WSC Assignment 3 – Question Part 2

PageRank:

PageRank Introduction:

Developed by Google founders to measure the importance of webpages from the hyperlink network structure. PageRank assigns a score of importance for each node. The important nodes are those with many in links from important pages. Page rank can be used for any type of network, but its mainly useful for Directed networks. A node's page rank depends upon the page rank of the other nodes. This is a kind of circular definition.

In simple words, PageRank says that nodes who are central are the ones that, if you were to take a random walk on this network, then you would pass by them a lot of times or you would end up landing on them a lot of times.

The Basic PageRank of a node can be interpreted as the probability that a random walk lands on the node after k Random steps. The problem with the basic page rank algorithm is that in some networks few nodes can "suck up" all the page rank from the Network. To fix this problem, Scaled PageRank introduces a parameter Alpha such that a random walker chooses a random node to jump to with probability 1 - Alpha. Typically, alpha value varies between 0.8 and 0.9.

Hits Algorithm:

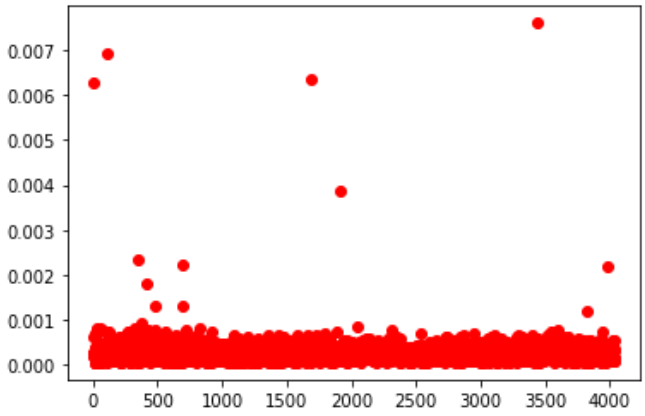
- The HITS algorithm starts by constructing a root set of relevant web pages and expanding it to a base set.
- HITS then assign an authority and hub score to each node in the network.
- Nodes that have incoming edges from **good hubs** are **good authorities**, and nodes that have outgoing edges to **good authorities** are **good hubs**.
- Authority and hub scores converge for most networks.

Important point to be noted here is, in order to understand what the HITS algorithm is saying, the Hub and authorities scores need to be looked together.

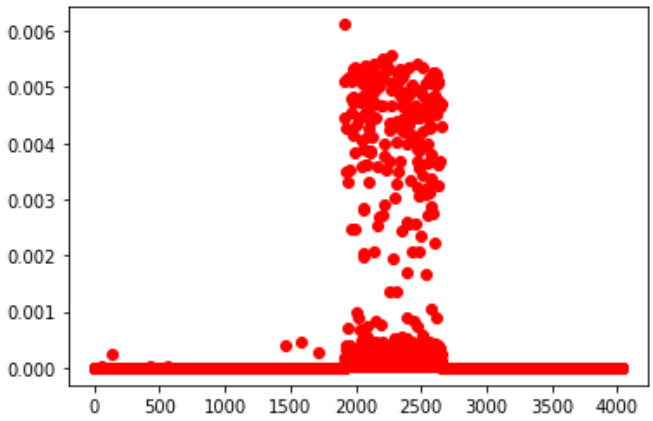
Analysis:

1. Medium sized Network

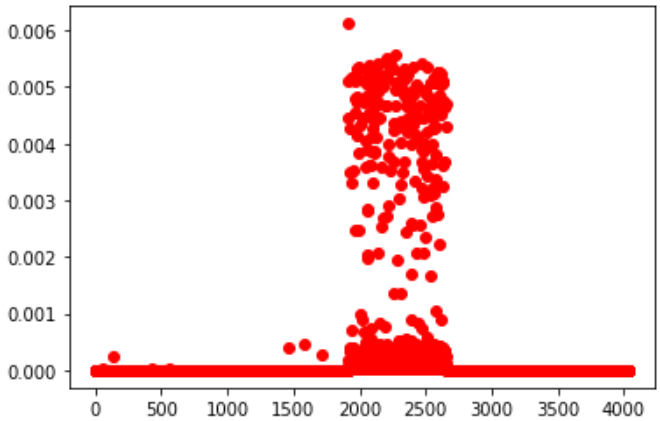
   a. PageRank Plot:



   b. Hubs Plot:



   c. Authorities:



Summary:

Ranks	Centralities				Page Rank		
	Degree	Close ness	Between ness	Eigen vector		Hubs	Hits
1	107	107	107	1912	3437	1912	1912
2	1684	58	1684	2266	107	2266	2266
3	1912	428	3437	2206	1684	2206	2206
4	3437	563	1912	2233	0	2233	2233
5	0	1684	1085	2464	1912	2464	2464

- It can be observed that no pair of centrality measures produces the exact same ranking of nodes, however they share some commonalities.
- Centrality measures make different assumptions about what it means to be a "central" node. Thus, they produce different rankings.
- The best centrality measure depends on the context of the network one is analysing.
- When identifying central nodes, it is usually best to use multiple centrality measures instead of relying on a single one.