Assignment 5: Markov Model for PageRank

Assigned: Nov. 18
Due: Dec. 9

Write a function \( \text{PageRank}(A,E) \) that computes page rank based on links, as described in the class notes. The input parameter \( A \) is an \( N \times N \) adjacency matrix; \( A(I,J) = 1 \) if page \( I \) has a link to page \( J \). The input parameter \( E \) corresponds to the probabilistic parameter \( e \) as described in the notes.

Ignore any self-loops. Treat a page with no outlinks as if it had an outlink to every other page, as in solution 3, section 10.2.2 of the notes.

Do as much as possible with matrix operations, keeping the use of explicit loops to a minimum.

Once you have constructed the transition matrix \( M \), you can find the stationary distribution \( \vec{P} \) as follows. The vector \( \vec{P} \) satisfies the equations \( M \vec{P} = \vec{P} \), and the equation \( \sum_{I=1}^{n} \vec{P}[I] = 1 \). We can rewrite the first equation as \( (M - I^n) \cdot \vec{P} = \vec{0} \), where \( I^n \) is the identity matrix. For a Markov model with a unique stationary distribution, \( M - I^n \) is always a matrix of rank \( n-1 \). We can incorporate the final constraint \( \sum_{I=1}^{N} \vec{P}[I] = 1 \) by constructing the matrix \( Q \) whose first \( n \) rows are \( M - I^n \) and whose last row is all 1’s. This is an \( (n+1) \times n \) matrix of rank \( n \). The vector \( \vec{P} \) is the solution to the equation \( Q \cdot \vec{P} = \langle 0,0\ldots,0,1 \rangle \).

For example suppose \( M = \begin{bmatrix} 1/2 & 2/3 \\ 1/2 & 1/3 \end{bmatrix} \):

\[
\begin{bmatrix}
0.5000 & 0.6667 \\
0.5000 & 0.3333
\end{bmatrix}
\]

\[
\begin{bmatrix}
-0.5000 & 0.6667 \\
0.5000 & -0.6667
\end{bmatrix}
\]

\[
\begin{bmatrix}
-0.5000 & 0.6667 \\
0.5000 & -0.6667 \\
1.0000 & 1.0000
\end{bmatrix}
\]

\[
\begin{bmatrix}
0.5714 \\
0.4286
\end{bmatrix}
\]

Incidentally, this is not the way that you find PageRank for a graph with billions of web pages; you use an iterative algorithm.