CS-E5740 Complex Networks,
Answers to exercise set 5

Alex Herrero Pons, Student number: 918697

October 20, 2020

Problem 1

a) The network given in file in `pagerank_network.edg` is represented in Figure 1.

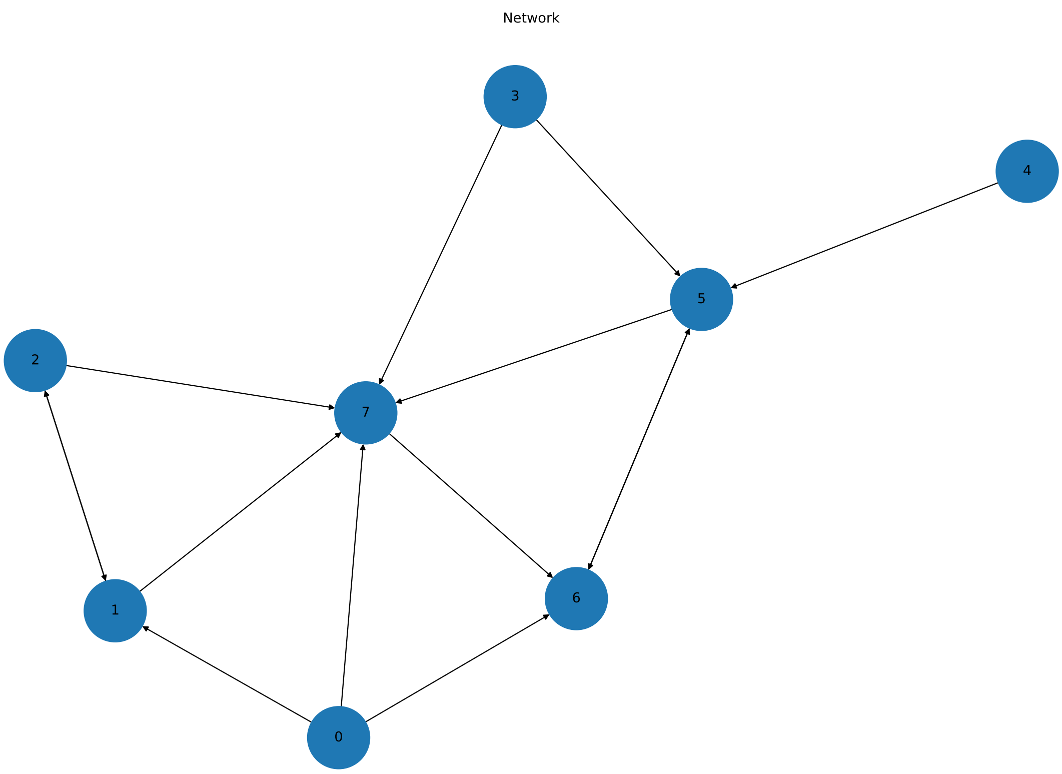

Figure 1: Pagerank network

b) Now in Figure 2 we can see the same network colored with the PageRank generated with our random walker function.

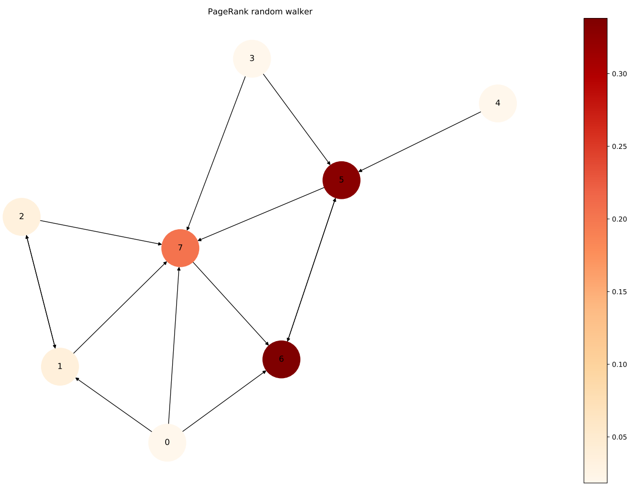

Figure 2: Pagerank network colored with the PageRank of each node generated with our random walker function

In Figure 3 we can observe how our function gives very similar results to the function `nx.pagerank`.

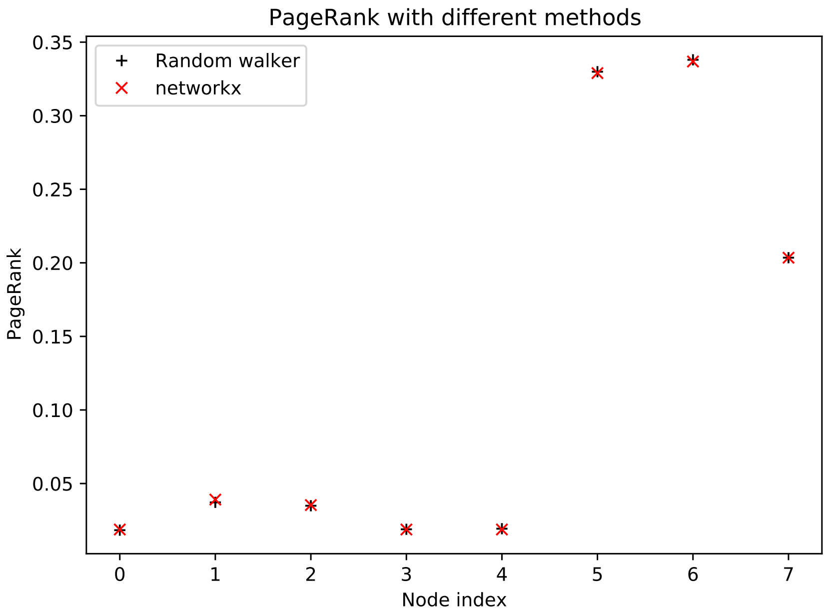

Figure 3: Comparison between the PageRank of our function and the one generated with `nx.pagerank`

c) As we can see in Figure 4 the PageRank generated with the power iteration function is the same as the one generated with the random walker function (Figure 2).

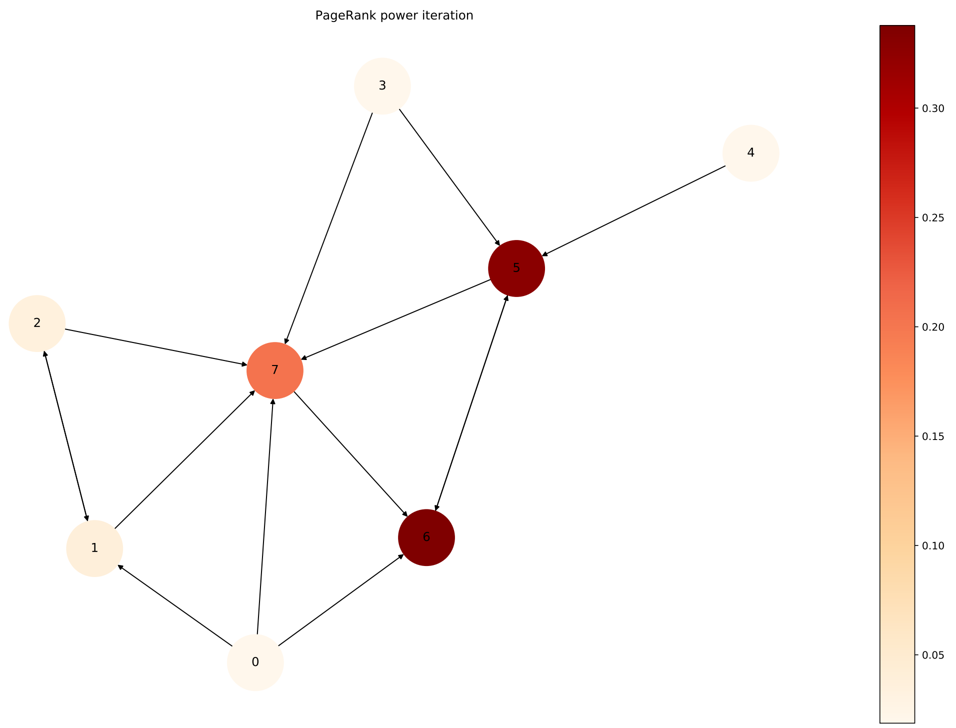

Figure 4: Pagerank network colored with the PageRank of each node generated with our power iteration function

d) We ran 3 tests for each function (`n_steps = ` 10^6 for the random walker function) with a network of size `n_nodes=`10^4 generated with `nx.directed_configuration_model(n_nodes*[5],n_nodes*[5],create_using=nx.DiGraph())`. We measured the running time of each test with `time.time()` and with the average running time of the three tests we computed the estimated running times for a network of size `N_nodes=`26×10^6 as `estimated_time = avg_time/n_nodes*N_nodes/3600`.

The obtained results are:

- Power iteration function:

  Average running time (`n_nodes=`10^4): 1.1098794142405193 seconds

  Estimated running time (`N_nodes=`26×10^6): 0.8015795769514862 hours

- Random walker function:

  Average running time (`n_nodes=`10^4): 78.05679607391357 seconds

  Estimated running time (`N_nodes=`26×10^6): 56.37435272004869 hours

As we can see there is a huge difference of running times between the power iteration and the random walker functions.

e) How the structure of a network relates to PageRank:

1) The bigger the in-degree k_{in} of a node, the bigger it's PageRank. This is because there will be higher probability that the random walker arrives to that node so there will be a higher neighbour contribution.
2) There is no connection.
3) If the node belongs to a strongly connected component (SCC) the PageRank will be lower. This is because the degree of the neighbors will be higher and therefore the probability that the random walker arrives to that node will be lower (there will be a lower neighbour contribution).
4) Knowing the relation between the in-degree and the belonging to a SCC with the PageRank, we could improve the power iteration algorithm of section c) initializing the PageRank of each nodes to a value proportionate to it's degrees and weather it belongs to a SCC.
5) Even though the node 7 has a higher in-degree the nodes 5 and 6 have higher PageRank because they have higher node contribution.

f) The resulting PageRank values for the different damping factors are represented in the Figure 5.

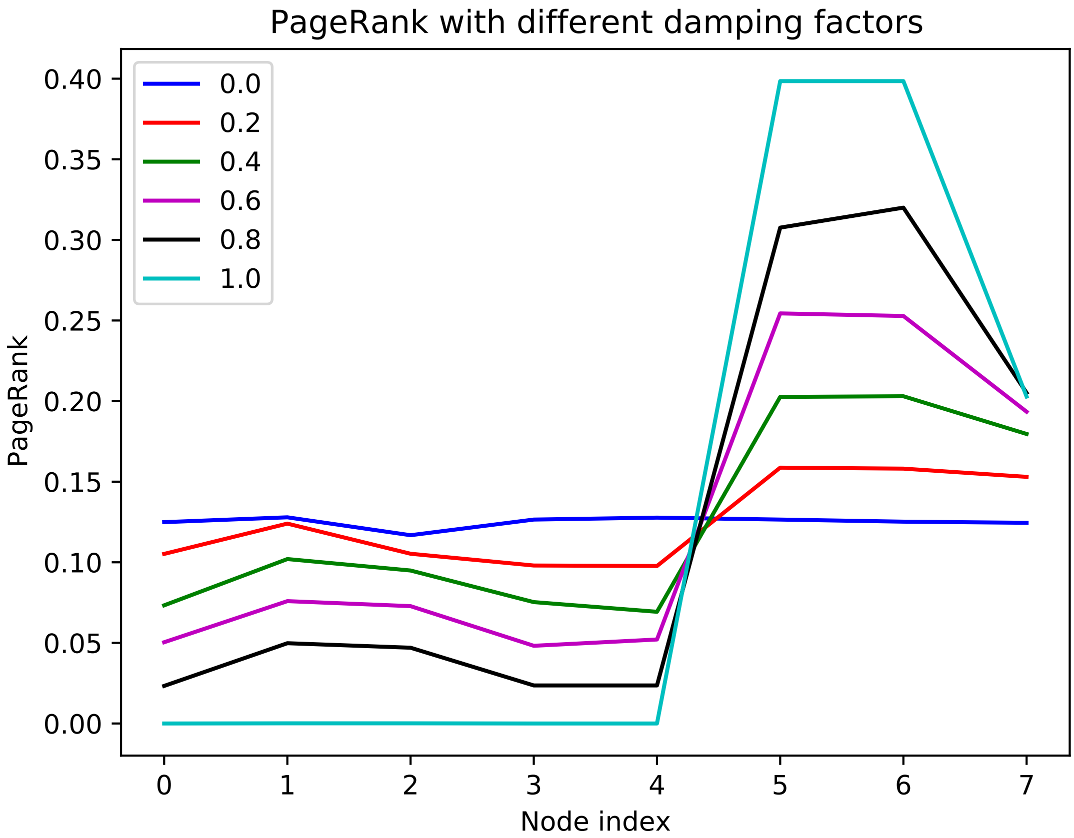

Figure 5: Effects of the damping factor on the PageRank values of the network

As we can observe in Figure 5 when the damping factor is closer to 0 all the nodes have similar rank and therefore the PageRank of the network is flattened. And as the dumping factor increases, the distance between the ranks of the nodes increases until $d = 1$ where we find that most of the nodes have rank 0.