COMP 6651: Algorithm Design Techniques
Fall 2017: Programming Assignment 3

Overall time complexity of Pagerank algorithm depends on

The running time of the Pagerank algorithm depends on three factors:

- number of iterations
- number of web pages (n)
- number of outgoing edges of each web page(On)

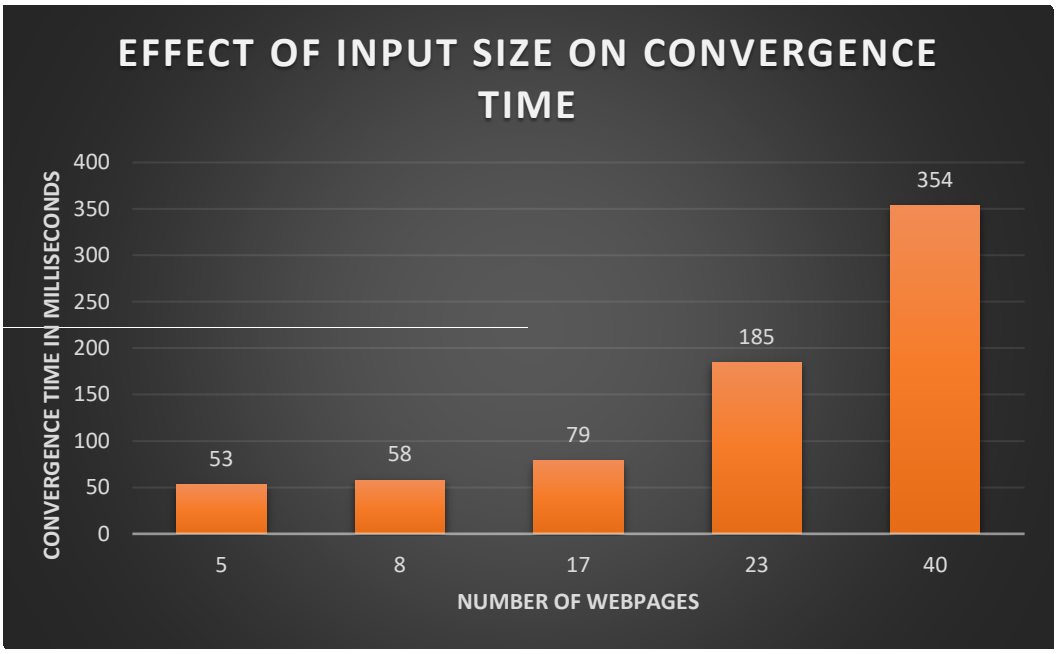
ANALYSIS OF PAGERANK ALGORITHM

- SIZE OF INPUT (WITH SAME SCALING FACTOR)
- SCALING FACTOR (WITH SAME INPUT SIZE)
- TYPE OF GRAPH (WITH SAME INPUT SIZE AND SCALING FACTOR)

SIZE OF INPUT (WITH SAME SCALING FACTOR)

Input size-NodeCount	Convergence time(ms)	Maximum iterations	Scaling factor
5	53	20	0.85
8	58	20	0.85
17	79	20	0.85
23	185	100(since large graph)	0.85
40	354	100(since large graph)	0.85

Based on the above analysis, as the size of the input graph i.e., the number of webpages increase, the convergence time also increases proportionally. Adding new webpages, increase the time complexity of the algorithm. Below is a pictorial representation of the same.

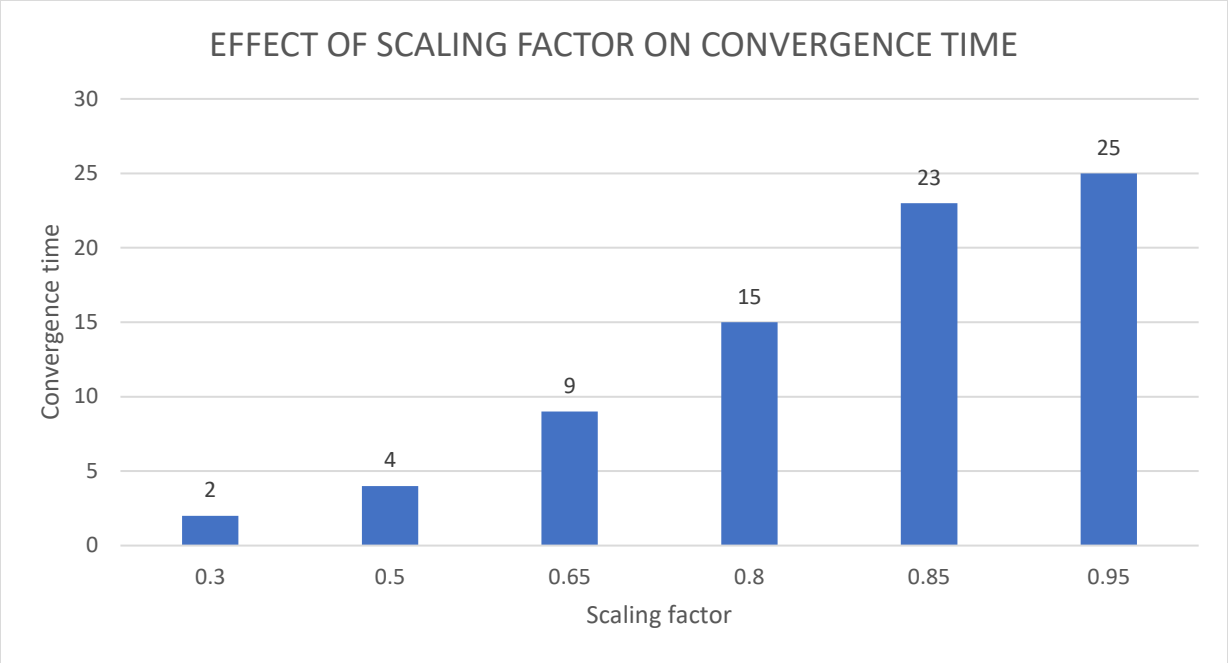


SCALING FACTOR (WITH SAME INPUT SIZE)

Scaling factor	Convergence time	Input size	Maximum iterations
0.60	4	5	20
0.65	8	5	20
0.80	15	5	20
0.85	23	5	20
0.95	25	5	20

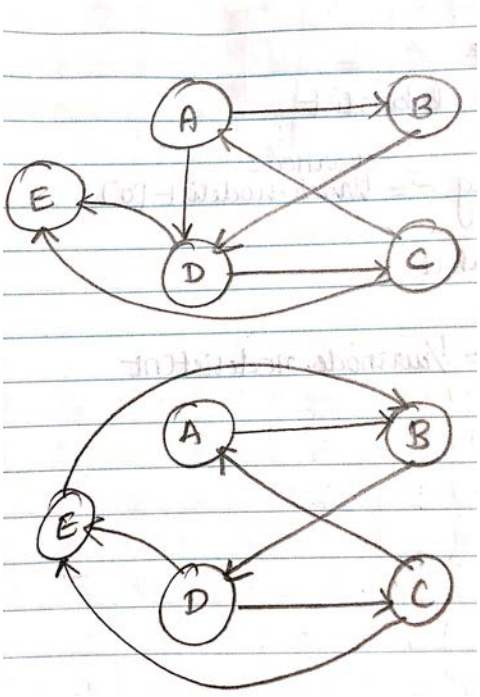
A low scaling factor (**much** damping involved) will make calculations easier. Since the flow of PageRank is dampened the iterations will quickly converge. Much damping means that the relative PageRank will be determined by PageRank received from external pages - rather than the internal link structure.

A high scaling factor (**little** damping) will result in the webpage's total PageRank higher and it takes time to compute.

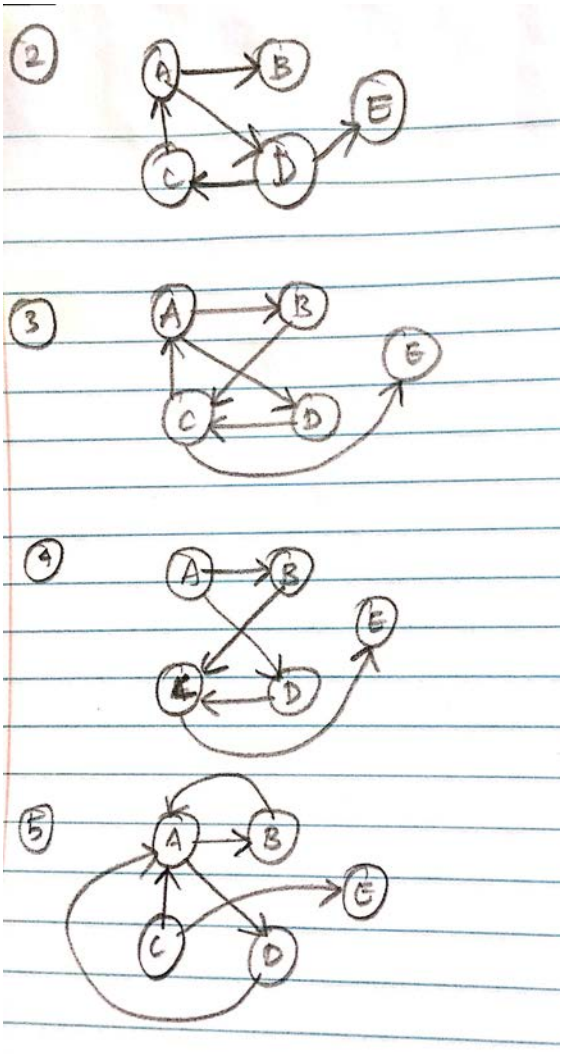


Based on the above analysis, as the scaling factor increases i.e., more dampening involved, the convergence time also increases proportionally.

TYPE I-2 MODELS OF SAME INPUT SIZE

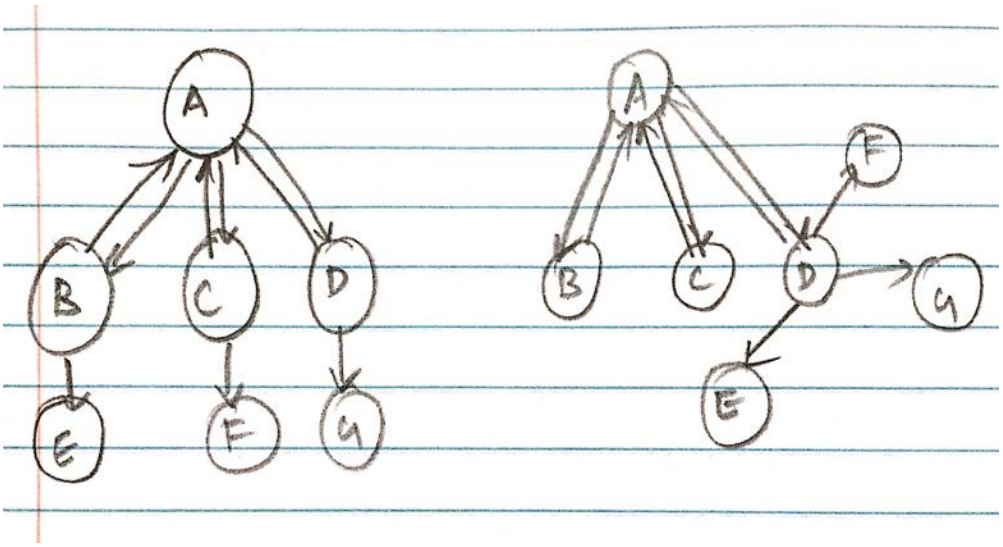


TYPE II-5 MODELS OF SAME INPUT SIZE



Based on the input models, the page ranks distribution follows a probabilistic distribution as explained below,

- Importance of a webpage ‘i’ as its pagerank-is the probability that a person surfing the internet follows the hyper links.
- In case 1, If user in Webpage A then he/she has ½ probability of going to webpage B or webpage D, and so the initial probability distribution or pagerank can be chosen as 1/N for each node. And based on its probability distribution, the pageranks vary in every iteration.
- Same follows in case 2, with 5 models of the same input graph’s size.
- Also, more the number of external outbound links, higher the pagerank values, therefore the best modelling of webpages with less external outbound links based on usage will have lesser
  Pagerank values. See the below testcase



Consider these graphs the graph 1 had lesser pagerank values when compared to second graph, increasing the external outbound links of one node, increases the overall pagerank distribution.

Therefore, the time complexity of pagerank algorithm mainly depends on the number of webpages and the external outbound linkage and number of iterations has very negligible effect on the complexity.