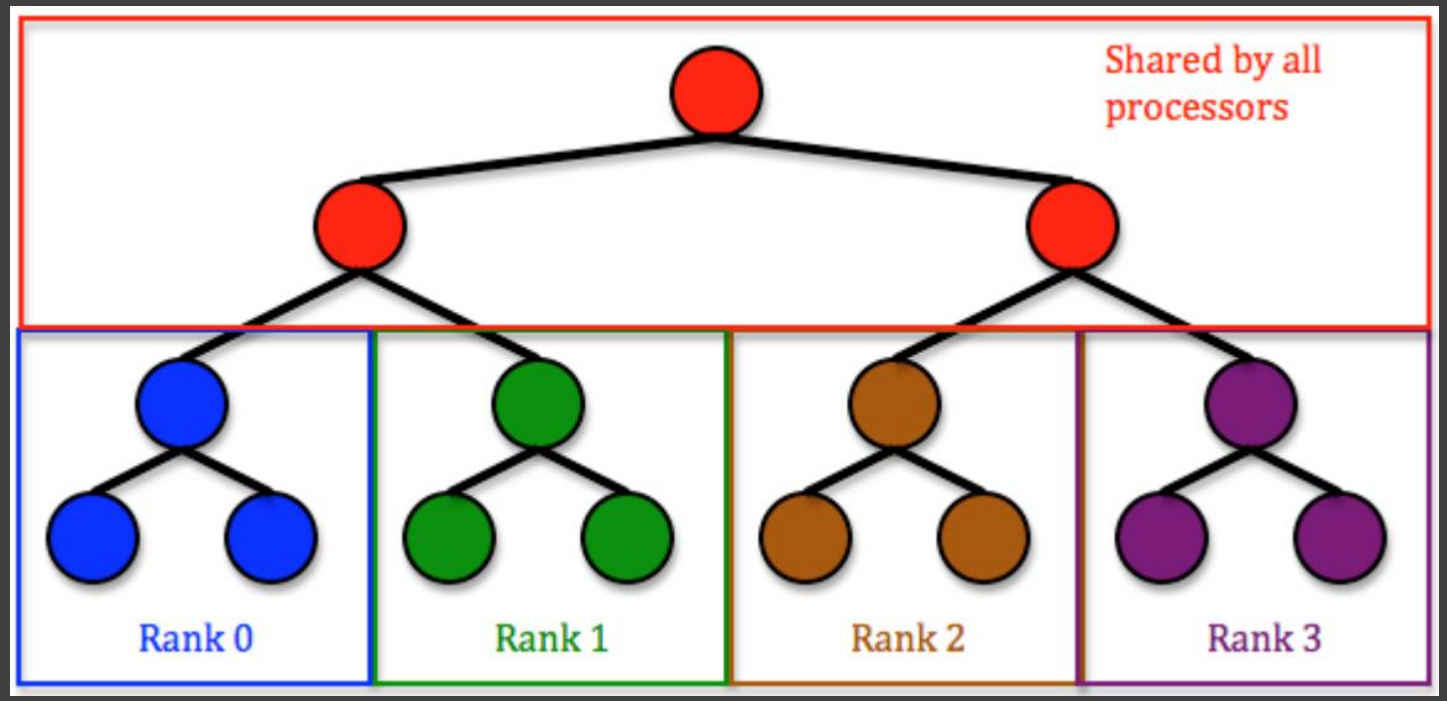


Key Challenges

Load Balancing: In parallel programming, the most important part is load balancing i.e.

- **All the processes have an approximately equal amount of task.**
- **In an efficient parallel implementation, all the process should complete the task approximately the same computation time.**
- **To achieve the proper load balance graph,**
- **We have stored the graph in the adjacent matrix but accessing the elements in the graph was not time efficient, then we tried using adjacency list, here all the elements/nodes can be divided equally into all the processes.**

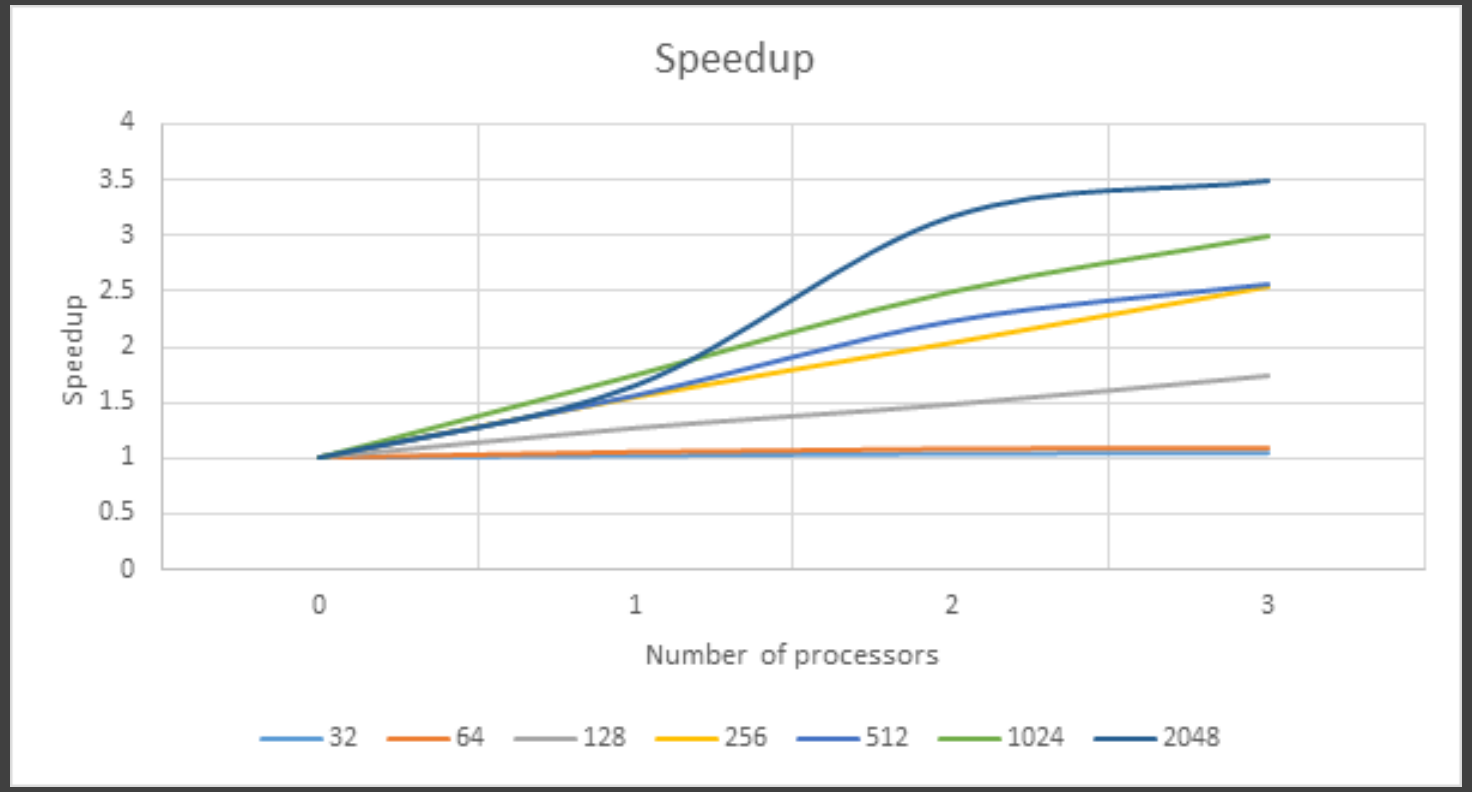
Load Balancing on graph



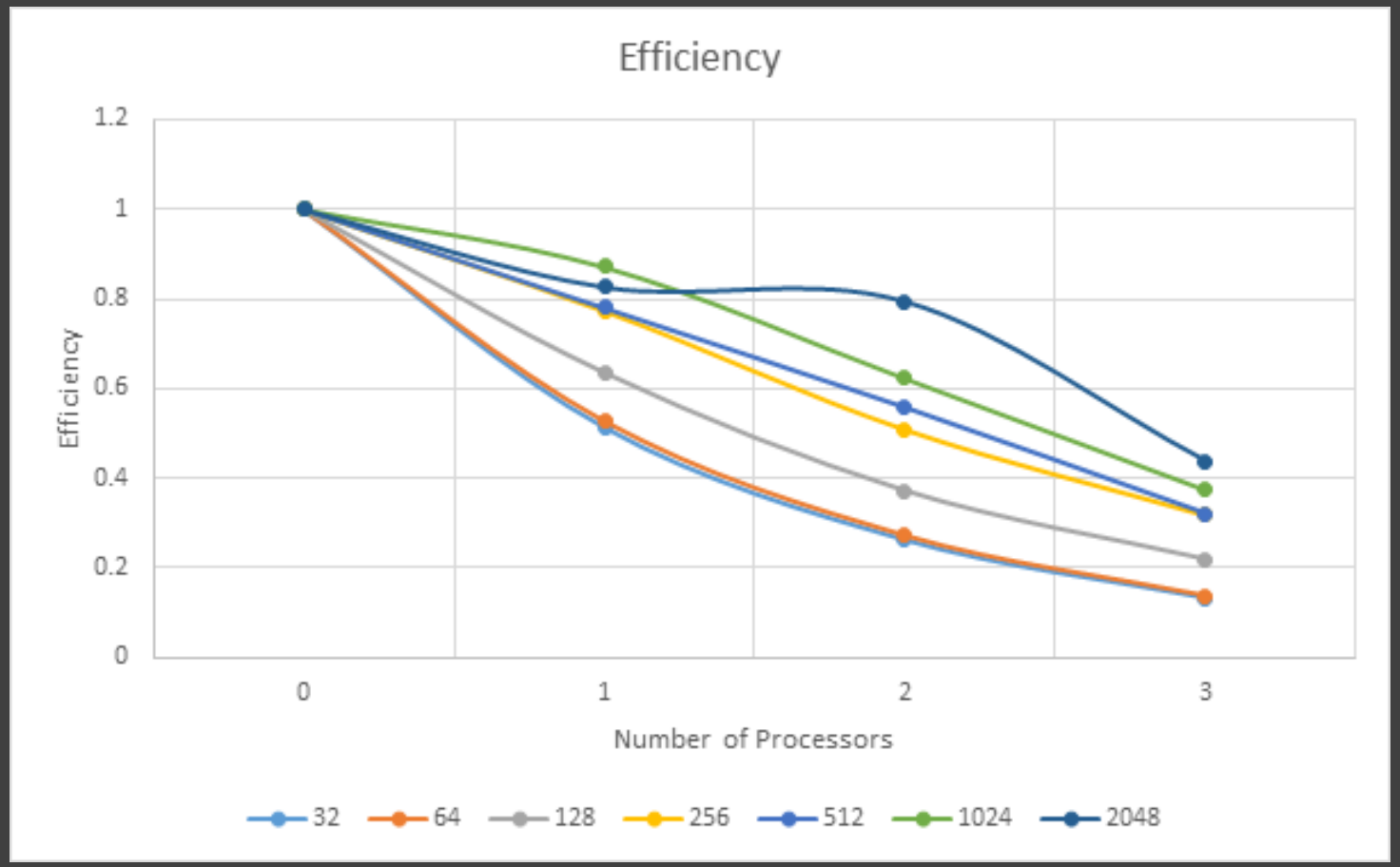
What is Speedup and Efficiency?

The speedup is defined as the ratio of the serial runtime of the best sequential algorithm for solving a problem to the time taken by the parallel algorithm to solve the same problem on p processors.

The efficiency is defined as the ratio of speed up to the number of processors. Efficiency measures the fraction of time for which a processor is usefully utilized.


Speedup
Speedup
4
3.5
3
2.5
2
1.5
1
0.5
0
0
1
2
3
Number of processors
32
64
128
256
512
1024
2048

Results


Efficiency
1.2
1
0.8
0.6
0.4
0.2
0
Efficiency
0
1
2
3
Number of Processors
32
64
128
256
512
1024
2048

Results

Applications of PageRank

Neuroscience- detect brain activity through examining changes in blood flow to the brain. PageRank can be used to generate a network in which the voxels or the subunits of the image of the brain are represented as nodes and the edges between the nodes represent the voxels that are time correlated. With this information, neuroscientist are able to conduct studies to assess the parts of the brain that have changed as the patient ages.

Toxic Waste Management- determine the position of water molecules which can provide an efficient method for waste management teams to remove nuclear waste/pollutants

Traffic Patterns- Each road is represented as a node on the graph, and each intersection where roads meet are the edges. Understanding traffic patterns can aid civil engineers in figure out how to improve overall traffic flow and prevent congestion.

References

[1] D. K. Sharma et al, “A Comparative Analysis of Web Page Ranking Algorithms,” International Journal on Computer Science and Engineering Vol. 02, No. 08, 2010.[Online] Available: http://inderscience.com

[2] R. Andersen, F. Chung, and K. Lang, “Local Graph Partitioning using PageRank Vectors,” 47th Annual IEEE Symposium on Foundations of Computer Science - 2006.

[3] S. Brin, and L. Page, “The Anatomy of a Large Scale Hypertextual Web Search Engine,” Proceedings of the 7th international conference on World Wide Web (WWW), 1998.

[4] Haotong Cao at el, “An Efficient Mapping Algorithm With Novel Node-Ranking Approach for Embedding Virtual Networks,” Access IEEE, vol. 5, pp. 22054-22066, 2017. [Online]. Available: http://ieeexplore.ieee.org

[5] S. Cheng and P. YunTao, “PageRank, HITS and Impact Factor for Journal Ranking,” WRI World Congress on Computer Science and Information Engineering-2009.

[6] “Networks,” More than just a Web Search algorithm: Google's PageRank in non-Internet contexts : Networks Course blog for INFO 2040/CS 2850/Econ 2040/SOC 2090. [Online]. Available: https://blogs.cornell.edu/info2040/2014/11/03/more-than-just-a-web-search-algorithm-googles-pagerank-in-non-internet-contexts/. [Accessed: 13-May-2019].

Question ?

Thank you for your time!