

Question 1.1, Homework 3, CS224W

Total nodes of email: 265214

IN of 2018 in email: 1

OUT of 2018 in email: 52104

Total nodes of epinions: 75879

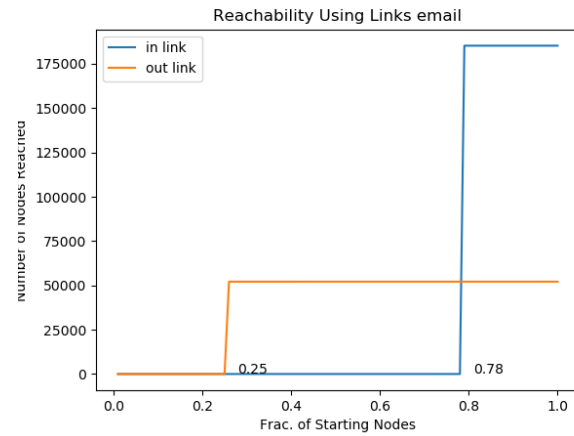
IN of 224 in epinions: 56459

OUT of 224 in epinions: 47676

According to the definition, node 2018 of email lies in IN and node 224 of epinions lies in SCC.

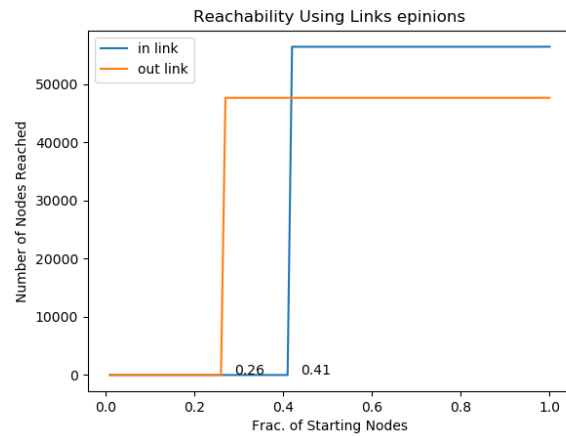
Question 1.2, Homework 3, CS224W

Plot for email is as below:



From the picture we can roughly estimate that relative sizes of SCC, IN and OUT are 16.5%, 58.5%, 5.5%.

Plot for epinions is as below:



From the picture we can roughly estimate that relative sizes of SCC, IN and OUT are 43.66%, 30.34%, 15.34%.

Firstly, we can use $(\text{total_nodes} - \text{MxWcc_size})$ and MxScc_size to calculate the sizes of DISCONNECT and SCC. Then starting from any node in SCC, do one forward BFS to calculate the size of OUT and one backward BFS to calculate the size of IN. The left is $\text{TENDRILS} + \text{TUBES}(\text{TT})$.

The results are as follows:

Total nodes of email: 265214

Size of MxWcc in email: 224832

Size of MxScc in email: 34203

Size of IN in email: 151023

Size of OUT in email: 17900

Size of TT in email: 21706

Size of DISCONNECT in email: 40382

Total nodes of epinions: 75879

Size of MxWcc in epinions: 75877

Size of MxScc in epinions: 32223

Size of IN in epinions: 24236

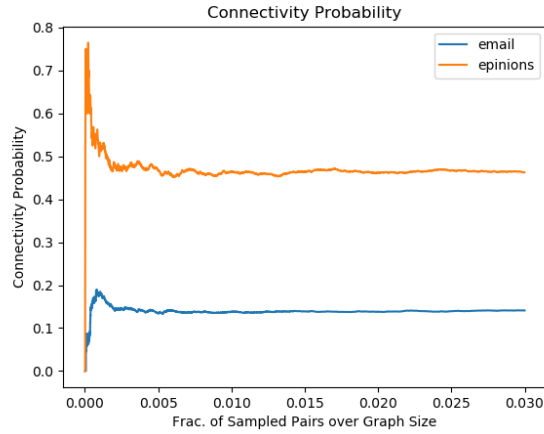
Size of OUT in epinions: 15453

Size of TT in epinions: 3965

Size of DISCONNECT in epinions: 2

Question 1.4, Homework 3, CS224W

The sampled probability plots are as follows:



We can see that both of email and epinions converge to 14% and 46%.

If we neglect the other parts, the probability will be converged to

$$p = \frac{size_{IN} + size_{SCC}}{size_{IN} + size_{SCC} + size_{OUT}} * \frac{size_{OUT} + size_{SCC}}{size_{IN} + size_{SCC} + size_{OUT}}$$

which is higher than (23%, 52% respectively for email and epinions), since other parts still take up a certain number of percentage of the graph.

Question 2.1, Homework 3, CS224W

Yes. We can compute the probabilities for these guys. Just set the restart point of random walk to be the points in the corresponding teleport set (uniformly or according to the weights).

Question 2.2, Homework 3, CS224W

As long as the choosing of restart points of the random walk is uniform, we can compute personalized pagerank vectors of whose teleport set can be added (or minus) by others in V .

The calculation procedure can be written as:

$$\begin{aligned} r &= Mr \\ &= \beta Mr + \frac{1-\beta}{N} \mathbf{1}^T r \end{aligned}$$

For every r_i in the left, the right part $\frac{1-\beta}{N} \mathbf{1}^T r$ contributes $\frac{1-\beta}{N} \sum_i r_i$, as $\sum_i r_i = 1$, then we have

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

.

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) _____