PageRank Project Report
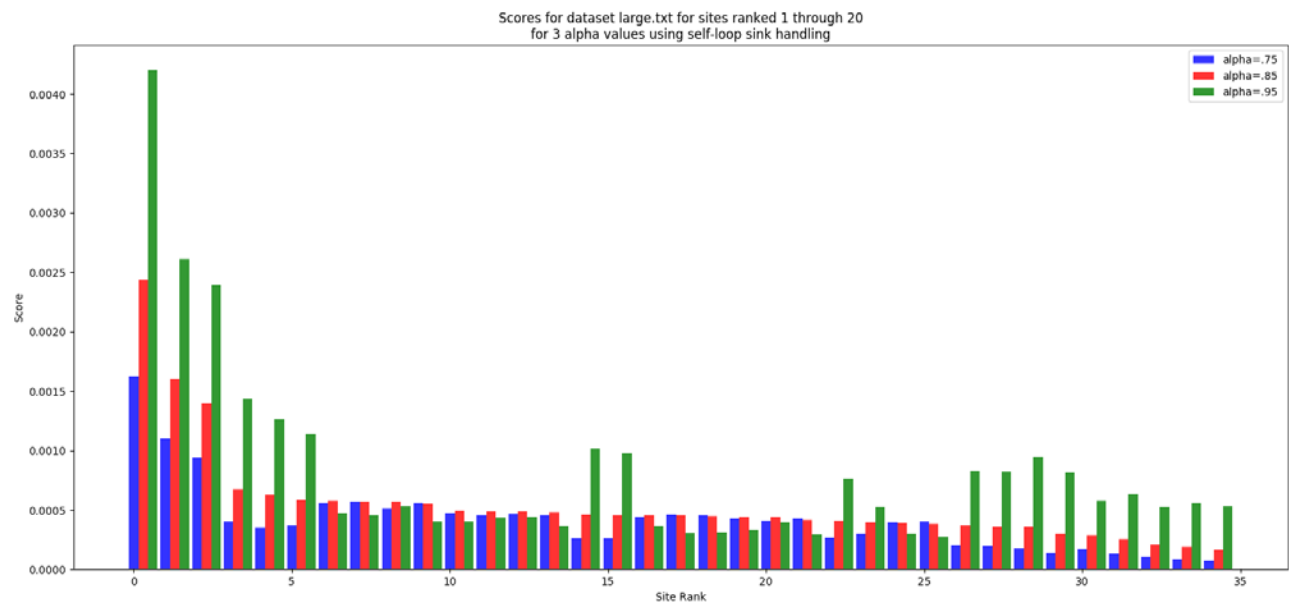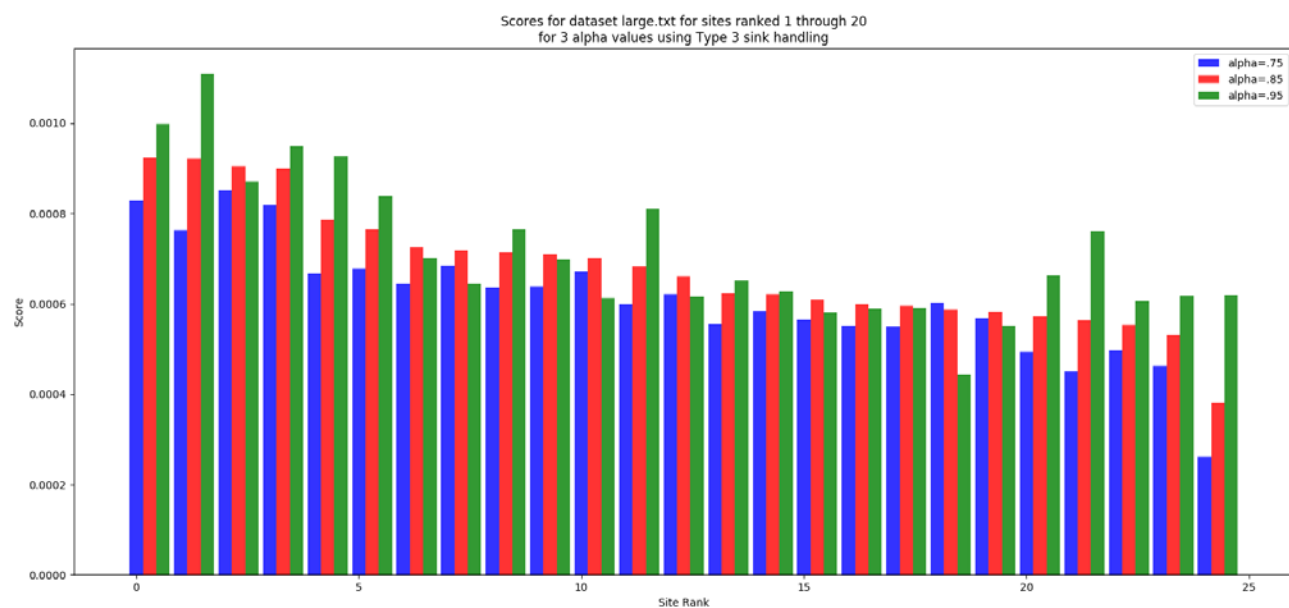
Name : Ayazhan Zhakhan

Graph of results for 3 alphas for self-loop sink handling

Scores for dataset large.txt for sites ranked 1 through 20
for 3 alpha values using self-loop sink handling

Graph of results for 3 alphas for Type 3 sink handling

Scores for dataset large.txt for sites ranked 1 through 20
for 3 alpha values using Type 3 sink handling

PageRank Project Report

Name : Ayazhan Zhakhan

Plots Discussion :

What are the results you see in the plots? What do the values represent? What does increasing alpha toward 1.0 mean? Why do the results change with changing alpha? Why does changing alpha affect the runtime? How does the sink handling strategy impact the results and the runtime?

Plots represent PageRank scores for the top 20 websites. From the plots, you can see that scores decrease with website ranks. Higher alpha values resulted in higher scores on average. Self-loop algorithm resulted in more extreme score distribution, where top 3 websites had especially high scores. Increasing alpha toward 1.0 increases website's scores. However, it also increases running time. The higher the alpha, the less random is the agent's behavior. Therefore, the scores distributed among websites highly depend on theirs interlinks, i.e. their hierarchical place in the web graph.
The higher the alpha, the less randomness is involved. This means that the agent has less options for exploration, which results in longer conversion time. Alpha if like an exploration/exploitation rate (in reinforcement learning language). The higher the alpha, the smaller is the exploration rate, therefore, the slower is the conversion time.
Self-loops had high scores for top 3 website and more uniformed scores for the rest. For alpha values of .75 and .95, score distribution was not monotonically decreasing, where websites with lower ranking sometimes had higher scores. For alpha of .85 the distribution was monotonically decreasing.
Type 3 sink handling resulted in a more uniformed score distribution with no extreme values. As with self-loops, alpha values of .75 and .95 resulted in non-monotonically decreasing score distribution, whereas alpha of .85 resulted in monotonically decreasing score distribution. Looking at these graphs, I can now understand why alpha value of .85 was chosen by Google.