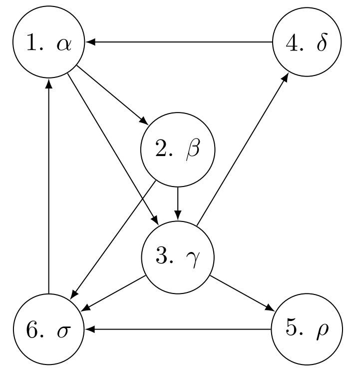
Lecturer: Barbara Re **TAs:** Davide Torlo, Stefano Colombo

Exercise session 9, 13 November, 2019

Numerical Methods in Informatics (L + E)

First part:
Correction of homework 4.

Exercise 1:
The picture on the right shows a toy network representing $n = 6$ web-pages with the links in between. The *stochastic matrix* A can be built as a sparse matrix whose entry a_{ij} is non-null only if the page j has a link to the page i. In this case, $a_{ij} = 1/N_j$ where N_j is the number of pages (including i) linked by j. Notice that the matrix is non-symmetric, indeed page j may link to page i even if i does not link to j. Notice also that the columns sum up to 1.
The PageRank algorithm is a basic algorithm, published in 1998 by Google's founders, Sergey Brin and Larry Page, which ranks a page, i.e. establishes the importance of a web-page, according to the number and the importance of pages that link to it. Mathematically, it can be expressed as an eigenvalue problem, with $\lambda = 1$:



Toy network of 6 pages.

$$\mathbf{x} = A\mathbf{x}$$

a. Due to the features of A, we know that the dominant eigenvalue is $\lambda_1 = 1$, and it is unique and distinct. The components of the associated eigenvector $\mathbf{x}_1$ are the ranks of each page. Write a MATLAB function that exploits the power method to compute it. Take into consideration the sparsity of the matrix A while developing the routine.

b. For the toy network in the picture:

- **b1.** write the stochastic matrix;
- **b2.** use the function developed in point **a.** to compute the eigenvector $\mathbf{x}_1$;
- **b3.** display in a bar plot (histogram) the rank of the pages, i.e. components of $\mathbf{x}_1$.
- **b4.** Which page has the maximum rank?

Exercise 2 Do at home!
The file `network_sample.mat` contains the sparse matrix A, representing a sample network[1]. More specifically, it is the connectivity matrix of the network, whose non-zeros entries are all 1, i.e. they are not normalized by N_j.

- Plot a diagram that shows the sparsity of the connectivity matrix.
- Build the stochastic matrix by dividing each non-zero entry of A by N_j. Then, use the function developed in exercise 1 to compute the eigenvector $\mathbf{x}_1$ of the stochastic matrix.
- Display in a bar plot (histogram), the rank of the pages listed in this network.

Hint: in MATLAB, you can load the matrix through `load network_sample.mat`; the command `sum(full(B)` returns a row vector containing the sum of each column of the sparse matrix `B`. Remind also that the matrix product `C*D`, with `D` diagonal, multiplies each row i of `C` by the i-th diagonal element of `D`.

[1] **Source:** David F. Gleich, `http://www.cs.purdue.edu/homes/dgleich/nmcomp/matlab/`, *wb-cs.stanford.mat*