NSD[2017] Practical 03 – PageRank

Professor Maximilien Danisch

Muti'ah Binti

Israel Márquez S.

Section 1 – Dataset.

Completed, files required to run the algorithm.

Section 2 - PageRank

Implement an efficient PageRank algorithm using the power iteration method.

Code has been created and it can be found on the following GitHub link:

https://github.com/imarquez21/NSD/tree/master/Practical3

How many iterations seem necessary to reach convergence?

From our analysis we can say that the more the iteration the more feasible for it to converge. We ran three test with different iterations on the wikipedia file and the values are:

- Iterations: 20
  - Number of Edges: 46092177
  - Number of Nodes: 13834639
  - P[12]: 0.0002227739126334962
  - P[25]: 0.0002227739126334962
  - P[39]: 0.0002227739126334962
  - P[309]: 0.0002227739126334962
- Iterations: 5
  - Number of Edges: 46092177
  - Number of Nodes: 13834639
  - P[12]: 0.0002657456294858770
  - P[25]: 0.0002657456294858770
  - P[39]: 0.0002657456294858770
  - P[309]: 0.0002657456294858770

- Iterations:2
  - Number of Edges: 46092177
  - Number of Nodes: 13834639
  - P[12]: 0.0002686347336606589
  - P[25]: 0.0002686347336606589
  - P[39]: 0.0002686347336606589
  - P[309]: 0.0002686347336606589

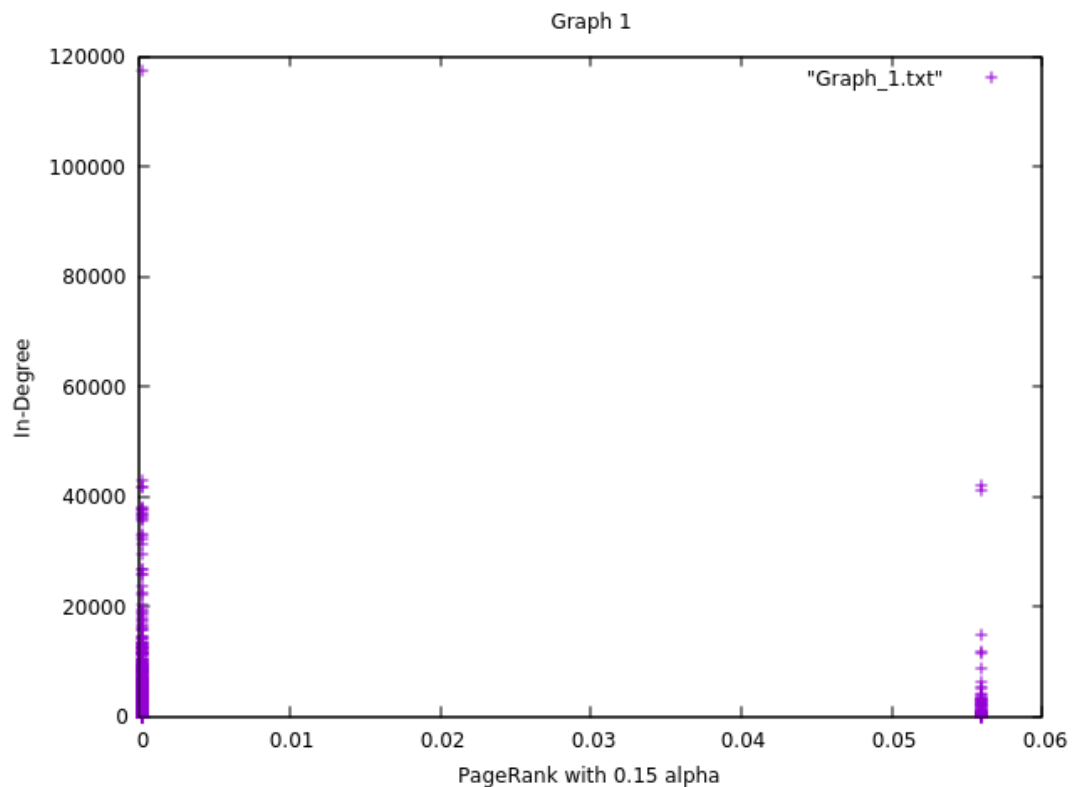Test it on the Wikipedia network with α = 0.15 and 20 iterations: which pages have the highest PageRank? The lowest?

We ran our algorithm and identified that the website witt:

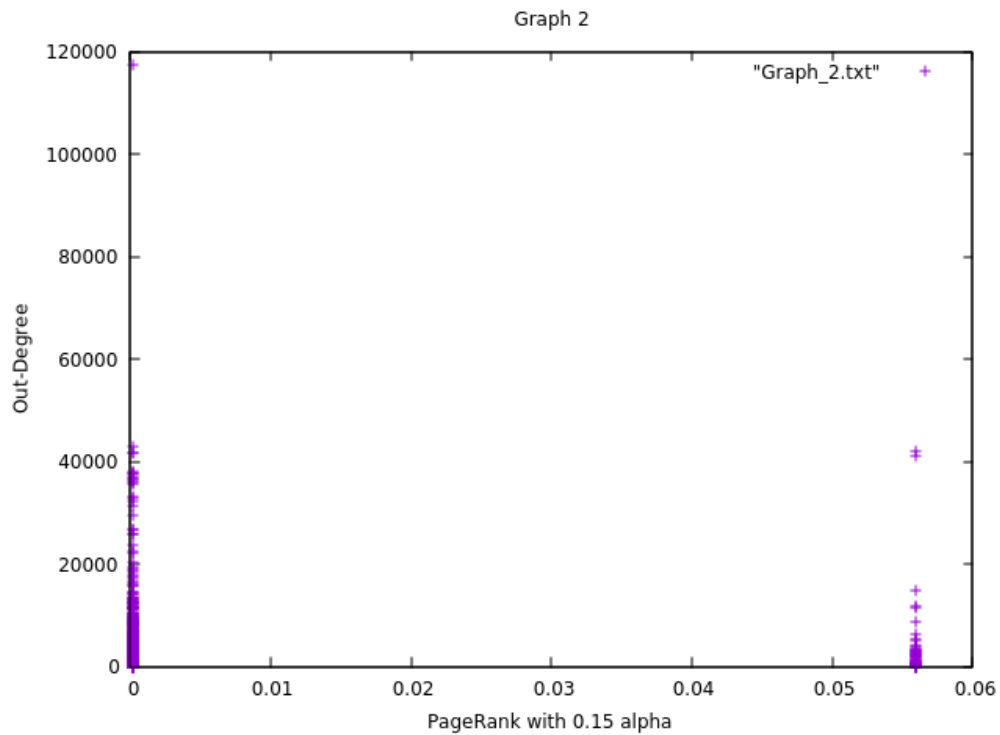The highest is "Agriculture" with a PR of 0.0040443497452324383.

The lowest is "Anarchisim" with a PR of 0.0002686347336606589.
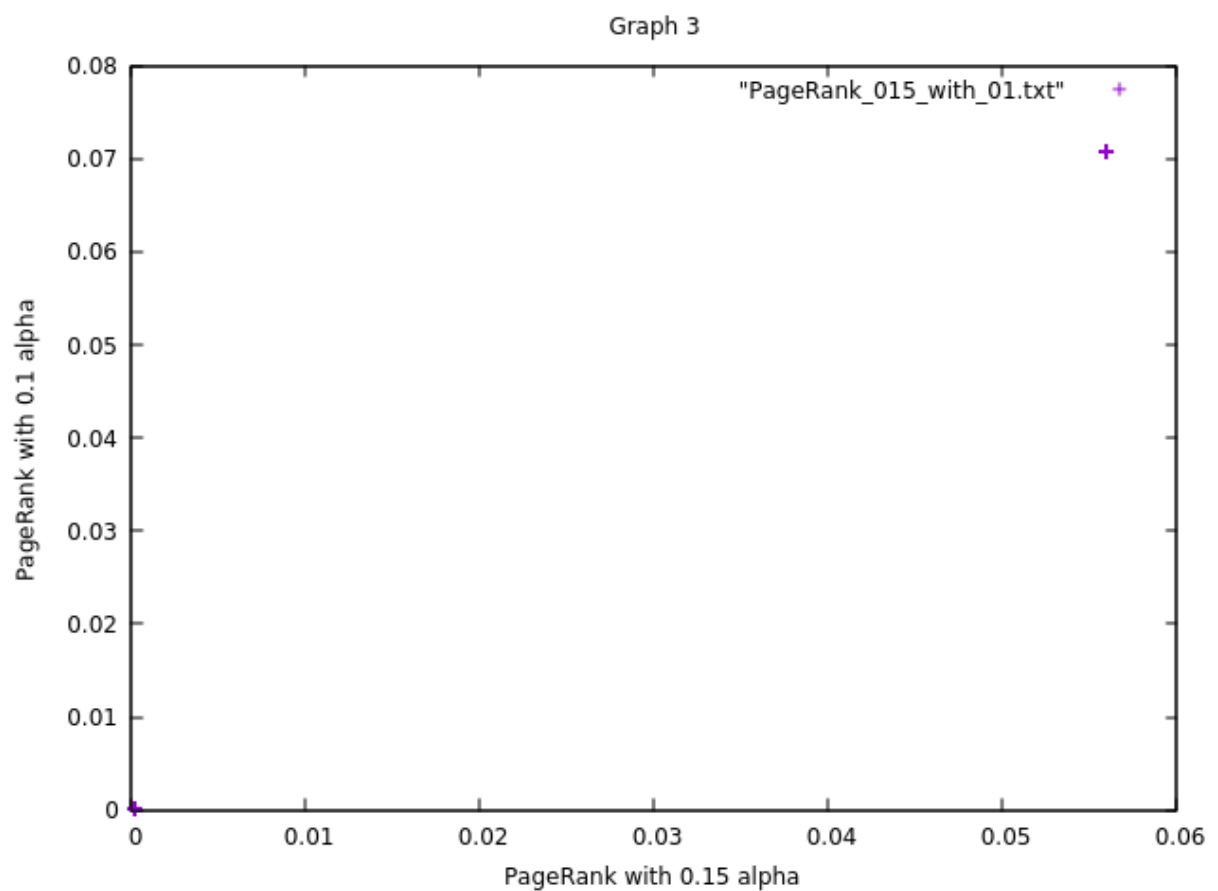
Section 3 – Correlations
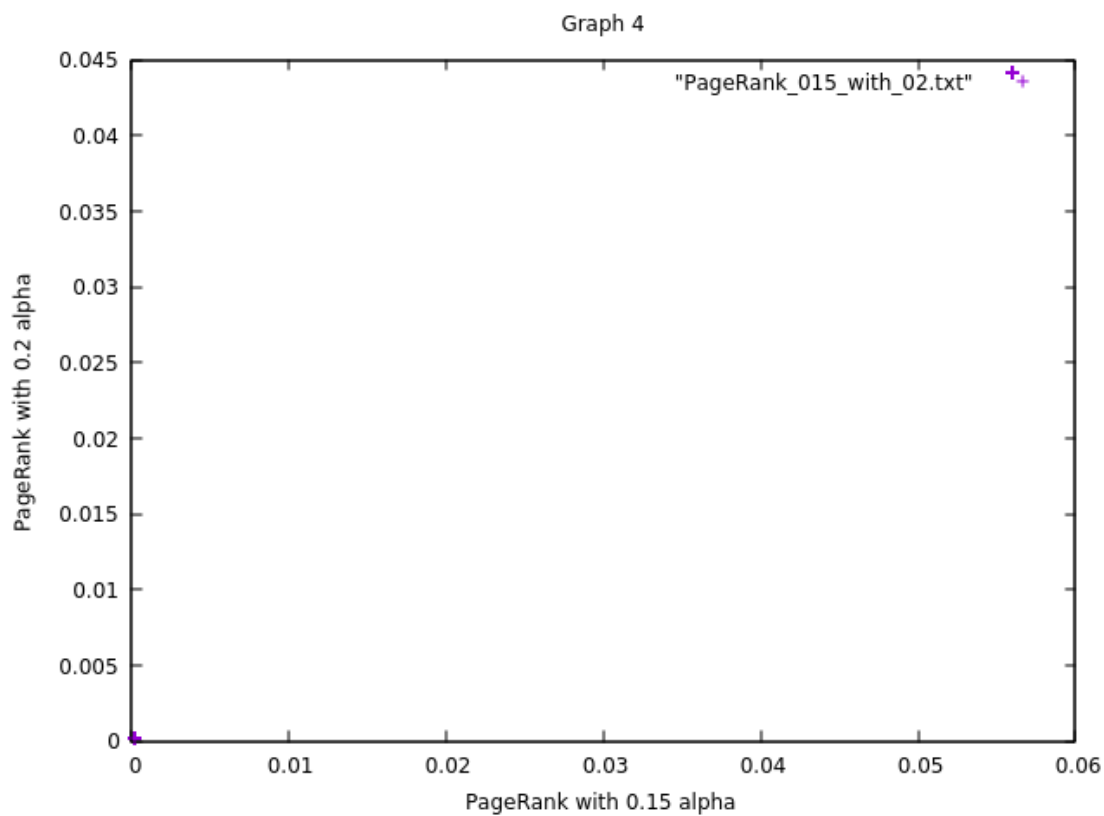
1. x = PageRank with α = 0.15, y = in-degree
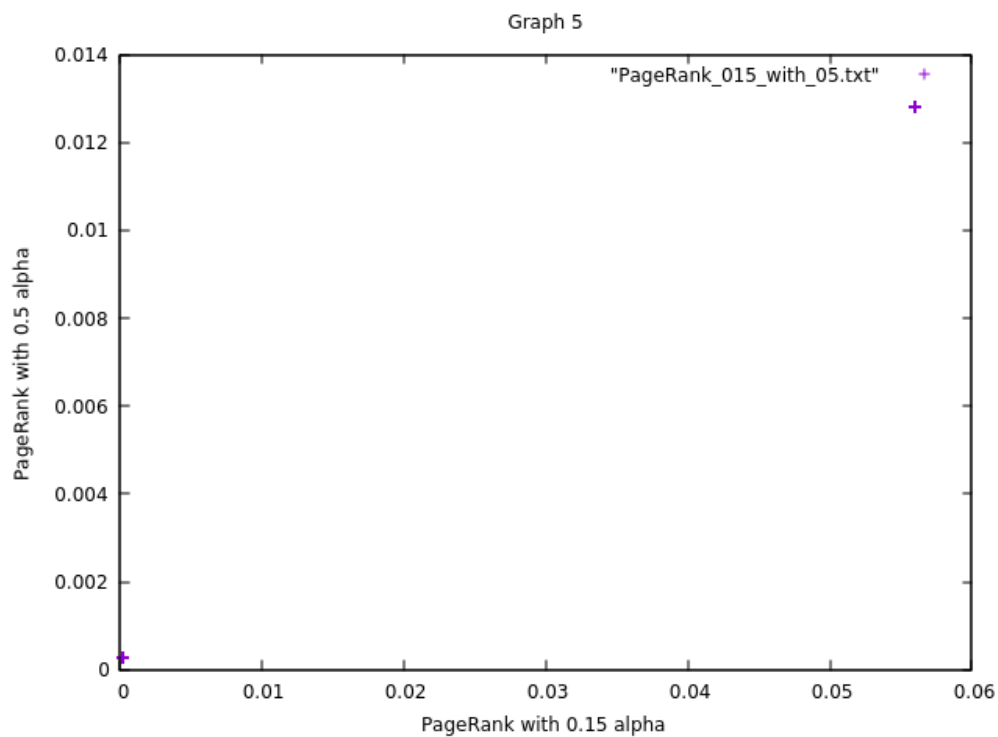

Graph 1

2. x = PageRank with α = 0.15, y = out-degree

3. x =PageRank with α = 0.15, y = PageRank with α = 0.1

4. x = PageRank with α = 0.15, y = PageRank with α = 0.2

Graph 4

"PageRank_015_with_02.txt"

PageRank with 0.2 alpha

PageRank with 0.15 alpha

5. x = PageRank with α = 0.15, y = PageRank with α = 0.5

Graph 5

"PageRank_015_with_05.txt"

PageRank with 0.5 alpha

PageRank with 0.15 alpha

6. x = PageRank with α = 0.15, y = PageRank with α = 0.9.
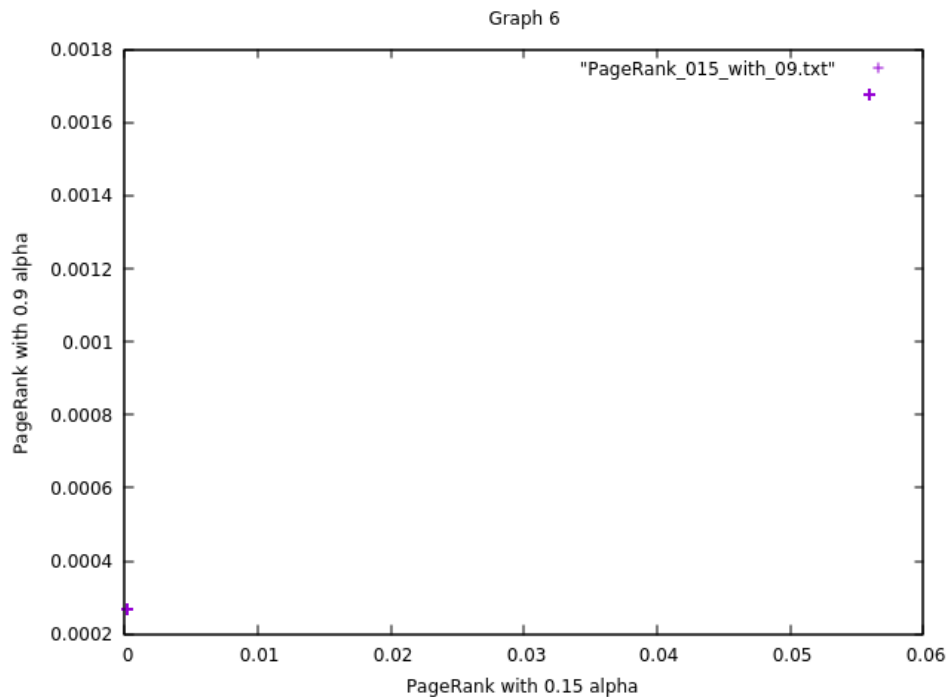
- Should you use linear or log scales?
  - We consider that linear scales should be used for this case as with logarithmic scales the plot behavior is similar. Additionally based on our results we haven't identified a behavior in which the growth or decay follows a logarithmic pattern.
- What can you say about the correlations between the values?
  - A particular behavior we have identified is that the greater the degree of the node the smaller the PageRank value is, we were expecting the opposite in which the greater the degree of the value the higher the PageRank.