https://www.theatlantic.com/national/archive/2013/05/why-high-school-rankings-are-meaningless-and-harmful/276122/

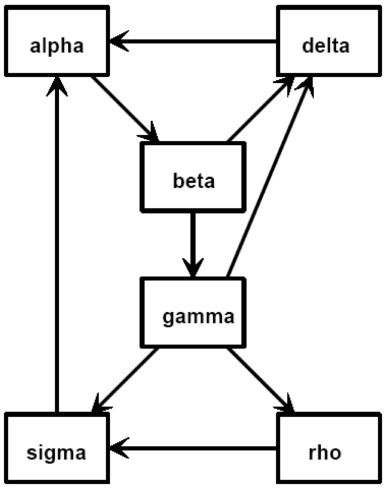
Homework 2: Googling Your Ranking

Please submit your completed work before 7 AM EDT on Thursday, July 15.

This homework will have you 1) running and adapting ranking code, 2) working with the PageRank algorithm and 3) reflecting on some dangers in ranking, which is important as you consider your research. Note, you will need to download codes from Session 2 to complete some of the problems.

1. Read *Why High-School Rankings Are Meaningless – and Harmful* at `https://www.theatlantic.com/national/archive/2013/05/why-high-school-rankings-are-meaningless-and-harmful/276122/`. In 2-3 paragraphs describe what you found most interesting about the article. Note, I'm asking for paragraphs not sentences in your response.

2. Consider the tiny network of web pages given below:



   For your PageRank model in this problem, assume there is an 85% chance that a surfer will follow a link on a web page.

   (a) (By hand) Find the Google matrix G for this network. Your answer should be the matrix G.

   (b) (Using code) By adapting the adjacency matrix within the codes given in our course. Find the PageRank vector for this network. Your answer should be a table with two columns. One column is a list of the webpages (alpha, beta, delta and such). The second column is the PageRank of the associated web page in that row. List the pages in descending order as sorted according to their corresponding PageRank value.

3. Choose a root page on Wikipedia and a size of a network. (Warning, large networks can take awhile so I'd advise an upper bound of 1000.) Then, use the python Jupyter notebook `wikipediaPageRankPython.ipynb` to find the top 10 pages in your network. Your answer should give your root page and the table of results. You will need to have beautiful soup installed which can usually be done with the command

```
pip install beautifulsoup4
```

   If you need me to run a network, let me know at least 3 days before the assignment is due.

4. Think of an application where you can create a network of at least 5 entities. Create the adjacency matrix and compute the PageRank of your network. In your write-up, describe your application, what the nodes in your graph are and what is represented by an arrow from one vertex to another in the digraph. Include a spy plot of the adjacency matrix and give the final PageRank of the network. If the network is larger than 10, just give the top 10. For your network, you could:
   (a) choose pages in Wikipedia and see which are connected to which, as I did with the example in class,
   (b) choose Twitter users and see who follows who,
   (c) work with sports teams and have the losing team in the digraph point at the winning team, or
   (d) pick your own application.
5. What are your current thoughts regarding your research? After seeing PageRank, what comes to mind? Do you have questions that will help you think about research into a ranking question? If what, what are your questions?

Submitting your work - All your responses should be contained in one document saved as a PDF. Upload the PDF and your programming files to Schoology.