Juan Zamudio
Graph Algorithms
William Devanny
November 12, 2017
SP4: Pagerank Write-Up
https://github.com/juanezamudio/Wikipedia-PageRank

I enjoyed implementing the Pagerank algorithm for this short project. It was interesting to see which Wikipedia pages were the top pages. I definitely encountered some implementation difficulties when it came to actually obtaining the Pagerank results. There were issues with pages having pagerank results of infinity which was a little troublesome. It took some time to find the bug that caused this to occur. Along with this issue were nullpointer bugs that kept appearing in the subtractMap() function. These bugs were hard to catch because the logic was always a little hard to follow and trace and print statements just sometimes did not suffice. Thus, sometimes debugging with breakpoints had to occur. As for when deciding when to terminate computation, I followed the *v* and *vOld* subtraction value up to the precision we discussed in class. This value ended up being 0.0001. The values obtained from my Pagerank implementation are on the next page.

By varying epsilon values, we can see that some rankings change. With a higher epsilon, we can see that the rankings for country categories like Italy and Australia went up while, for other country categories like India and Poland, they went down. What was most interesting from raising the epsilon value to 0.30 was that Russia, World War I, New York, and Brazil were not ranked in the top 20. With a lower epsilon, we can see that the top pages that we have seen with higher epsilons are now being ranked lower and the lower ranked pages are being ranked higher. For instance, Canada goes from ranking 4th to 6th and UK goes from ranking 3rd to 2nd. This happens in many instances across the rankings for the epsilon value of 0.05. We also see, again, categories that once ranked in the top 20 do not rank in the top 20 anymore: Departments of France, Communes of France, and Brazil. It is interesting that Departments of France and Communes of France both do not rank with this epsilon value. It seems that if epsilon is increased, the probability of landing on a certain page is decreased and so pages that switched rankings in this position create a ranking that is more accurate as it is likely for pagerank to get stuck in a cycle with the lower epsilon. On the contrary, the lower epsilon value may cause pagerank to get stuck in a cycle and that is why we see the normally higher ranking pages ranked lower than the normally lower ranking pages. Some categories switch positions in this manner with the lower epsilon because pagerank gets stuck in a cycle and so there is more amount of time spent on these lower ranking nodes. This makes them rank higher than they actually should be ranked.

I thought the results of the algorithm and kind of surprised me for a moment but they do make sense. I think since the United States is one of the most influential superpowers in the world, if not the most, it makes sense that it is at the top of the list. It

also makes sense that other countries are there, especially the most influential countries in Europe. Although I knew soccer was the most popular sport in the world, I did not know it was under Association Football so that was a little surprising. I surely thought that it would be under futbol or the like, but it does make sense that that is in the top 20. The only two hits that I was honestly surprised about were the Departments of France and the Communes of France. I thought these two referred two the same thing until I looked it up. Nonetheless, it was interesting to see that these two categories are in the top 20. As for the epsilon values of the data, I thought it was interesting that the United States had an epsilon value that was almost twice the epsilon value of the second place category. I guess the United States is really that important when it comes to Wikipedia page rankings. That was interesting to note. Even with a change in epsilon values, the United States page remains supreme so that was kind of weird.

Node #	Name	Rank & ε = 0.15	Rank & ε = 0.30	Rank & ε = 0.05
279122	United States	1) 5.51 E-3	1) 5.30 E-3	1) 5.37 E-3
987583	France	2) 2.37 E-3	2) 2.44 E-3	3) 2.14 E-3
541013	United Kingdom	3) 2.16 E-3	3) 1.83 E-3	2) 2.36 E-3
230038	Canada	4) 1.72 E-3	4) 1.68 E-3	6) 1.61 E-3
896828	Germany	5) 1.67 E-3	5) 1.57 E-3	4) 1.68 E-3
121347	World War II	6) 1.50 E-3	6) 1.26 E-3	5) 1.65 E-3
1055792	English Language	7) 1.37 E-3	9) 1.19 E-3	7) 1.47 E-3
610154	Australia	8) 1.31 E-3	7) 1.34 E-3	9) 1.24 E-3
98332	Italy	9) 1.29 E-3	8) 1.20 E-3	8) 1.31 E-3
1118496	India	10) 1.20 E-3	12) 1.10 E-3	10) 1.23 E-3
362517	Japan	11) 1.14 E-3	10) 1.14 E-3	11) 1.03 E-3
1313468	Poland	12) 9.68 E-4	14) 1.07 E-3	17) 7.43 E-4
82313	Association Football	13) 9.58 E-4	13) 1.09 E-3	19) 6.80 E-4
446994	Spain	14) 8.75 E-4	20) 8.11 E-4	13) 8.75 E-4
987710	Departments of France	15) 8.43 E-4	11) 1.12 E-3	**NOT RANKED**[1]
1457593	Russia	16) 8.24 E-4	**NOT RANKED**[1]	14) 8.42 E-4
121348	World War I	17) 7.99 E-4	**NOT RANKED**[1]	12) 8.98 E-4
987709	Communes of France	18) 7.88 E-4	15) 1.06 E-3	**NOT RANKED**[1]
303904	New York	19) 7.78 E-4	**NOT RANKED**[1]	15) 7.81 E-4
1362343	Brazil	20) 7.28 E-4	**NOT RANKED**[1]	**NOT RANKED**[1]

[1] In the Top 20