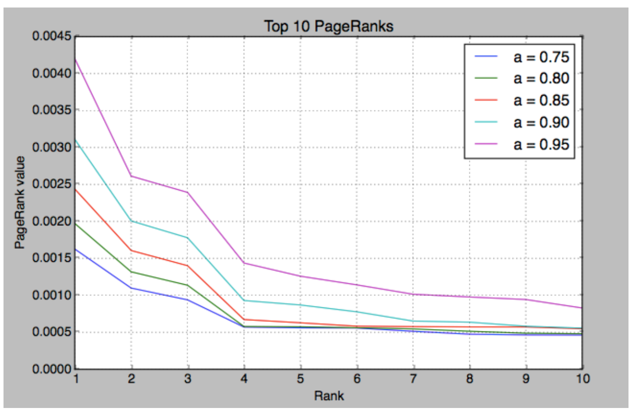
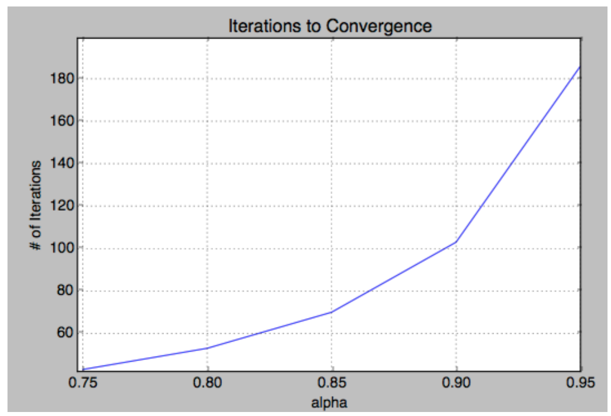
Karel Klein Cardeña

PageRank αlgorithm

The PageRank algorithm's resulting vector largely depends on the parameter α. Its complement, 1-α, represents the probability that a surfer will go to a webpage chosen at random from the N pages in the web graph. Interestingly, the top 10 pageranks all display exponential decay. This leads one to believe that the top sites are likely to remain the top sites, regardless of a change in α. Furthermore, a positive correlation is observed between the value of α and the pagerank of each site: higher α yields higher pagerank. This is seen from a different perspective in graph at the bottom-right. Here, we see that a higher α raises average pagerank exponentially (or at least non-linearly). However, this comes at a high price. As seen below, the number of iterations PageRank takes to converge is a function exponential in the α value. This suggests that while a higher α may increase the importance of a webpage, the computational cost quickly becomes intractable, especially for the internet's enormous web graph.





Let's look at the top 10 sites in terms of their number of incoming edges. The top 3 sites have unchanged number of in-edges for various alpha- corroborating the earlier proposition that top sites keep their 'importance' regardless of alpha. Interestingly, however, the farther down the rankings one goes, the more the plot bifurcates. This seems to suggest that the farther a site is from the top, the more susceptible its pagerank is to the damping factor. It is also worth noting that the most 'important' site does not necessarily have the most incoming hyperlinks. Given the tradeoff between pagerank and convergence time, it becomes apparent why an intermediary α such as 0.85 was chosen by PageRanks's creators.