Lab 2: Link Analysis

1) Personalizing PageRank is an important real-world problem. Since different users find different pages relevant, search engines can provide better results if they tailor their page relevance scores to the users they are serving.

PageRank can be specialized with modifications of the teleport vector.

PageRank

$$r = \beta * \boldsymbol{M} * r + \frac{(1-\beta)}{n} * p$$

Where $\boldsymbol{M}$ is the stochastic adjacency matrix, $r$ is the PageRank vector, and $p$ is a stockastic vector (sum == 1) like *r* and represents the 'teleport' vector. Example:

$$p = [0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1]$$

In the case of the standard PageRank algorithm, $p$ has equal probabilities for a web page in the network.

For Personalized PageRank, *p* is a weighted set of web pages based on user interest. Example:

$$p = [0.1, 0, 0, 0.2, 0, 0, 0.5, 0, 0, 0.2]$$

For SimRank, *p* is a weighted set of web pages incorporating web page similarity metrics.

For random walk with restarts, *p*, would be all zeros except for one value equal to one to bring you back to class from TikTok.

$$p = [0, 0, 0, 0, \mathbf{1}, 0, 0, 0, 0, 0, 0]$$

In this question, we will explore how this can be applied to personalize the PageRank algorithm. Consider a graph of web pages connected by hyperlinks as described in lectures. Assume that people's interests are represented by a set of representative web pages. For example, if Chad is interested in sports and movies, then we could represent his interests with the set of pages {*www.espn.com*, *www.rottentomatoes.com*}. We will represent users with letters, e.g., A, and web pages with a number, e.g., 2.

Assume you have already computed the personalized PageRank vectors for the following users: 1. User A, whose interests are represented by the teleport set {1,2,3}
2. User B, whose interests are represented by the teleport set {3,4,5}
3. User C, whose interests are represented by the teleport set {1,4,5}
4. User D, whose interests are represented by the teleport set {1}.

Assume that the weights for each node in a teleport set are uniform.

a) Define a graph with nodes 1, 2, 3, 4, and 5. Add an arbitrary set of edges to your graph.
b) Define your initial PageRank vector, your stochastic adjacency matrix, and teleport vectors for each user.
c) Without looking at the graph or actually running the PageRank algorithm, can you compute the personalized PageRank vectors for the following users? If so, how? If not, why not and how would you make it work?

d) Compute the Personalized PageRank values for your graph. You can start with the code provided in the PageRank notebook. You can use the custom implementation or Networkx. If you are using Networkx check out the “personalized” parameter. Include a copy of your implementation and sample results.

2) Assume a fixed teleport parameter β. For the following questions, assume that the PageRank vector for user *i* is $v_i$. Express your answer in terms of *v* if you can. For example, PageRank for user A whose teleport set is {1,2,3} is $v_A$.

Express the flow equations for each node in the graph.

3) Provide a formulation for PageRank using vector/matrix notation that include personalized and relevance (similarity) information.