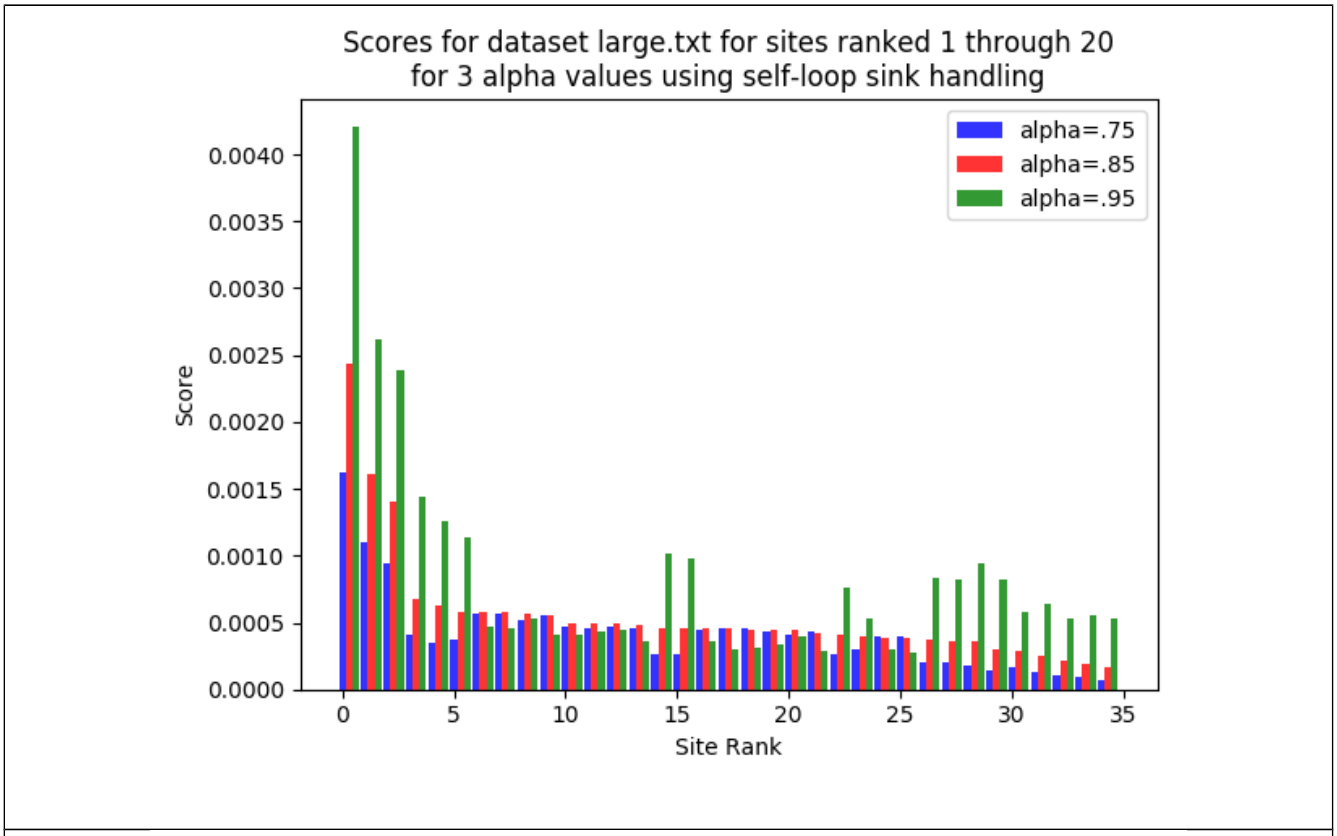
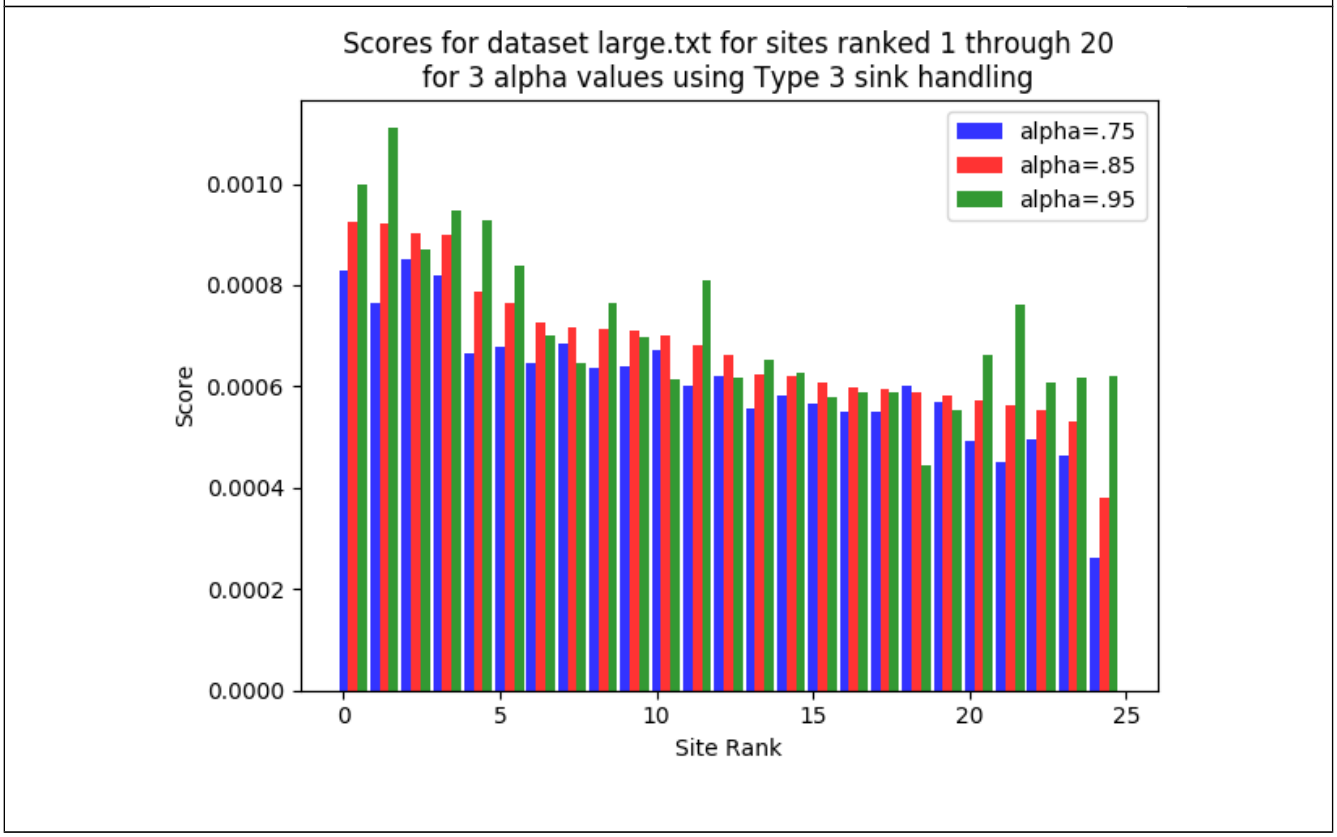
PageRank Project Report

Name : Magahet Mendiola GID: mmendiola3





PageRank Project Report

Name : Magahet Mendiola GID: mmendiola3

Plot Results

The plots show the combined set of top 20 page rank values for the three values of alpha. These page rank values represent the probability that random traversal of the weighted combined graph (web graph and complete graph) will end on a given vertex.

Alpha's Effect on the Model

Alpha models the probability that a random surfer will continue to follow outlinks from the current page. An alpha close to 0 means there is a high probability that the modeled surfer will stop following outlinks and instead transition to a completely random site on the graph. Conversely, an alpha near 1 means it is more likely the surfer will continue to follow outlinks.

The purpose of alpha in the algorithm is to transform the original web graph into one which is ergodic. As alpha moves closer to 1.0, this transformed graph moves closer to the original web graph. Alpha determines the mix between the original web graph and the complete graph (all vertices connected).

Alpha's Effect on PageRank Value

The pagerank algorithm uses the mixed model of the original web graph and the complete graph with alpha as the mixing factor between the two. As alpha changes, so does the graph the algorithm uses to determine rank values. The ranking value of a given node and the relative rankings between nodes can, and does, change as the combined graph changes.

Alpha's Effect on Runtime

Pagerank values propagate by updating each node's value based on contributions from every other node. As alpha moves lower, more weight is placed on the complete graph and thus pagerank values are propagated throughout the graph more quickly.

As an illustration, if we construct a linearly connected graph from nodes 0 to 10, it would take 10 iterations for node 0's pagerank to propagate to node 10 (with alpha = 1). On the same graph, if alpha = 0, all of node 0's pagerank value would propagate to node 10 in a single iteration. This demonstrates the effect the damping factor plays in the convergence of the pagerank algorithm.

Sink Strategy Effect on PageRank Value and Runtime

Empirically, self-loop sink handling increases the variance between site rankings and between ranking values by alpha. Intuitively, this is due to the model placing more transition probability on the sinks pointing back to themselves. This reduces the amount sinks distribute their pagerank to other nodes. The effect is that pagerank in sinks are higher in self loop handling.

The variance in rank between alphas can similarly be explained, as the combined graph (web and complete) with high alpha will place more transition probability on the self loops, again keeping value from propagating to the rest of the graph. Alpha values then have a greater effect on how much rank value is distributed out of sink nodes. Type 3 handling evenly distributes sink pagerank to the rest of the graph. This distribution of sink value isn't suppressed with higher alpha, so the variance of rank values between alphas for type 3 handling is lower.