Homework 6 for SM 233 Section 4031

This homework assignment has two parts. One part involves writing code, the second part involves proving some facts about stochastic matrices.

Part 1

Due Thursday 3/8/2012 at 2359

Given a collection of webpages $W = \{w_1, \ldots, w_N\}$ a particularly concise way to represent the set of links is as follows. Suppose that there is a link from the page w_i to the page w_j, then we represent this by the ordered pair (i, j). The set is denoted E.

We define F_i to be the set of pages to which the page w_i links, i.e., $F_i = \{w_j : (i, j) \in E\}$. We also let B_i denote the set of webpages that link to w_i, $B_i := \{w_j : (j, i) \in E\}$. We let N_i denote the number of outgoing links from page w_i, i.e., N_i is the number of elements in the set F_i.

Using this notation we define two versions of the link matrix $L = l_{i,j}$.

In both versions we set $l_{i,j} = 1/N_i$ if F_i is non-empty, i.e., there are links leading out of w_i. The only way in which the two link matrices differ is in the way we handle pages with no outgoing links. We now describe this difference.

Suppose that F_i is empty, that is, w_i has no links going out. In version 1, we let $l_{i,i} = 1$ and $l_{i,j} = 0$ for $i \neq j$. In version 2, we let $l_{i,j} = 1/N$ for all j. Hence, in version 1, we treat w_i as an absorbing state. In version 2, we assume that once we reach page w_i, we then choose another page on the web at random and move to it.

Now let J denote the $N \times N$ matrix of all 1s. Note that $\frac{1}{N}J$ is stochastic. Given a number α such that $0 < \alpha < 1$ let

$$G_\alpha = \frac{\alpha}{N} J + (1 - \alpha) L,$$

where L is the link matrix.

Write a function called `google` that has the following signature

```
function [f1, f2] = google(E, alpha)
...CODE...
end
```

The matrix `E` is $p \times 2$ and each row represents a link. So if the first row is `[1, 2]` then there is a link from w_1 to w_2. The program needs to do the following:

- compute both versions of the the link matrix;
- compute the corresponding versions of the matrix G_α;
- compute the corresponding stationary distrbutions `f1, f2` (using eigenvalues).

After doing this you should run your function on the small set of webpages shown in Figure 1 and compute the two stationary distributions. Which version do you think produced better results?

Part 2

Submit in-class Friday 3/9/2011

Let P, Q be stochastic matrices, let $0 \leq \alpha \leq 1$ and let f be a probability vector (distribution). Show that

(1) $\alpha P + (1 - \alpha) Q$ is stochastic;
(2) fP is a probability vector.

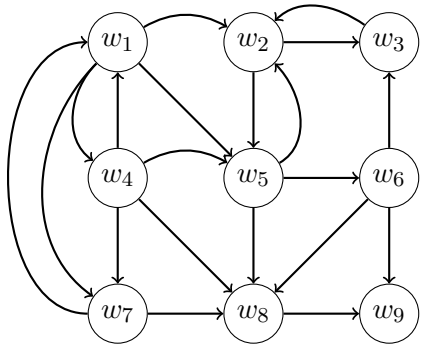

FIGURE 1. A small WWW