Last time: hints about importance ranking of webpages.

1. Very naive: a page is important if there are many links to it.

2. Less naive: a page is important if there are many links to it from important pages.

Pages 1, 2, ..., K.
Importance ranking vector \( (x_1, \ldots, x_K) \)

\[
x_i = \sum_{j \text{ such that there is a link from page } j \text{ to page } i} \frac{1}{n_j} x_j
\]

all \( j \) such that there is a link from page \( j \) to page \( i \),
\( n_j = \text{total number of links from page } j \).

Informally, each page \( j \) has one vote. It gives \( \frac{1}{n_j} \) of it to each page it links to.

Another view: let \( x_p \) be the probability to be on page \( p \). Then \( \sum_{j \rightarrow i} \frac{1}{n_j} x_j \) is the probability of ending on page \( i \) in one click!

And our equation just describes the stable probability distribution after many clicks.
Example.

Naively:

1. has 2 backlinks
2. has 1 backlink
3. has 3 backlinks
4. has 2 backlinks

Our "less naive" approach:

\[ A = \begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix} \]

because 4 links to 0 and 3.

For eigenvalue 1,
every eigenvector proportional to \( \begin{pmatrix} 12 \\ 4 \\ 9 \\ 6 \end{pmatrix} \)

normalizing, get \( \begin{pmatrix} \frac{12}{31} \\ \frac{4}{31} \\ \frac{9}{31} \\ \frac{6}{31} \end{pmatrix} \) \( \approx \) \( \begin{pmatrix} 0.387 \\ 0.129 \\ 0.290 \\ 0.194 \end{pmatrix} \) so page 1 is

the most important.
Example

\[
\begin{pmatrix}
0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}
\]

\[
\begin{pmatrix}
\lambda_1 \\
\frac{1}{\lambda_1} \\
0 \\
0 \\
0
\end{pmatrix}
\text{ and }
\begin{pmatrix}
0 \\
0 \\
\frac{1}{\lambda_2} \\
\frac{1}{\lambda_2} \\
0
\end{pmatrix}
\]

are eigen-vectors

non-uniqueness $\leftrightarrow$ disconnectedness of the network
Modification

\[ M = (1-p)A + p \frac{1}{k} \left( \begin{array}{cccc} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \end{array} \right) \]

with probability \( p \) incorporate chaos (move to another page randomly)

Usually, \( p \approx 0.15 \)

\[ M \] has positive matrix elements

\( M \) is column-stochastic: numbers in each column add up to 1.
Some theory.

1. Always have eigenvalue 1.

   \[ \{ \text{Eigenvalues of } A^T \} = \{ \text{Eigenvalues of } A \} \]

   \[ \det(A - cI_n) = \det(A^T - cI_n) \]

   And \( A^T \) has eigenvalue 1 because \( (\frac{1}{1}) \) is an eigenvector.

2. If all entries of \( A \) are positive, then every eigenvector with eigenvalue 1 has all coordinates of the same sign. Otherwise

   \[ |x_i| = \sum_j A_{ij} |x_j| < \sum_j A_{ij} |x_j| \]

   \[ \sum_i |x_i| = \sum_i \sum_j A_{ij} |x_j| < \sum_i \sum_j A_{ij} |x_j| = \left[ \sum_i |x_i| \right] \text{ contradiction} \]

3. Up to proportionality, just one eigenvector for eigenvalue 1.

   Idea: If there are two linearly independent eigenvectors for eigenvalue 1, then can find a combination of these that has both positive and negative coefficients, contradicting (2).