**EXPERIMENT - 1**

**AIM**: Implementation of Page rank estimation

**THEORY**:

PageRank (PR) is an algorithm used by Google Search to rank websites in their search engine results. PageRank was named after Larry Page, one of the founders of Google. PageRank is a way of measuring the importance of website pages. According to Google:

*PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites.*

It is not the only algorithm used by Google to order search engine results, but it is the first algorithm that was used by the company, and it is the best-known.

The above centrality measure is not implemented for multi-graphs.

The formula for calculating PageRank is:



Where the PageRank value for a page u is dependent on the PageRank values for each page v contained in the set Bu (the set containing all pages linking to page u), divided by the number L(v) of links from page v. The algorithm involves a damping factor for the calculation of the PageRank.

**CODE:**

| import numpy as np  import networkx as nx  import requests  from bs4 import BeautifulSoup  def compute\_pagerank(url, num\_iterations=100, d=0.85):  response = requests.get(url)  soup = BeautifulSoup(response.content, "html.parser")  links = []  # First round  for link in soup.find\_all("a"):  href = link.get("href")  if href and href.startswith("http"):  links.append(href)  G = nx.DiGraph()  G.add\_nodes\_from(links)  # Here we go again  for i, u in enumerate(links):  response = requests.get(u)  soup = BeautifulSoup(response.content, "html.parser")  for link in soup.find\_all("a"):  href = link.get("href")  if href and href.startswith("http") and href in links:  G.add\_edge(u, href)  pr = nx.pagerank(G, alpha=d, max\_iter=num\_iterations)  return pr  pr = compute\_pagerank("https://www.team-bhp.com/")  for url, score in pr.items():  print(url, score) |
| --- |

**OUTPUT:**

| https://www.team-bhp.com/ 0.00782992767585498  https://www.team-bhp.com/forum/ 0.07463188563968107  https://www.team-bhp.com/hot-threads 0.06399958870407477  https://www.team-bhp.com/news 0.0395026543146141  https://www.cars24.com/buy-used-car?utm\_source=teambhp&utm\_medium=header\_nav\_bar&utm\_campaign=desktop 0.04748246777845908  https://oriparts.com/?utm\_source=team-bhp&utm\_medium=link&utm\_campaign=main&back\_url\_id=https%3A%2F%2Fboodmo.com%2f Catalog%2 Part-p-%7Bitem\_id%7D%2F&back\_url\_pn=https%3A%2F%2Fboodmo.com%2Fsearch%2F%7Bpn%7D%2F 0.04748246777845908 |
| --- |

**CONCLUSION**: PageRank algorithm is one of the main algorithms that links different websites over the internet, making the work of search engines highly efficient. Without such algorithms, the internet won’t be the way it is. In this experiment, we have implemented the PageRank algorithm using python.