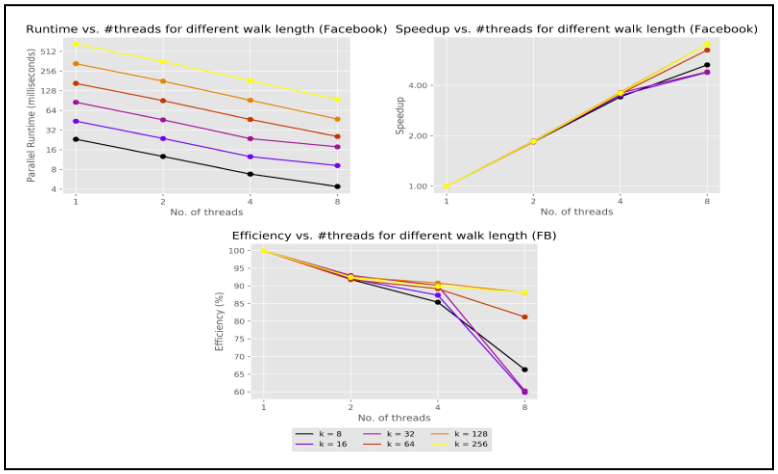
Fantastic 4(11): Rise of the Random Surfer

Reet Barik, School of EECS, Washington State University, Pullman WA (reet.barik@wsu.edu)



Proposed Approach

1. PageRank of a node is the number of times it is visited by a random surfer across many random walks (higher the value, more important it is).
2. Random walks are independent of each other and hence can be parallelized by multi-threading.
3. Random Walk Simulation: Given a node, next one is either a random neighbor or a random node in the graph (based on probability D).
4. PageRank value of each node is updated atomically by each thread.

Problem and challenges

PROBLEM STATEMENT:

Given a directed graph G(V,E) conduct random walks of length 'k' with a damping ratio of d and rank the nodes based on the number of times they were visited. (Input: Graph G, Walk-Length k, Damping ratio d. Output: List of all nodes sorted based on their pageranks).

SIGNIFICANCE:

1. Increasing 'k' increases computation, hence the need for parallelization.
2. Speedup will be of the order of number of threads.

CHALLENGES:

1. Efficient storage of the graph (Adjacency List to find neighbors easily).
2. Thread safe update of each node's pagerank.

Main Outcomes

Project's Main Outcomes:

1. Design choices: Static scheduling, atomic updates
2. Linear Speedup achieved
3. Runtime increases linearly w.r.t. 'k'.
4. Output stabilizes as 'k' increases.
5. Validity of output checked against established PageRank implementation in Networkx's library [1].

REFERENCES:

[1] Networkx link-analysis pagerank. https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.link_analysis.pagerank_alg.pagerank.html