CS6890: Fraud Analytics Using Predictive and Social Network Techniques

Trust Rank Algorithm

# Soumi Chakraborty

ES19BTECH11017

# Nandita Lakshmi Tunuguntla

CS19BTECH11051

# What is the TrustRank algorithm?

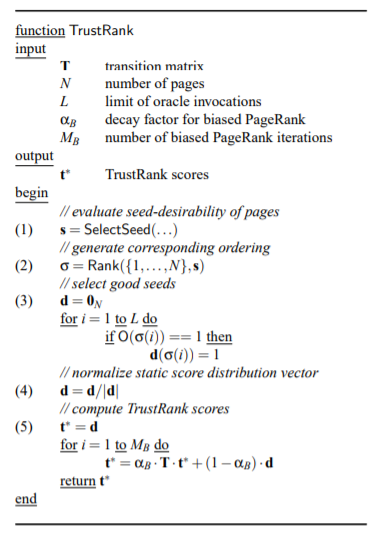
Many websites on the internet are spam websites which serve no other purpose than to drive traffic towards a particular website. While humans are capable of recognising when a website is spam, it is harder for a machine to be able to do so. The TrustRank algorithm was developed to do exactly this and it classifies a set of websites as spam or not spam.

## Working Principle

The working principle behind identifying whether a website is a spam one is that most legit websites will never point to a spam website at any point of time. The TrustRank algorithm tracks where the links in a particular website points to, and then creates a link-chain from one website to another. For a website to have a good score, none of its links in the link chain should ever point to a spam website. Another important component of this algorithm is the PageRank algorithm which is based on the fact that a website is considered to be important if several other websites point to it.

## The Algorithm

The pseudocode for the TrustRank algorithm is as follows, as taken from the original paper:



The algorithm requires a transition and inverse transition matrix which is built from a web graph. Then, the SelectSeed() function assigns scores to websites using the PageRank algorithm. Rank() arranges the websites in order of their scores after which the Oracle function is used to determine the vector *d*. The TrustRank is then determined using the iterative formula.

This algorithm very effectively manages to replicate human discretion by taking the whole labour intensive part out of the process.