# **Assignment 2**

### Authors:

|  |  |
| --- | --- |
| **Name** | **ID** |
| Sangi Manish Rao | 2015A7PS0235H |
| Tanuj Gupta | 2015A7PS0159H |
| Krishna Bharadwaj | 2015A7PS0076H |
| Ashrith Grandi | 2015A7PS0285H |

### Code Files:

1. make\_data.py
2. calculate\_rankmatrix.py
3. reset\_rankmatrix.py
4. make\_graph.py

### Working:

Described in README file

### Language:

Python

### Packages Used:

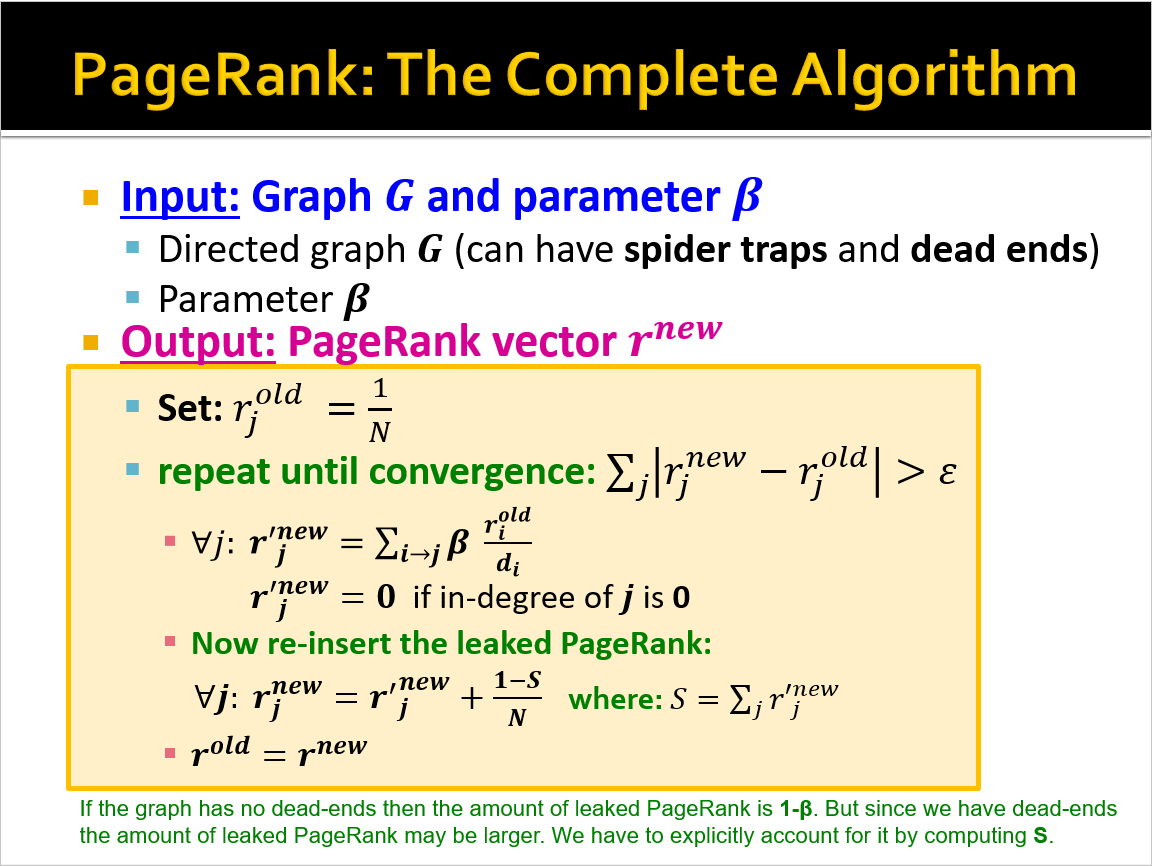
#### IGraph

#### Numpy

#### Sqlite3

### Project Description:

The algorithm implemented in this assignment was Page Rank. PageRank (PR) is an algorithm used by Google Search to rank websites in their search engine results.The algorithm used is described by the picture below.



### Running Time:

* 3.17 Seconds

### Implementations:

* Topic-Specific Page-Rank
* Implemented on very large Data Sets making use of sqlite database.

### 

### DataSet:

A sequence of snapshots of the Gnutella peer-to-peer file sharing network from August 2002. There are total of 9 snapshots of Gnutella network collected in August 2002. Nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts.

### Pros:

* Topic Specific PageRank has been implemented so choice is given to users to select the set to be used for Topic Specific PageRank and also the number of iterations.
* Can handle very large Graphs
* Visualization can be achieved
* Since database has been implemented extraction of values gets easier.

### Cons:

* Takes lot of time for large number of iterations
* Requires maintenance of a database