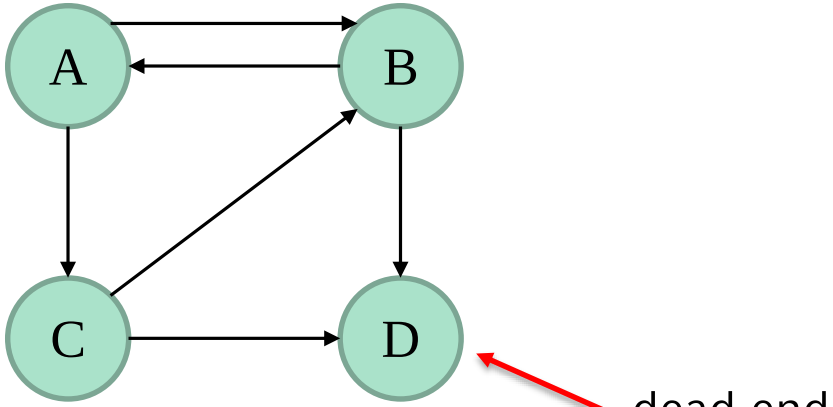
Big Data Management and Analytics
Assignment 11

Assignment 11-1

a) Explain how the PageRank algorithm avoids to get stuck in a „dead end".

What is a „dead end"?

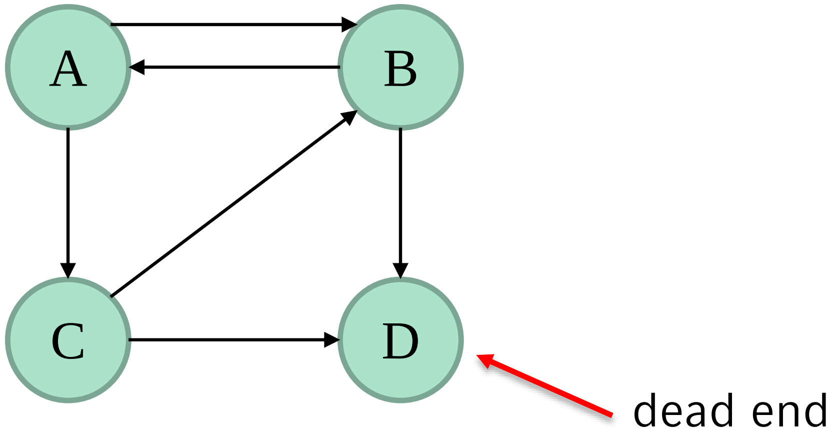
→ Pages which do not have any outgoing link



Assignment 11-1

a) Explain how the PageRank algorithm avoids to get stuck in a „dead end".

- PageRank avoids getting trapped in such „dead ends" by randomly surfing on any other page → Teleport
- Let n be the number of pages in the graph. The probability to surf on a specific page is $^1/_n$
- In the graph below, being trapped in D, the next pages A,B,C or D itself can be visited with a probability of $^1/_4$

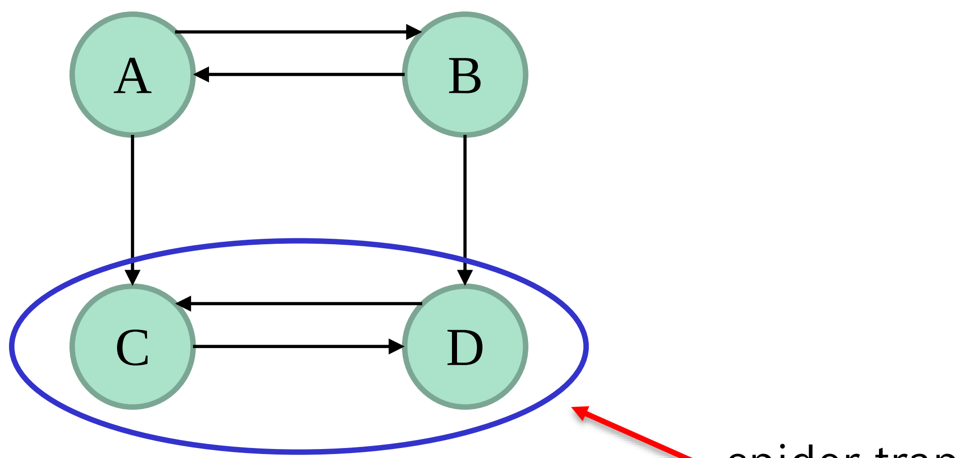


Assignment 11-1

a) Explain how the PageRank algorithm avoids to get stuck in a „dead end“.

What is a „spider trap“?

→ All outgoing links of a group of pages are **within** the group



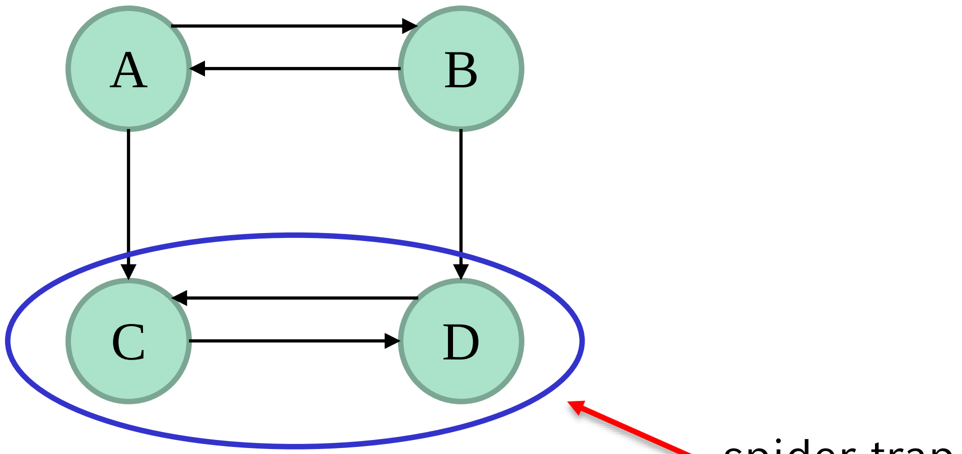
Assignment 11-1

a) Explain how the PageRank algorithm avoids to get stuck in a „dead end".

How can a spider trap being avoided?

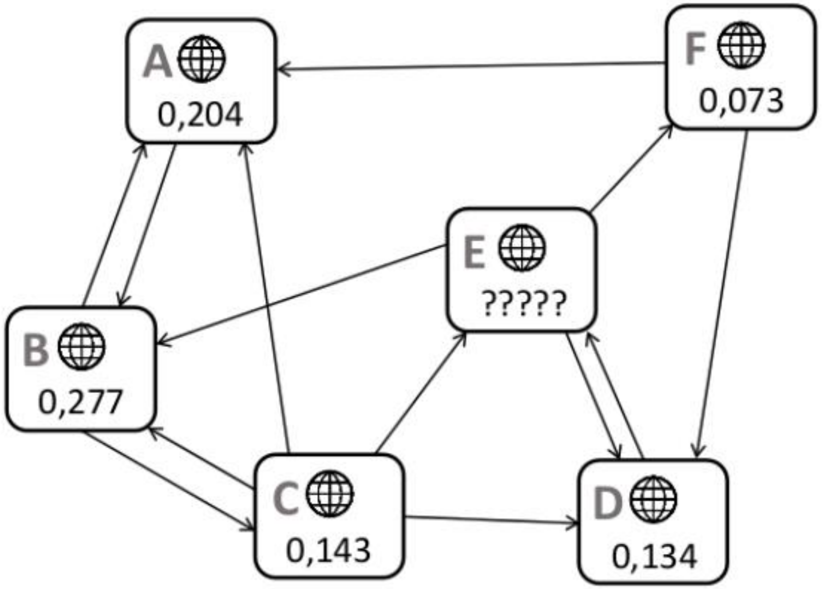
→ Follow a link with a probability of β

→ Teleporting randomly to any page is done with a probability of $(1 - \beta)$



Assignment 11-1

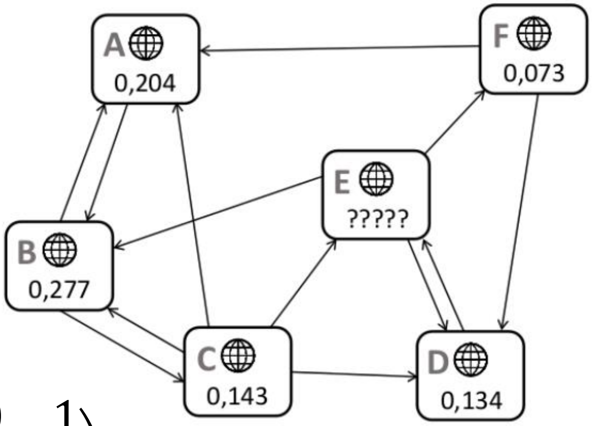
b) Given the graph below, compute the Google Matrix with $\beta = 0.85$



Assignment 11-1

b) Given the graph below, compute the Google Matrix with $\beta = 0.85$

1. Create the adjacency matrix A and A^T



$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \quad A^T = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Assignment 11-1

b) Given the graph below, compute the Google Matrix with $\beta = 0.85$

2. Create a matrix M by norming A^T

$$A^T = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\mathrm{M} = \begin{pmatrix} 0 & 1/2 & 1/4 & 0 & 0 & 1/2 \\ 1 & 0 & 1/4 & 0 & 1/3 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 1/3 & 1/2 \\ 0 & 0 & 1/4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 0 \end{pmatrix}$$

Assignment 11-1

b) Given the graph below, compute the Google Matrix with $\beta = 0.85$

2. Create a matrix M by norming A^T

$$
\mathrm{M} = \begin{pmatrix}
0 & ^1/_2 & ^1/_4 & 0 & 0 & ^1/_2 \\
1 & 0 & ^1/_4 & 0 & ^1/_3 & 0 \\
0 & ^1/_2 & 0 & 0 & 0 & 0 \\
0 & 0 & ^1/_4 & 0 & ^1/_3 & ^1/_2 \\
0 & 0 & ^1/_4 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & ^1/_3 & 0
\end{pmatrix}
$$

- This matrix is alread real, non-negative and its columns fulfill the stochastic property
- In case of a dead end these properties would not hold
- → a column would be zero → All values of that column would be set to $^1/_6$

Assignment 11-1

b) Given the graph below, compute the Google Matrix with $\beta = 0.85$

3. Compute the Google matrix $\boxed{G = \beta \cdot M + \frac{1-\beta}{n} \cdot \mathbf{1}}$

$$0.85 \cdot \begin{pmatrix} 0 & {}^1/_2 & {}^1/_4 & 0 & 0 & {}^1/_2 \\ 1 & 0 & {}^1/_4 & 0 & {}^1/_3 & 0 \\ 0 & {}^1/_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & {}^1/_4 & 0 & {}^1/_3 & {}^1/_2 \\ 0 & 0 & {}^1/_4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & {}^1/_3 & 0 \end{pmatrix} + (1 - 0.85) \begin{pmatrix} {}^1/_6 & \cdots & {}^1/_6 \\ \vdots & \ddots & \vdots \\ {}^1/_6 & \cdots & {}^1/_6 \end{pmatrix} =$$

Assignment 11-1

b) Given the graph below, compute the Google Matrix with $\beta = 0.85$

3. Compute the Google matrix $\boxed{G = \beta \cdot M + \frac{1-\beta}{n} \cdot \mathbf{1}}$

$$= \frac{1}{60} \cdot \begin{pmatrix} 0 & 25.5 & 12.75 & 0 & 0 & 25.5 \\ 51 & 0 & 12.75 & 0 & 17 & 0 \\ 0 & 25.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 12.75 & 0 & 17 & 25.5 \\ 0 & 0 & 12.75 & 51 & 0 & 0 \\ 0 & 0 & 0 & 0 & 17 & 0 \end{pmatrix} + \frac{1}{60} \begin{pmatrix} 1.5 & \cdots & 1.5 \\ \vdots & \ddots & \vdots \\ 1.5 & \cdots & 1.5 \end{pmatrix} =$$

Assignment 11-1

b) Given the graph below, compute the Google Matrix with $\beta = 0.85$

3. Compute the Google matrix $\boxed{G = \beta \cdot M + \frac{1-\beta}{n} \cdot \mathbf{1}}$

$$= \frac{1}{60} \cdot \begin{pmatrix} 1.5 & 27 & 14.25 & 1.5 & 1.5 & 27 \\ 52.5 & 1.5 & 14.25 & 1.5 & 18.5 & 1.5 \\ 1.5 & 27 & 1.5 & 1.5 & 1.5 & 1.5 \\ 1.5 & 1.5 & 14.25 & 1.5 & 18.5 & 27 \\ 1.5 & 1.5 & 14.25 & 52.5 & 1.5 & 1.5 \\ 1.5 & 1.5 & 1.5 & 1.5 & 18.5 & 1.5 \end{pmatrix}$$

Assignment 11-1

c) How can the actual PageRank values be computed by using the Google Matrix?

- The Google matrix has the stochastic property
  → There exists an eigenvector of G with the eigenvalue 1

- Looking at the eigenvalue problem $G \cdot x = x$,
  vector x is a stochastic vector which consists of the PageRank values

- For getting the eigenvector x_i to the corresponding eigenvalue λ_i
  we can solve the following equation system: $(G - \lambda_i \boldsymbol{E} x) = 0$

- In our case $\lambda_i = 1$, thus we need to compute $(G - \boldsymbol{Ex}) = 0$
  in order to get the eigenvector

Assignment 11-1

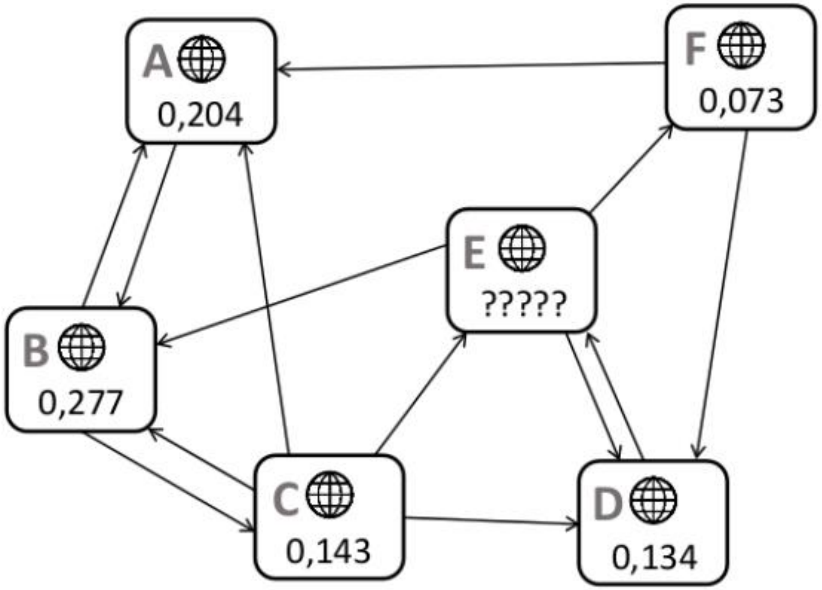
c) How can the actual PageRank values be computed by using the Google Matrix?

- Solve the following equation system $(G - \boldsymbol{E}\boldsymbol{x}) = 0$:

$$\left(\begin{array}{cccccc|c} -0.975 & 0.450 & 0.238 & 0.025 & 0.025 & 0.450 & 0.000 \\ 0.000 & -0.571 & 0.451 & 0.047 & 0.331 & 0.429 & 0.000 \\ 0.000 & 0.000 & -0.605 & 0.064 & 0.293 & 0.383 & 0.000 \\ 0.000 & 0.000 & 0.000 & -0.943 & 0.462 & 0.662 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.376 & -0.873 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \end{array}\right)$$

Assignment 11-1

d) Given the graph as seen below, compute the missing PageRank value of node E by using the PageRank equation



Assignment 11-1

d) Given the graph as seen below, compute the missing PageRank value of node E by using the PageRank equation

$$\boxed{PR_E = \frac{1-\beta}{n} + \beta \sum_{i \to E} \frac{PR_i}{d_i}}$$

$$PR_E = \frac{0.15}{6} + 0.85(\frac{0.143}{4} + 0.134)$$

$$PR_E = 0.169$$

Assignment 11-1

d) Given the graph as seen below, compute the missing PageRank value of node E by using the PageRank equation

- Alternatively it can also be computed by using the equation system in (c):

$$
\begin{array}{c}
\begin{array}{cccccc} \text{A} & \text{B} & \text{C} & \text{D} & \text{E} & \text{F} \end{array} \\
\begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \\ \text{E} \\ \text{F} \end{array}
\left(\begin{array}{cccccc|c}
-0.975 & 0.450 & 0.238 & 0.025 & 0.025 & 0.450 & 0.000 \\
0.000 & -0.571 & 0.451 & 0.047 & 0.331 & 0.429 & 0.000 \\
0.000 & 0.000 & -0.605 & 0.064 & 0.293 & 0.383 & 0.000 \\
0.000 & 0.000 & 0.000 & -0.943 & 0.462 & 0.662 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & \boxed{0.376} & \boxed{-0.873} & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000
\end{array}\right)
\end{array}
$$

$$0.376 \cdot PR_E + (-0.873) \cdot PR_F = 0$$
$$0.376 \cdot PR_E + (-0.873) \cdot 0.073 = 0$$
$$PR_E = \frac{0.873 - 0.073}{0.376}$$
$$PR_E = 0.169$$