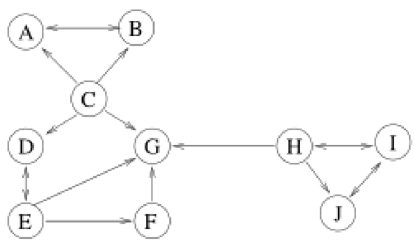
MDM2023, Exercise 4.1

April 2023

This is the complete graph with which (or it's subgraphs) we are working with.



Adjacency matrix

This adjacency matrix can be used as a root for both Page Rank and the HITS tasks. The matrix is of the form $\mathbf{A}[from, to]$:

$$\mathbf{A} = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \\ I \\ J \end{array} \begin{array}{c} \begin{array}{cccccccccc} A & B & C & D & E & F & G & H & I & J \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

Task a)

We can work directly with the transpose of this matrix (this allows normal matrix*vector products). So, we transpose $\mathbf{A}$, and deal with the dead-end/dangling node G by adding links to all other nodes. By normalizing all columns to sum to 1, ,we get the following stochastic matrix:

$$\hat{\mathbf{A}} = \begin{array}{cc} & \begin{array}{cccccccccc} A & B & C & D & E & F & G & H & I & J \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \\ I \\ J \end{array} & \begin{pmatrix} 0 & 1 & 1/4 & 0 & 0 & 0 & 1/10 & 0 & 0 & 0 \\ 1 & 0 & 1/4 & 0 & 0 & 0 & 1/10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/10 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 1/3 & 0 & 1/10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1/10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 1/10 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 1/3 & 1 & 1/10 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/10 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/10 & 1/3 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/10 & 1/3 & 1/2 & 0 \end{pmatrix} \end{array}$$

Now, we define teleportation probability ϵ, which gives us state transition $x_t \to x_{t+1}$:

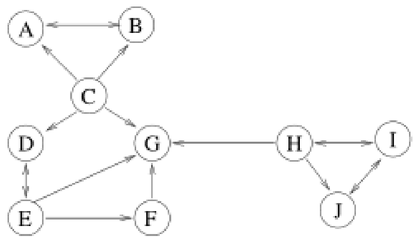
$$x_{t+1}[i] = (1-\epsilon)\sum_{j=1}^{10} \hat{\mathbf{A}}_{ij} x_t[j] + \epsilon \frac{1}{10} \sum_{j=1}^{10} x_t[j]$$

or in matrix form:

$$\begin{aligned} \mathbf{x_{t+1}} &= (1-\epsilon)\hat{\mathbf{A}}\mathbf{x_t} + \epsilon \frac{1}{10} \mathbf{1}\mathbf{1}^T \mathbf{x_t} \\ &= \mathbf{P}\mathbf{x_t} \end{aligned}$$

The ranks can be calculated using iteration or eigenvalues of $\mathbf{P}$ ($\mathbf{P}$ remains a stochastic matrix, with spectral radius of 1). Both methods produce the same results (roughly 100 iterations).

SOLUTIONS (normalized to sum to 1):



"PageRank" appears in documents {C,D,G}, "teleportation" in {G,H,J}.

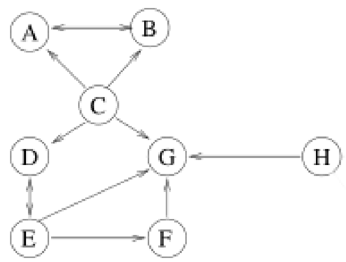
$$ranks_{0.15} = \begin{array}{ll} A & 0.1946 \\ B & 0.1946 \\ C & 0.0241 \\ D & 0.0474 \\ E & 0.0644 \\ F & 0.0423 \\ G & 0.1068 \\ H & 0.0825 \\ I & 0.1374 \\ J & 0.1058 \end{array}$$

In descending order: {A,B,I,G,J,H,E,D,F,C}

$$ranks_{0.10} = \begin{array}{ll} A & 0.2246 \\ B & 0.2246 \\ C & 0.0183 \\ D & 0.0383 \\ E & 0.0528 \\ F & 0.0342 \\ G & 0.0926 \\ H & 0.0785 \\ I & 0.1338 \\ J & 0.1021 \end{array}$$

In descending order: {A,B,I,J,G,H,E,D,F,C}

Task b)



Root Set for term **"PageRank"** is {C,D,G}.
Base Set: {A,B,C,D,E,F,G,H}

We can use the adjacency matrix **A**, and just remove the rows/columns for {I,J}:

$$\mathbf{M_{pr}} = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \end{array} \begin{array}{c} \begin{array}{cccccccc} A & B & C & D & E & F & G & H \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

Hubs **u** and authorities **a** can be solved from the update rule:

$$\mathbf{a} = \mathbf{M}^T\mathbf{u}$$
$$\mathbf{u} = \mathbf{Ma}$$

We can again solve using either iteration or first eigenvector from:

$$\mathbf{a} = \mathbf{M}^T\mathbf{u} = \mathbf{M}^T\mathbf{Ma}$$
$$\mathbf{u} = \mathbf{Ma} = \mathbf{MM}^T\mathbf{u}$$

For iteration, hubs can be initialized to constant vector.

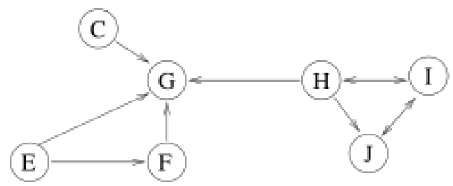
SOLUTION (normalized so that the highest score is 1)

For term **"PageRank"**, which appears in documents {C,D,G}.

$$a_{pr} = \begin{array}{cc} A & 0.4675 \\ B & 0.4675 \\ C & 0 \\ D & 0.6985 \\ E & 0 \\ F & 0.3015 \\ G & 1 \\ H & 0 \end{array}$$

$$u_{pr} = \begin{array}{cc} A & 0.1775 \\ B & 0.1775 \\ C & 1 \\ D & 0 \\ E & 0.7595 \\ F & 0.3797 \\ G & 0 \\ H & 0.3797 \end{array}$$

Task c)



Root Set for term **"teleportation"** is {G,H,J}.
Base Set: {C,E,F,G,H,I,J}

Again, using the adjacency matrix after removing rows/columns {A,B,D}.

$$
\mathbf{M_{tp}} = \begin{array}{c} \\ C \\ E \\ F \\ G \\ H \\ I \\ J \end{array}
\begin{array}{c}
\begin{array}{ccccccc} C & E & F & G & H & I & J \end{array} \\
\begin{pmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}
\end{array}
$$

(solutions in the following page)

SOLUTION (normalized so that the highest score is 1)

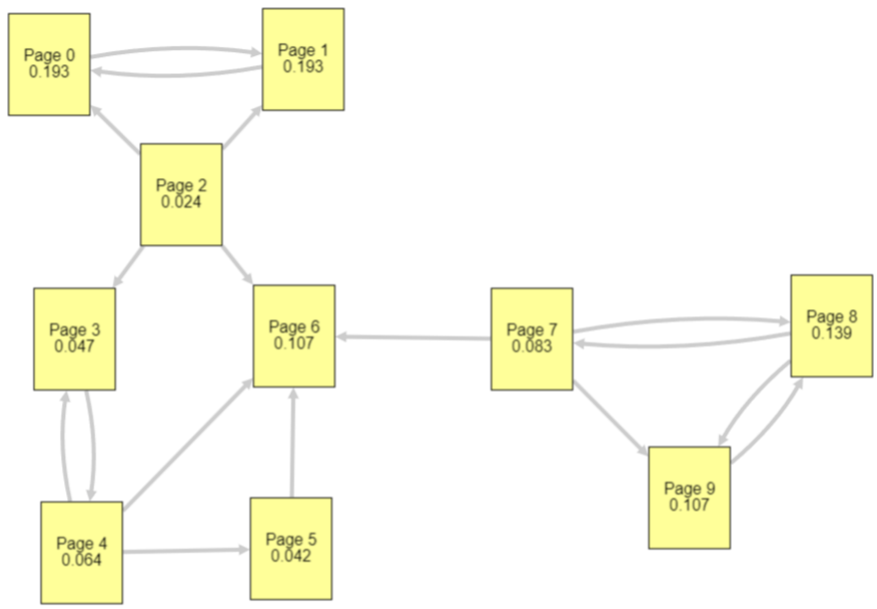
For term **"teleportation"**, which appears in documents in {G,H,J}.

$$a_{tp} = \begin{array}{cc} C & 0 \\ E & 0 \\ F & 0.2381 \\ G & 1 \\ H & 0.1179 \\ I & 0.4672 \\ J & 0.4953 \end{array}$$

$$u_{tp} = \begin{array}{cc} C & 0.5096 \\ E & 0.6309 \\ F & 0.5096 \\ G & 0 \\ H & 1 \\ I & 0.3124 \\ J & 0.2381 \end{array}$$

PageRank ONLINE

From an online PageRank tool (here), which uses teleportation factor of roughly 0.15



we get ranking (descending):

{A,B,I,G,J,H,E,D,F,C}

or, equally

{A,B,I,J,G,H,E,D,F,C}

HITS ONLINE

Online tool is available here. It gives slightly different scores, but the same ordering:

$$a_{pr} = \begin{array}{cc} A & 0.5555 \\ B & 0.5555 \\ C & 0 \\ D & 0.7777 \\ E & 0.1111 \\ F & 0.3333 \\ G & 1 \\ H & 0 \end{array}$$

$$u_{pr} = \begin{array}{cc} A & 0.2 \\ B & 0.2 \\ C & 1 \\ D & 0.1 \\ E & 0.7 \\ F & 0.4 \\ G & 0 \\ H & 0.4 \end{array}$$

$$a_{tp} = \begin{array}{cc} C & 0 \\ E & 0 \\ F & 0.2857 \\ G & 1 \\ H & 0.2857 \\ I & 0.5714 \\ J & 0.7143 \end{array}$$

$$u_{tp} = \begin{array}{cc} C & 0.5 \\ E & 0.625 \\ F & 0.5 \\ G & 0 \\ H & 1 \\ I & 0.375 \\ J & 0.25 \end{array}$$