Analysis of PageRank Results Across Various Graph Structures

Question 4

(c)

Question

Where did most of the score tend to end up in your experiments? Look at the nodes that have the highest or lowest scores; is there a consistent pattern among your trials? Include your analysis in the pdf write-up.

Answer

1 Graph Descriptions and PageRank Results

1.1 Graph 15.1 Left

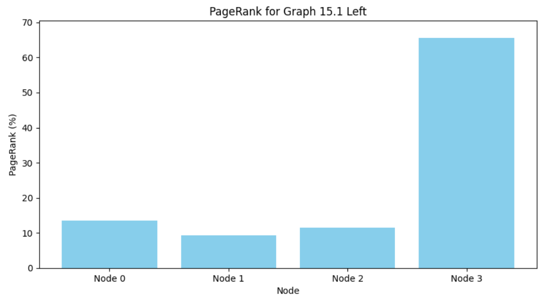
This graph is a simple cyclic graph with an additional self-loop at one node:

- Nodes: A, B, C, Z
- Edges: A $\rightarrow$ B, B $\rightarrow$ C, C $\rightarrow$ A, A $\rightarrow$ Z, Z $\rightarrow$ Z

Results:

- Node 0 (A): 0.1352
- Node 1 (B): 0.0937
- Node 2 (C): 0.1160
- Node 3 (Z, with self-loop): 0.6551

As we can see, because of the graph construction, all the most of the flow eventually goes to Z but because of the the 3 cycle ABC transferring same "score to each other" however A always Set aside extra score for Z. and Z does not gives extra score to no one but himself.
As to the **Theorem 15.2** , the ε scaled page rank converges to a unique solution.



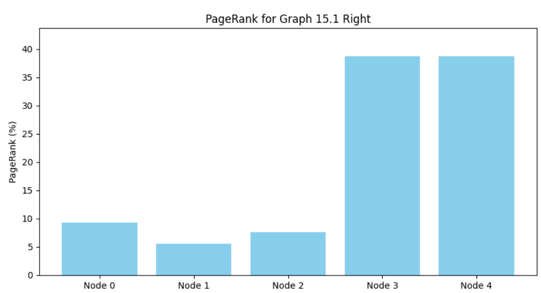
1.2 Graph 15.1 Right

This graph introduces an additional complexity with two interconnected nodes:

- Nodes: A, B, C, Z1, Z2
- Edges: A $\rightarrow$ B, B $\rightarrow$ C, C $\rightarrow$ A, A $\rightarrow$ Z1, A $\rightarrow$ Z2, Z1 $\rightarrow$ Z2, Z2 $\rightarrow$ Z1

Results:

- Node 0 (A): 0.0937
- Node 1 (B): 0.0554
- Node 2 (C): 0.0760
- Node 3 (Z1): 0.3875
- Node 4 (Z2): 0.3875



As we can see, because of the graph construction, all the most of the flow eventually goes to Z1 and Z2 because of the the 3 cycle ABC transferring same "score to each other" however A always Set aside extra score for Z1 and Z2 while they are just **giving and taking** the same score out to each other. And not giving no body else score.

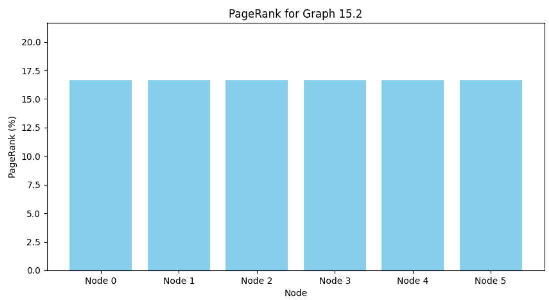
1.3 Graph 15.2

This graph consists of two separate cycles:

- Nodes: A, B, C, A', B', C'
- Edges: A $\rightarrow$ B, B $\rightarrow$ C, C $\rightarrow$ A and A' $\rightarrow$ B', B' $\rightarrow$ C', C' $\rightarrow$ A'

Results:

- All nodes have equal PageRank values of 0.1667.

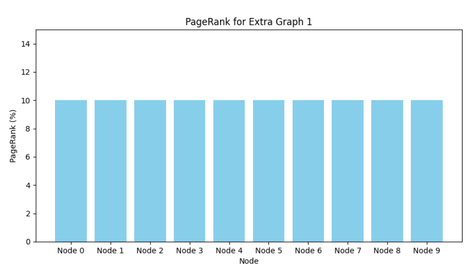


As we can see, because of the graph construction, it is constructed by to perfect cycles of 3, therefore there is no transferring each score from one cycle to other cycle.
Also for each node in the cycle, the score that he is "getting" is the score that he is "transferring" because we are in a cycle and every node has 1 degree in and 1 degree out.

1.4 Extra Graph 1

A perfect cycle with 10 nodes.

- Results: Each node has a PageRank of 0.1000.



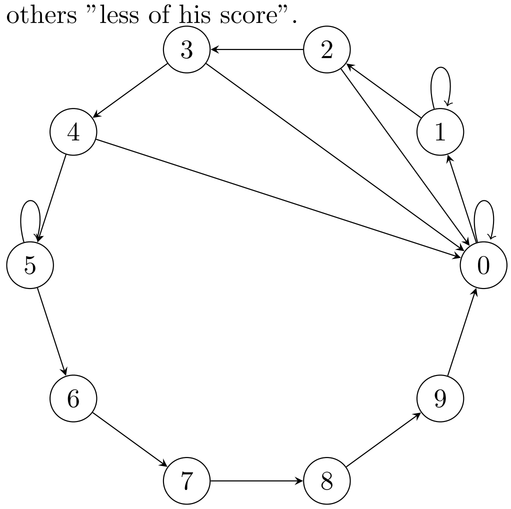
As we can see, because of the graph construction, it is the same logic as graph 15.2 however there is a 1 perfect cycle of size 10.

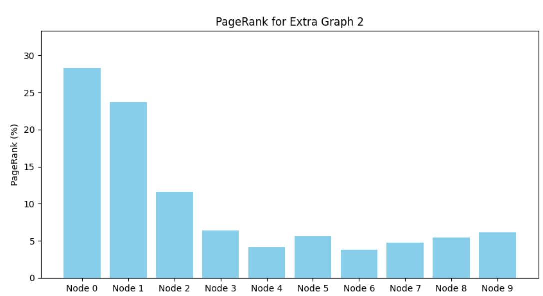
1.5 Extra Graph 2

A complex structure with self-loops and multiple edges to one node.

- Node 0, which has multiple incoming edges, shows the highest PageRank at 0.2831.
- after Node 0, Node 1 shows the highest PageRank at 0.2373.
- all of the other nodes are eventually taking much less percentage in the page rank

As we can see, because of the graph construction, the Node 0 as a lot in degree , has a loop to himself, and an edge to 1 therefore, he is getting "a lot of score" and giving "keeping to himself score because of the loop" therefore he has a bigger pageRank score than others.
Also the Node 1 gets relatively big score because he receives relatively big score from 0 who has a big score , and also has an inner loop witch make him give to others "less of his score".





1.6 Facebook graph

In the facebook graph there are 4039 nodes, after doing the PageRank algorithm, we have seen that the Node 0 had the most Rank of **0.0062 percent**.
His inner degree (and outer because it is a non directed graph) was the highest by far then the others with in degree of 347.
From the other hand the node with the least rating was node number 2079 with PageRank 0.0000 and innerDegree 1.

(d)

Question

Intuitively explain your results in terms of a measure of influence in a social network. Do you think that this is an accurate measurement? How could we try to improve it (for instance, by incorporating link strengths or other measures of popularity)? Include your answers in the pdf write-up.

Answer

We think that the first problem with social networks (and we talk about their representation as undirected graphs) is that they do not capture that relationships aren't necessarily equal, if I'm following Kanye West on Instagram, he probably doesn't follow me back, moreover, Kanye probably has much more influence over me than I have over him. This asymmetric link strength can be captured easily in social networks (and now we talk about the actual social network web-site), for example for every person, p, we can maintain a number, n_p, that represents the number of meaningful interactions that p had with the social network. A meaningful interaction is just an interaction that can indicate us how much p values the interaction's target, which is another person, p', e.g.:

- Following p'.
- Sending p' a chat message.
- Becoming friends with p'.
- Liking p''s post.
- Commenting on p''s post.

The value that we'll add to n_p for such interaction isn't necessary equal for every type of interaction, for example, following p' probably means that p' has much more influence over p than it would mean if p liked p''s post. Also, for every such target, p', we'll maintain an accumulator, $n_{p,p'}$, that will help us to decide how much influence p' has over p, every time n_p was incremented due to an action that p did, relating p', we'll increment $n_{p,p'}$ by the same number, for every p', $n_{p,p'}$ will be initiated to zero (of course we need to actually save $n_{p,p'}$ only when it's incremented for the first time to make things efficient.). We can determine the influence that p' has over p by calculating the next number:

$$\frac{n_{p,p'}}{n_p}$$

Now we can refine the PageRank algorithm by computing the score of p in every iteration likewise:

$$\text{score}(p) = \frac{\epsilon}{n} + (1-\epsilon)\sum_{n_{p',p}>0} \frac{n_{p',p}}{n_{p'}}\text{Score}(p')$$

The problem with that refinement is that it doesn't work if there are nodes that had no meaningful interaction with the site (their score gets lost), but as long we remove those nodes from the calculation or will spread their score evenly among the other nodes, defining:

$$\text{score}(p) = \frac{\epsilon}{n} + (1-\epsilon)\left(\sum_{n_{p',p}>0} \frac{n_{p',p}}{n_{p'}}\text{Score}(p') + \sum_{n_{p'}=0} \frac{1}{\text{number of nodes}}\text{Score}(p')\right)$$

We should be good (there's a problem with the marked lines in lecture 9, slide 26) and theorems 15.1, 15.2 hold. The nice thing about this improvement is that it does not only capture how relationships can be asymmetrical but also how not all of the people that influence me,
influence me the same amount,
the more p' influences me, the bigger $\frac{n_{p',\text{me}}}{n_{p'}}$ will be.