Kunal Mehta
50352960
kunalkam

Quiz 9

1.

a. The page rank vectors for the different damping constants are shown below

```
> pg <- page.rank(g, damping = 0.05)
> pg$vector
        A         B         C         D         E         F
0.1683271 0.1639395 0.1718214 0.1681380 0.1680380 0.1597361
> pg <- page.rank(g, damping = 0.25)
> pg$vector
        A         B         C         D         E         F
0.1786588 0.1544288 0.1848587 0.1758772 0.1737324 0.1324441
> pg <- page.rank(g, damping = 0.5)
> pg$vector
         A          B          C          D          E          F
0.19227231 0.14719411 0.18583257 0.19135235 0.18399264 0.09935603
> pg <- page.rank(g, damping = 0.75)
> pg$vector
         A          B          C          D          E          F
0.19399617 0.14778661 0.17077331 0.21832113 0.20320659 0.06591619
> pg <- page.rank(g, damping = 0.95)
> pg$vector
         A          B          C          D          E          F
0.17305017 0.15761096 0.14454445 0.25658531 0.23247617 0.03573294
>
```

The rankings of importance are shown below

For 0.05 – C, A, D, E, B, F
For 0.25 – C, A, D, E, B, F
For 0.50 – A, D, C, E, B, F
For 0.75 – D, E, A, C, B, F
For 0.95 – D, E, A, B, C, F

I don't exactly understand what the question means by how sensitive, I have laid out below what the different values of the damping constant imply, and for this case how they are impacting the PageRank vector. Won't the numerical sensitivity of the vector to the damping constant differ on a case to case basis? I think what the question is trying to as is if there is a significant change, or a marginal one, or a negligible one in this case for the different constant values, and I have answered that as well.

Greater values of the damping constant indicate greater chances of aborting and restarting on a random page, therefore for higher values of the damping factor, like 0.75 and 0.95, it essentially means that the significant factor in the page rank resolution is the random component of opening that page rather than the links of the webgraph. For 0.05 and 0.25, which are tolerably low values of the damping factor, the PageRank vector is sensitive to the damping factor, for sure, which accounts for the different values of the vector elements, but it is still tolerable, where the significant factor contributing to page rank is still the webgraph connections. Get to 0.5, and there is now an equal degree of randomness and order to the page rank calculations, beyond which for 0.75 and 0.95, the significant factor shifts to being the random element of aborting and starting on a new page, which is therefore not helpful in terms of figuring out how important each page is in relation to the others.

As regards sensitivity to the constant, I believe the overall ranking of importance stays valuable to infer from as long as the constant remains in the range of say a max of 0.30, maybe 0.35. Once you start exceeding that range, the random component of the

calculation becomes too significant for the PageRank vector to yield any reliable insights.

b. For the initial lower values of damping constant, which indicate that the webgraph architecture determines more of the ranking, yes, it does seem intuitively correct. However, once it goes to 0.5 and higher, I sort of see why D, E, A have a high rank, but for the graph, I still think that the order of importance that we got for 0.05 and 0.25 is more intuitively correct.

2. The PageRank vector for Webgraph B with damping constant = 0.15 is shown below

```
> pg2 <- page.rank(g2, damping = 0.15)
> pg2$vector
        A         B         C         D         E         F
0.1541610 0.1418827 0.1582538 0.1091405 0.1091405 0.1091405
        G         H
0.1091405 0.1091405
>
```

The only nodes with incoming links are A, B and C. For node A, A has 2 direct incoming nodes B and C, both of which in turn have 2 and 3 incoming nodes respectively. Based on the architecture, D, E, F, G and H are all analogous to leaf nodes of a tree, and since they only have outgoing nodes, I can see why they all have a low and equal values in the converged vector. I expected C to have a higher rank than B since it has more incoming nodes, and a greater one than A as well for the same reason. The main question in my mind was whether A would be more important than C, since it has 2 incoming nodes granted, but both of those also have incoming nodes themselves, and D, E lead to B, which leads to A, so that would increase A's importance, while F,G,H lead all to C, which also leads only to A. Either ways I knew it would be close, and like I thought, C edged it, but barely. This intuitively is also correct, since The D,E,B sector leading to A does increase it's importance, but numerically there are more nodes that lead to C directly than to B.