**Hariom Mehta BIA-660 Assignment-4** [**hmehta8@stevens.edu**](mailto:hmehta8@stevens.edu) **CWID:10453223**

1. Code:

import networkx as nx

import matplotlib.pyplot as plt

g = nx.read\_edgelist('web-EPA.edges',create\_using=nx.Graph())

print (nx.info(g))

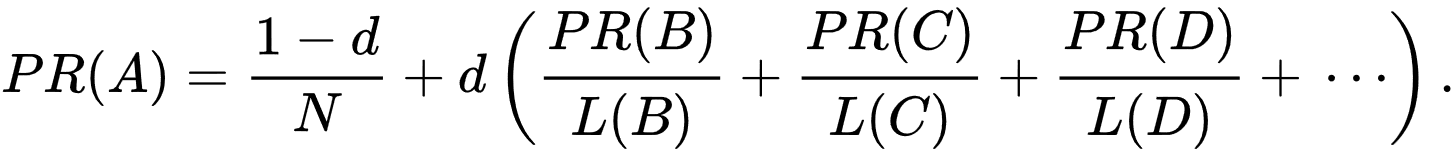
pr = nx.pagerank(g)

print(pr)

1. PageRank Algorithm:

PageRank is an iterative algorithm used by Google Search to rank web pages. It’s way of measurement to identify the importance of website pages. It measures the quality of pages and counts the number of links to determine the rank of page.

PageRank Formula is :



Here d is dumping factor, N total number of documents, L is referred to outbound links.

PageRank Algorithm outputs a probability distribution that a person clicking on the page randomly arrives on the web page. But here there is assumption that distribution is evenly divide among the all data in beginning of process. Algorithm goes through many iterations in order to get the appropriate PageRank values.

1. My Observation:

I used the data of web-EPA. There are 4271 nodes and 8909 edges. We are trying to identify the number by checking the number of links received by the other nodes. PageRank identify each node and rank them by observing the number of links received by node.