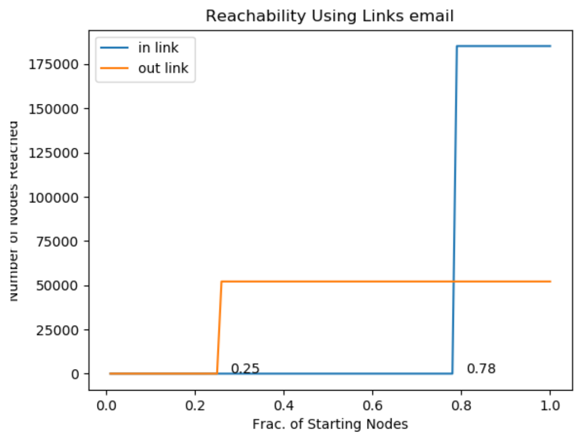
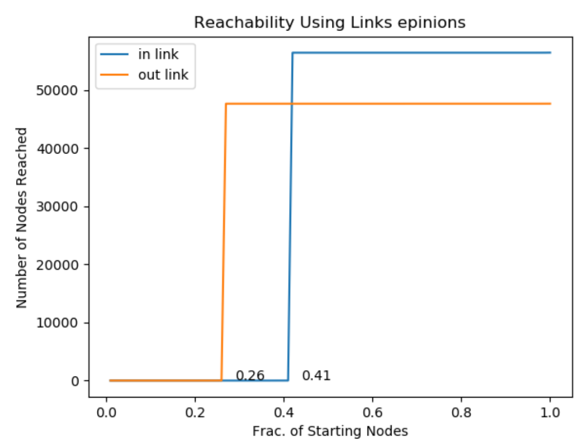
Total nodes of email: 265214
IN of 2018 in email: 1
OUT of 2018 in email: 52104
Total nodes of epinions: 75879
IN of 224 in epinions: 56459
OUT of 224 in epinions: 47676

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

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:

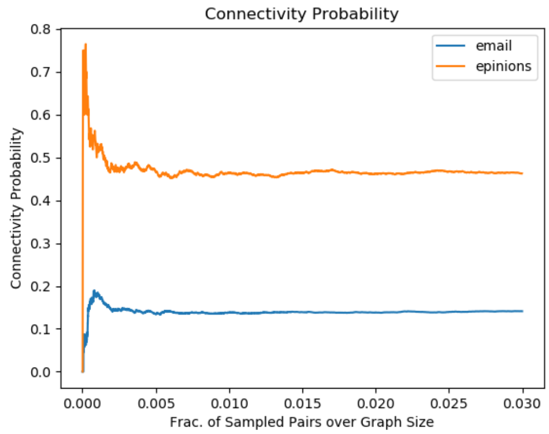


From the picture we can roughly estimate that relative sizes of SCC, IN and OUT are 43.66%, 30.34%, 15.34%.

Firstly, we can use (total_nodes - 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 TENDRILS+TUBES(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

The sampled probabiltiy 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 probabiltiy 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.

Yes. We can compute the probabilties 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).

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:

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

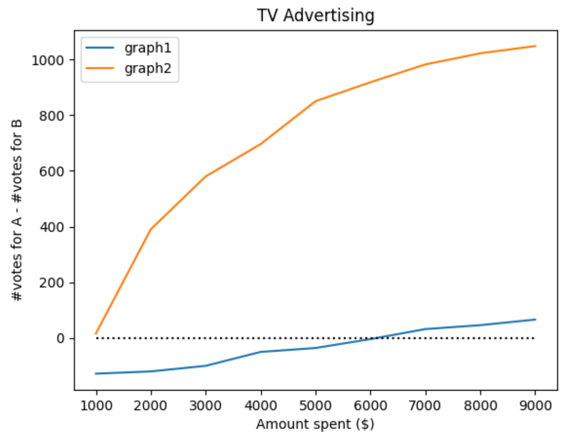
For every r_i in the left, the right part $\frac{1-\beta}{N} 1 1^T r$ contributes $\frac{1-\beta}{N} \Sigma_i r_i$, as $\Sigma_i r_i = 1$, then we have

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

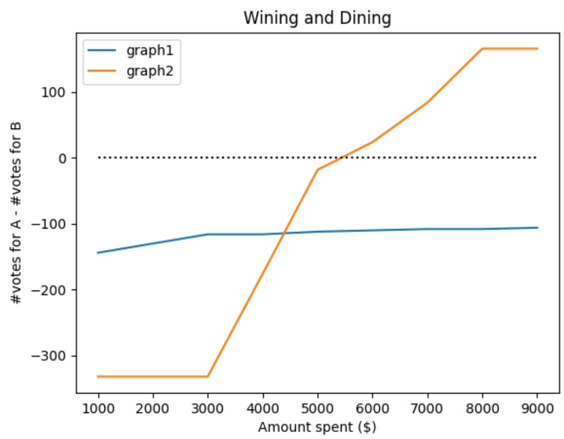
.

Sadly, to the best of my knowledge, my implementation tells that
In graph 1, candidate B wins by 162 votes
In graph 2, candidate B wins by 332 votes

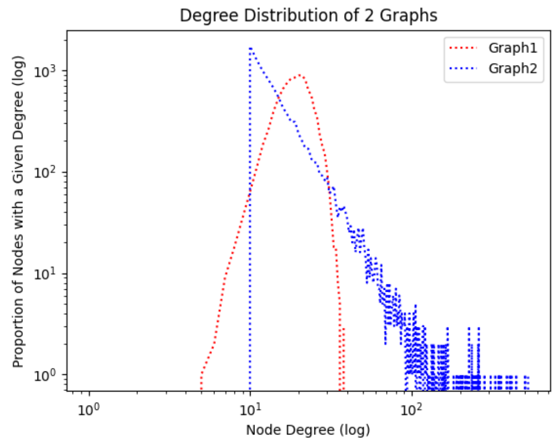
Results are
On graph 1, the minimum amount you can spend to win is 7000
On graph 2, the minimum amount you can spend to win is 1000



Results are
On graph 1, the minimum amount you can spend to win is *INF*
On graph 2, the minimum amount you can spend to win is 6000

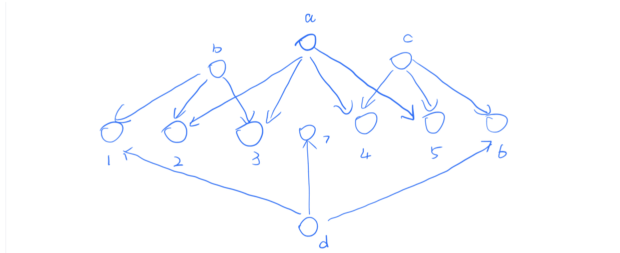


The log-log distribution is like below



We can see that Graph2 have more nodes with high degrees, so it's effective to invite high rollers in order to win vote. However, in case of advertising, changing nodes with less neigbors may bring limited contribution to win the vote.

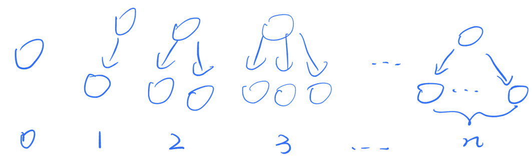
Take the below graph as an example:



when $k = 2$, $|X_u| = |\{b, 1\}| = 7$, $S = \{a, b\}$, $T = \{b, c\}$ and $f(T) = |\{b, 1, 2, 3, c, 4, 5, 6\}| = 8$.

Here is the graph, we have $T = \{b, c, d\}$ and $S = \{a, b, c\}$, where each a, b, c, d connect with $3k$ nodes, and $Nbr(a)$ shares different k nodes with $Nbr(b), Nbr(c), Nbr(d)$. Then we have $f(T) = 9k$ and $f(S) = 7k + 1$, let $0.8f(T) = f(S)$, we have $k = 20$.

$$|X_u| = 1$$


0
1
2
3
n

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) ______________________________