

Question 1.1, Homework 3, CS224W

~~Using outgoing and incoming BFS search for both Email and Epinions graphs for nodes 2018 and 224 respectively I determined the following:~~

- Node 2018 in Email graph has 52104 out nodes and 1 in node (which is a self loop) thus this node is in the OUT of this graph.
- Node 224 in Epinions graph has 47676 out nodes and 56459 in nodes. In addition, the intersection of the two sets size is 32223, thus, this node is in SCC of this graph.

Question 1.2, Homework 3, CS224W

From the desired figures we can learn about the IN, OUT and SCC of each of the Email and Epinions graphs.

Regarding the Email graph: around 20% of the nodes are in IN or SCC set while about 66% of the nodes are in OUT of SCC set. Regarding the Epinions graph: around 56% of the nodes are in IN or SCC while about 59% are in OUT or SCC.

Question 1.3, Homework 3, CS224W

Regarding the Email graph, since the SCC is 13% of the nodes, IN and OUT sets are 7% AND 53%, respectively. The disconnected (total nodes - number of nodes in max wcc) are 15% thus TENDRILS and TUBES are 12%.

Regarding the Epinions graph, since the SCC is 42.5%, the IN and OUT sets are 13.5% and 16.5% respectively. There are only 2 nodes in the DISCONNECTED set thus the TENDRILS and TUBES are 27.5%.

Question 1.4, Homework 3, CS224W

The probability a path exists between start node u and end node v :

$$\Pr(u \rightarrow v) = (\Pr(u \in SCC) + \Pr(u \in IN)) \cdot (\Pr(v \in SCC) + \Pr(v \in OUT)).$$

For the Email graph $\Pr(u \rightarrow v) = 0.2 \cdot 0.66 = 0.132$.

For the Epinions graph $\Pr(u \rightarrow v) = 0.56 \cdot 0.59 = 0.33$.

Question 2.1, Homework 3, CS224W

In order to calculate the personalized PageRank vectors for the given users we will use the equation from class:

$$r = \beta M \cdot r + (1 - \beta)t$$

where, t is the teleport vector. Thus, we can calculate a user u rank r_u if we can find a linear combination of the given users' ranks. In other words, if the equation $T \cdot (a, b, c, d)^T = t_u$ has solution than $r_u = R \cdot (a, b, c, d)^T$. Where, $T = (t_A; t_B; t_C; t_D)$ and R the same.

(i) Eloise: $t_E = (0, 1, 0, 0, 0)^T = 3t_A - 3t_B + 3t_C - 2t_D$.

(ii) Felicity: $t_F = (0, 0, 0, 0, 1)^T$ cannot be computed with a linear combination of the given users' teleport.

(iii) Glynnis: $t_G = (0.1, 0.2, 0.3, 0.2, 0.2)^T = 0.6t_A + 0.3t_B + 0.3t_C - 0.2t_D$.

where T is the transposed operation.

Question 2.2, Homework 3, CS224W

All the set of users with a solution to $t_u = T \cdot a$, where, as before, $T = (t_1; t_2; \dots t_n)$, $a = (a_1, a_2, \dots a_n)^T$ is vector of parameters and t_u is the teleport vector of user u . Then, $r_u = R \cdot a$.

We need to prove the following:

$$\mathbf{r} = \beta \mathbf{M} \mathbf{r} + \frac{1 - \beta}{N} \mathbf{1} \tag{1}$$

As we've seen in class:

$$\mathbf{r} = \mathbf{A} \mathbf{r} = \beta \mathbf{M} \mathbf{r} + \frac{1 - \beta}{N} \mathbf{1} \mathbf{1}^T \mathbf{r}. \tag{2}$$

Since, $\sum_i r_i = \mathbf{1}^T \mathbf{r} = 1$ then, $\mathbf{1} \mathbf{1}^T \mathbf{r} = \mathbf{1}$. So, equation (2) becomes (1).

Information sheet

CS224W: Machine Learning with Graphs

Assignment Submission Fill in and include this information sheet with each of your assignments. This page should be the last page of your submission. Assignments are due at 11:59pm and are always due on a Thursday. All students (SCPD and non-SCPD) must submit their homework via GradeScope (<http://www.gradescope.com>). Students can typeset or scan their homework. Make sure that you answer each (sub-)question on a separate page. That is, one answer per page regardless of the answer length. Students also need to upload their code on Gradescope. Make sure to upload all of your code as .py files.

Late Homework Policy Each student will have a total of *two* late periods. *Homework are due on Thursdays at 11:59pm PT and one late period expires on the following Monday at 11:59pm PT.* Only one late period may be used for an assignment. Any homework received after 11:59pm PT on the Monday following the homework due date will receive no credit. Once these late periods are exhausted, any assignments turned in late will receive no credit.

Honor Code We strongly encourage students to form study groups. Students may discuss and work on homework problems in groups. However, each student must write down their solutions independently, i.e., each student must understand the solution well enough in order to reconstruct it by him/herself. Students should clearly mention the names of all the other students who were part of their discussion group. Using code or solutions obtained from the web (GitHub/Google/previous year's solutions etc.) is considered an honor code violation. We check all the submissions for plagiarism. We take the honor code very seriously and expect students to do the same.

Your name: _____

Email: _____ **SUID:** _____

Discussion Group: _____

I acknowledge and accept the Honor Code.

(Signed) _____