PageRank

Maximilien Danisch

Exercise 1 — *PageRank (directed graph)*

Download the Wikipedia dataset from:
`http://cfinder.org/wiki/?n=Main.Data#toc1`.
Download two files: (i) the list of directed hyperlinks and (ii) the name of the pages (corresponding to each node ID).
Implement PageRank using the power iteration method and test your algorithm on the Wikipedia network with $\alpha = 0.15$.
How many iterations seem necessary to reach convergence?
Report the 5 pages with the highest PageRank and the 5 pages with the lowest PageRank.

Exercise 2 — *Correlations*

Draw six scatter plots such that, for each node in the graph:

1. x = PageRank with $\alpha = 0.15$, y = in-degree;
2. x = PageRank with $\alpha = 0.15$, y = out-degree;
3. x = PageRank with $\alpha = 0.15$, y = PageRank with $\alpha = 0.1$;
4. x = PageRank with $\alpha = 0.15$, y = PageRank with $\alpha = 0.2$;
5. x = PageRank with $\alpha = 0.15$, y = PageRank with $\alpha = 0.5$;
6. x = PageRank with $\alpha = 0.15$, y = PageRank with $\alpha = 0.9$.

Should you use linear or log scales?
What can you say about the correlations between x and y values?

Exercise 3 — *Personalized PageRank*

Implement the personalized PageRank using the power iteration method.
Reproduce the plot seen in class (rooted pagerank of Magnus Carlsen).
Download the Wikipedia categories from the url given in Exercise 1.
Suggest a relevant restart vector for a user interested in Chess and Boxing.
Run your personalized PageRank algorithm with this vector and comment the result.

Exercise 4 — *(Optional) Push method*

Implement the personalized PageRank using the push method, page 6 of: `http://www.leonidzhukov.net/hse/2016/networks/papers/andersen06localgraph.pdf`.
Compare the runing time and precision of the power iteration and the push method.