

Homework 5 for CMSC 498U/644

Due May 2

April 26, 2018

1 Problem 1

Consider the following graph as shown in Figure 1. The current probability distribution is as follows: the probability of being at point v_i is p_i . $p_1 = p_2 = p_3 = p_4 = 0.25$. Or say uniform probability

1.1 Part A

What will the probability distribution be if we simulate the naive random surfing process for one step? Two steps?

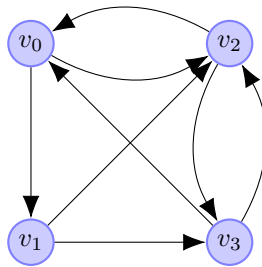


Figure 1: Pic 1

1.2 Part B

What will the probability distribution be if we simulate the random surfing processing with “taxation” for one step? Two steps? Please choose β to be 0.9.

2 Problem 2

Please find the densest subgraph in the graph as shown in Figure 2. Recall the definition of density of a subgraph G_S in a graph $G(V, E)$ is,

$$\frac{|\{(u, v) | (u, v) \in E, u, v \in S\}|}{|S|}$$

3 Problem 3

Please implement PageRank with the following iterative step

$$\vec{v}^j = \beta M \vec{v} + (1 - \beta) \vec{e}/n$$

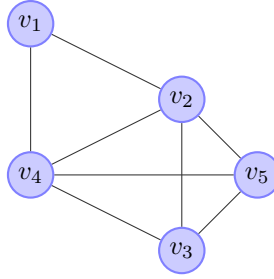


Figure 2: Pic 2

where $\beta = 0.85$. You can stop iteration if $\sum_{i=1}^n \text{abs}(v'_i - v_i) \leq n \cdot \epsilon$, where $\epsilon = 1.0 * 10^{-6}$.

Input will be provided as follows: the first line contains two integers n, m separated by a space. n is the number of vertices, and m is the number of edges. Vertices are labeled from 0 to $n - 1$. The remaining m lines are the m directed edges. Each line will consists two integers u, v separated by a single space, indicating an directed edge from u to v .

```
4 8
0 1
0 2
1 2
1 3
2 0
2 3
3 0
3 2
```

For output, each lines contains an id for a node, and its probability. We will manually check answers, so do not worry too much about output length and format of floating number. Sample output:

```
0: 0.2781243750596915
1: 0.15570219294056742
2: 0.3245609358176832
3: 0.24161249618205802
```

You can use your favourite language. We want to see the actual calculation, so do NOT use functions or libraries that calculate PageRank for you. You can use library to do matrix or vector calculations, of course. If you are not sure which language to use, here are some nice languages to try:

1. Python, with `numpy`, `scipy`. Here is a nice tutorial: [tutorial](#).
2. R, R is a free software environment for statistical computing and graphics. You are encouraged to try RStudio, which is a nice environment for R. You can find nice tutorial here: [tutorial](#).
3. MATLAB. The famous and powerful MATLAB, though commercial software.

4 Homework Extra

Recall Charikar's greedy algorithm on finding a dense subgraph:

Algorithm 1: Charikar's Greedy Algorithm

```
1  $S \leftarrow V$ ;  
2  $SOL = V$ ;  
3 while  $S \neq \emptyset$  do  
4    $v \leftarrow$  a vertex in  $S$  whose degree in  $G[S]$  (subgraph induced by  $S$ ) is minimized;  
5   Remove  $v$  from  $S$ ;  
6   if density of  $G[S]$  > density of  $G[SOL]$  then  
7      $SOL \leftarrow S$ ;  
8   end  
9 end  
10 return  $SOL$ 
```

Please find a graph that makes the maximum density algorithm fail to give the optimal solution. (It is possible to construct a graph where there exists a dense subgraph of density d , but Charikar's algorithm only outputs a subgraph of density roughly $\frac{d}{2}$.)