Lecture 21: PageRank and Markov chains

Thursday, November 5, 2015 9:30 AM

Today: Spectral theory for general matrices
Stochastic matrices/Markov chains & PageRank

Recall: λ is an eigenvalue of A

$\Updownarrow$

$A\vec{v} = \lambda\vec{v}$ for a nonzero vector $\vec{v}$

$\Updownarrow$

$(A - \lambda I)\vec{v} = 0$ for $\vec{v} \neq 0$

$\Updownarrow$

$N(A - \lambda I) \neq \{0\}$.

$\Updownarrow$

$A - \lambda I$ is singular!

$\Updownarrow$

$\det(A - \lambda I) = 0$

we still need to prove this step!

Proof that $\det(A) \neq 0 \iff A$ is nonsingular:

① This is true in the case that A is upper triangular.

$$A = \begin{pmatrix} \text{(nonzero)} & & \\ & \ddots & \\ 0 & & \text{(nonzero)} \end{pmatrix}$$

A is nonsingular $\iff$ $\text{rank}(A) = n$
$\iff$ all diagonal entries are nonzero
$\iff$ $\det A$ (product of diagonal entries) nonzero

② To reduce to the upper-triangular case, use Gaussian elimination! (Sorry.)
Gaussian elimination has one basic operation:
Add a multiple of one row to another row.
How does this affect the determinant?

Lemma 1: If a row is repeated $\Rightarrow \det(A) = 0$.

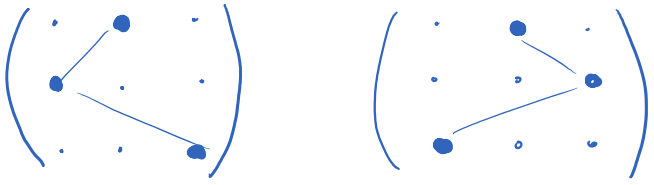
Example: $\det\begin{pmatrix} a & b \\ a & b \end{pmatrix} = ab - ba = 0$ ✓

```
>> n = 10;
>> A = randn(n-1,n);
>> A = [A; A(5,:)];
>> det(A)

ans =

  -1.6929e-14
```

Why?

$$\det(A) = \sum_{\text{permutations } \sigma} \text{sign}(\sigma) \cdot \prod_{i=1}^{n} A_{i,\sigma(i)}$$

If rows j and k are the same, then every term appears twice!



$a_{12}a_{21}a_{33}$ $\qquad$ $a_{12}a_{23}a_{31}$

the same if row 2 = row 3

But the permutations for matching terms differ by one transposition — so have opposite signs, and cancel out. ✓ □

Observation 2: The determinant function is definitely **not** linear,

e.g., $\det(10 \cdot A) = 10^n \cdot \det(A)$.

But each summand in

$$\det(A) = \sum_{\text{permutations } \sigma} \text{sign}(\sigma) \cdot \prod_{i=1}^{n} A_{i,\sigma(i)}$$

involves exactly one entry from any given row of A.

$$\Rightarrow \det\begin{pmatrix} 5a_{11} & 5a_{12} \\ a_{21} & a_{22} \end{pmatrix} = 5 \cdot \det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

and, more importantly,

$$\det\begin{pmatrix} \text{— row 1 —} \\ \text{— row 2 —} \\ \vdots \\ \text{— row n —} \end{pmatrix} + \det\begin{pmatrix} \text{— a different row 1 —} \\ \text{— row 2 —} \\ \vdots \\ \text{— row n —} \end{pmatrix}$$

$$\det\begin{pmatrix} \text{row } 1 \\ \vdots \\ \text{row } n \end{pmatrix} + \det\begin{pmatrix} \text{row } 2 \\ \vdots \\ \text{row } n \end{pmatrix}$$

$$= \det\begin{pmatrix} \text{row } 1 + \text{different row } 1 \\ \text{row } 2 \\ \vdots \\ \text{row } n \end{pmatrix}$$

Corollary: Adding one row to another does not change the determinant (by Obs. 2 and Lemma 1).

$$\det\begin{pmatrix} \text{row } 1 + \text{row } 5 \\ \text{row } 2 \\ \vdots \\ \text{row } n \end{pmatrix} = \det\begin{pmatrix} \text{row } 1 \\ \text{row } 2 \\ \vdots \\ \text{row } n \end{pmatrix} + \det\begin{pmatrix} \text{row } 5 \\ \text{row } 2 \\ \text{row } 3 \\ \vdots \\ \text{row } n \end{pmatrix}$$

$= 0$ since row 5 is repeated

$\Rightarrow$ Applying Gaussian elimination to A does not change whether $\det A = 0$ or $\det A \neq 0$.

After Gaussian elimination, $\det A' = 0 \Leftrightarrow \text{rank}(A') < n$.
Hence, before Gaussian elimination, $\det A = 0 \Leftrightarrow \text{rank}(A) < n$.

✓ □

SUMMARY OF EIGENVALUES & EIGENVECTORS

To find the eigenvalues of an $n \times n$ matrix A, find the roots λ where $\det(A - \lambda I) = 0$.

$\det(A - \lambda I)$ is a polynomial in λ of degree n

$\Rightarrow$ it can be factored

$$\det(A - \lambda I) = (\lambda - \lambda_1)^{\alpha_1} \cdot (\lambda - \lambda_2)^{\alpha_2} \cdots (\lambda - \lambda_k)^{\alpha_k}$$

with $\alpha_1 + \alpha_2 + \cdots + \alpha_k = n$.
(hence $k \leq n$)

The eigenvalues of A are $\lambda_1, \lambda_2, \ldots, \lambda_k$.

For each distinct eigenvalue λ_i,

$\dim N(A - \lambda_i I) \geq 1$ (since it is an eigenvalue)

and $\dim N(A - \lambda_i I) \leq \alpha_i$

$\Rightarrow$ There are at least k independent eigenvectors.

There are n independent eigenvectors (i.e., A is diagonalizable) if and only if $\dim N(A-\lambda_i I) = \alpha_i$ for every $i = 1, \ldots, k$

Why is $\dim N(A-\lambda_i I) \leq \alpha_i$?

Proof: We want to show that $\text{rank}(A-\lambda_i I) \geq n-\alpha_i$; the claim $\dim N(A-\lambda_i I) \leq \alpha_i$ then follows by the Rank-Nullity Theorem.

Useful claim: If U is invertible, then the eigenvalues of A are the same as those of UAU^{-1}.

Proof: If $A\vec{v} = \lambda\vec{v}$, then

$$(UAU^{-1})(U\vec{v}) = UA\vec{v}$$
$$= \lambda(U\vec{v}) \quad \checkmark$$

Now apply Gaussian elimination to A; assuming no row interchanges are required, this gives

$$A = (L)(U)$$

where L has 1s along the diagonal and U has each λ_i on its diagonal α_i times. (Think about it....)

Thus we can also write

$$A = L U' L^{-1},$$

this means applying the same operations to the columns of U as were done to the rows of A in Gaussian elimination. It leaves the diagonal entries unchanged.

But then

$$\text{rank}(A-\lambda_i I) = \text{rank}(U' - \lambda_i I)$$
$$\geq \#\text{ of nonzero entries along the diagonal, since } U'-\lambda_i I \text{ is upper triangular}$$
$$= n-\alpha_i. \quad \checkmark \quad \square$$

Note: The recipe

1. Compute $p(\lambda) = \det(A - \lambda I)$
2. Find its factors $\lambda_1, \ldots, \lambda_k \Rightarrow$ eigenvalues of A
3. Compute nullspaces $N(A - \lambda_i I) \Rightarrow$ eigenspaces

works in theory, and for 2×2, 3×3 matrices.

But it quickly becomes impractical. Writing down $p(\lambda)$, and then finding its roots, is very time-consuming.

Next time we'll learn a faster way...

In fact, a common way of solving for the roots of a polynomial

$$p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$$

is to compute the eigenvalues of a matrix!

For

$$A = \begin{pmatrix} 0 & 0 & \cdots & -a_0 \\ 1 & 0 & & -a_1 \\ & 1 & \ddots & \vdots \\ 0 & & 1 & -a_{n-1} \end{pmatrix},$$

$$\det(A - \lambda I) = p(\lambda).$$

Corollary: A and its transpose A^T have the same eigenvalues.

Proof: $\mathrm{Det}(A - \lambda I) = \mathrm{Det}(A^T - \lambda I)$ since $\mathrm{Det}\, M = \mathrm{Det}\, M^T$.

$\Rightarrow$ same roots (eigenvalues) ✓ □

But the eigenvectors can be different!

Google

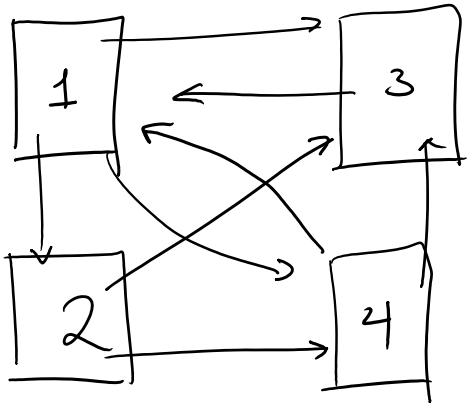
Search engine problems:

- Crawl and index the web
- Decide what ads to show you for each search query
- Decide what web pages are "important" to show you

"PageRank" web page ranking

- Rates every web page, compared to all others (independent of any search query)

Example:



Which webpage is most important?

Page	#incoming links
1	2
2	1
3	3
4	2

⇒ Page 3 > Page 1
" Page 4
> Page 3

But Page 1 should be > Page 4, since its extra link comes from 3, which is more important than 2.

Proposal:

Importance of page k is

$$x_k = \sum_{\substack{\text{pages } j \\ \text{linking to } k}} x_j$$

This is no good! First of all, there is no solution!

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

But also, this kind of ranking would make pages with more external links more influential, letting them game the system.

Better proposal:

Importance of page k is

$$x_k = \sum_{\substack{\text{pages } j \\ \text{linking to } k}} \frac{x_j}{\#\text{ of } j\text{'s links}}$$

Basically, a pages votes are divided evenly among its outgoing links. [Self links don't count.]

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

Importance scores are an **eigenvalue-one eigenvector** of the **link matrix**.

In this case,

```
>> A = [
0   0   1   1/2;
1/3 0   0   0;
1/3 1/2 0   1/2;
1/3 1/2 0   0
];

x = null(A - eye(4));
x = x / sum(x)

x =

   3.8710e-01
   1.2903e-01
   2.9032e-01
   1.9355e-01
```

4×4 identity

here I scale so the ℓ_1 norm is 1 (of course this is arbitrary)

```
>> format rat
>> x

x =

   12/31
    4/31
    9/31
    6/31
```

Check the answer:

```
>> A * [12 4 9 6]'

ans =

   12
    4
    9
    6
```

Note: Now page 1 is most important!

Definition: A matrix is stochastic if its entries are ≥ 0, and sum to 1 in each column.

Theorem: Every stochastic matrix A has 1 as an eigenvalue.

Proof: $A^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, and A's eigenvalues are the same as A^T. □

Intuition/philosophy behind PageRank:

Browsers spend more time on more important pages.

More precisely:

Consider a typical user, randomly browsing the web. The pages that he visits more often are more important.

Claim: Let $\vec{p}$ be a probability distribution (eg., over webpages), A a stochastic matrix (eg., the link matrix). Then $A\vec{p}$ is a probability distribution.

Proof:

$\vec{p}$ is prob. distn $\Rightarrow$ all $p_j \geq 0$ and $\|p\|_1 = \sum_j |p_j| = 1$.

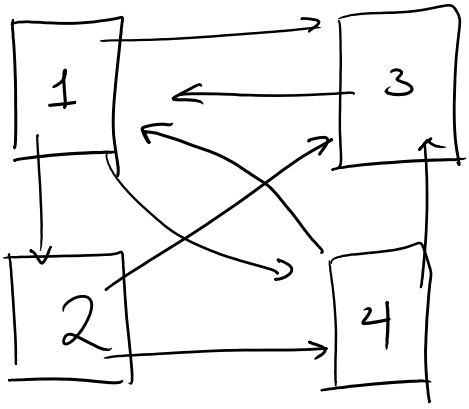
$$\|A\vec{p}\|_1 = \sum_i (Ap)_i$$
$$= \sum_{ij} A_{ij} p_j$$
$$= \sum_j p_j \left(\sum_i A_{ij}\right)$$
$$= \sum_j p_j \quad \text{since } A \text{ is stochastic}$$
$$= 1 \ \checkmark$$

□

Model: From the current webpage, user selects

uniformly a random outgoing link, and follows it.

$$\begin{pmatrix} p_1(t+1) \\ p_2(t+1) \\ p_3(t+1) \\ p_4(t+1) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix} \begin{pmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \\ p_4(t) \end{pmatrix}$$



An eigenvalue-one eigenvector — $A\vec{p} = \vec{p}$ — is a steady-state/equilibrium distribution.

Definition: A square, stochastic matrix (columns add to 1) is also called a **Markov chain**.

prob. dist^n at time 0 —multiply by A→ prob. dist^n at time 1 —multiply by A→ prob. dist^n at time 2 → .....

⇒ prob. dist^n at time 0 —A^t→ prob. dist^n at time t

and equilibrium distribution —multiply by A→ equilibrium distribution

[Most things we're talking about hold for any Markov chain, not just the link matrix.]

Exercise 1: Prove that the product of stochastic matrices is also stochastic.

(Homework.)

Obvious extension: Not all links are equally likely to be followed. So don't weight them evenly.

top stories

Mathematical problems with this "importance" measure:

① The link matrix is not necessarily stochastic!

webpage with no outgoing links!

$$A = \begin{pmatrix} 0 & 0 \\ \cdot & \cdot \end{pmatrix}$$

↑ sums to 0

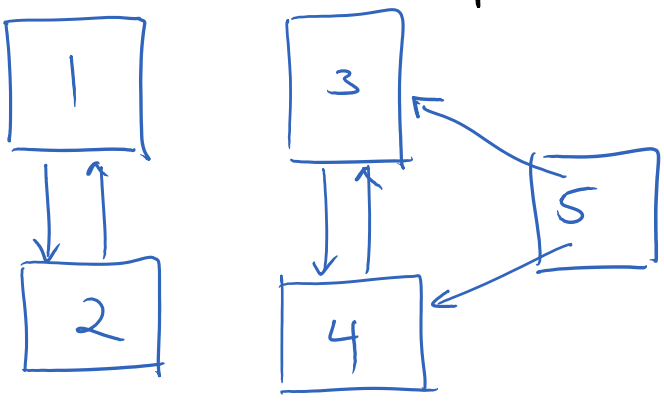
It is substochastic.

The theory is similar; one can prove that the largest eigenvalue of A is $\lambda \in [0,1]$, and use the corresponding eigenvector to rank pages. See the Perron-Frobenius Theorem.

(2) The web is disconnected

$\Rightarrow$ multiple independent eigenvalue-one e-vectors ($\dim N(A-I) > 1$)

$\Rightarrow$ rating is not uniquely defined

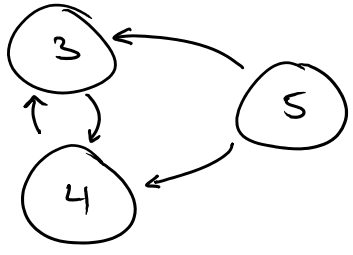


$$A = \left(\begin{array}{cc|ccc} 0 & 1 & & & \\ 1 & 0 & & & \\ \hline & & 0 & 1 & 1/2 \\ & & 1 & 0 & 1/2 \\ & & 0 & 0 & 0 \end{array}\right) \Rightarrow \text{any combination of } \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 0 \\ 1/2 \\ 1/2 \\ 0 \end{pmatrix} \text{ works!}$$

In general, $\dim N(A-I) \geq$ # of connected components

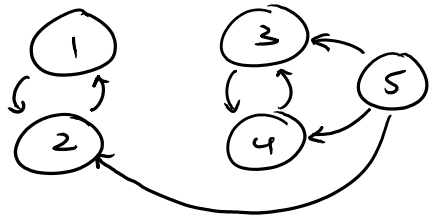
e.g., $A = \left(\begin{array}{c|c|c} A_1 & 0 & 0 \\ \hline 0 & A_2 & 0 \\ \hline 0 & 0 & A_3 \end{array}\right)$ ← disconnected webs

Exercise 2: Why is page 5's importance 0?



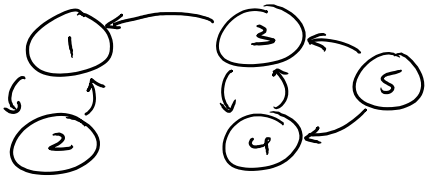
Exercise 3: What is $\dim N(A-I)$ for

Exercise 3: What is $\dim N(A-I)$ for



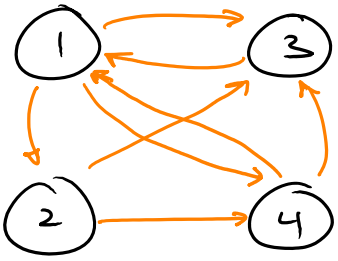
?

What about

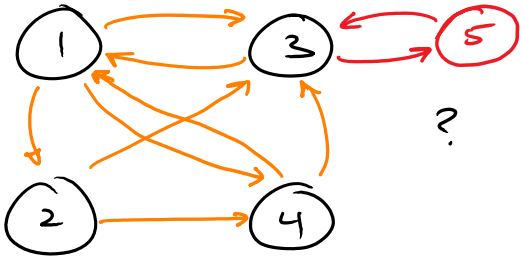


?

Exercise 4: How is importance changed from



to



?

$\begin{pmatrix} 12 \\ 4 \\ 9 \\ 6 \end{pmatrix}$ ← page 1 most important

```
>> A = [
0   0   1/2 1/2   0;
1/3 0   0   0     0;
1/3 1/2 0   1/2   1;
1/3 1/2 0   0     0;
0   0   1/2 0     0
];
x = null(A - eye(5));
x = x / sum(x)

x =

   12/49
    4/49
   18/49
    6/49
    9/49
```

Now page 3 is the most important!

Solution: CHANGE THE LINK MATRIX

Let $J = \frac{1}{n}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$

and $M = (1-\alpha)A + \alpha J$.

$\alpha = 0$: $M = A$

$\alpha = 1$: $M = J$

(ignore the links and choose a uniformly

random webpage ⇒ all are equally important)

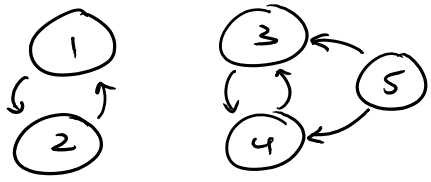
Google used $\alpha = 0.15$.

Interpretation:

With 85% probability choose a random outgoing link.

With 15% probability, restart at a uniformly random page.

Example:




```
>> A = [
0 1 0 0   0;
1 0 0 0   0;
0 0 0 1 1/2;
0 0 1 0 1/2;
0 0 0 0   0
];
n = 5;
J = ones(n,n)/n;
alpha = 0.15;
M = (1-alpha) * A + alpha * J;
x = null(M-eye(n));
x = x / sum(x)
```

← it is one-dimensional now!

```
x =

   1/5
   1/5
  57/200
  57/200
   3/100
```

1/5, 1/5 } sums to 2/5

57/200, 57/200, 3/100 } sums to 3/5

WHY?

Observe: $M = (1-\alpha)A + \alpha J$ is still stochastic.

And every entry is > 0.

Exercise 5: If nobody links to a page, its importance is α/n.

(Think about it.)

Theorem: Let M be any matrix that is:

1. Stochastic: $\sum_i M_{ij} = 1 \quad \forall i$
2. Entry-wise positive: $M_{ij} > 0 \quad \forall i,j$

Then, up to scaling, M has a unique e-value 1 e-vector, ie., $\dim N(M-I) = 1$.

Corollary: PageRank is uniquely defined.

Proof:

Lemma 1: If M is stochastic, entry-wise positive, then any e-value 1 e-vector has either all coefficients ≥ 0, or all coefficients ≤ 0.

Intuition: $M\vec{x} = \vec{x} \Rightarrow \|M\vec{x}\|_1 = \|\vec{x}\|_1$.

But if there are opposite signs, cancellation will make $\|M\vec{x}\|_1 < \|x\|_1$.

Proof:

Assume otherwise; let $\vec{x}$ satisfy $M\vec{x} = \vec{x}$, with some entries >0 and some <0.

$$\|M\vec{x}\|_1 = \sum_i \left| \sum_j M_{ij} x_j \right|$$

$$< \sum_i \sum_j |M_{ij} x_j| \quad \text{triangle inequality}$$

strict since some terms are >0, others <0 (since $M_{ij} > 0$)

$$= \sum_{ij} M_{ij} |x_j|$$

$$= \sum_j |x_j| \left(\sum_i M_{ij} \right)$$

$$= \sum_j |x_j| \quad \text{(since } \textstyle\sum_i M_{ij} = 1 \text{, } M \text{ is stochastic)}$$

a contradiction ✗ □

Lemma 2: If $\vec{x}, \vec{y} \in \mathbb{R}^n$ are linearly independent vectors, then some linear combination $\alpha\vec{x} + \beta\vec{y}$ has both >0 and <0 components.

Proof:

Assume $\vec{x}$ and $\vec{y}$ both have only ≥ 0 components.
(Otherwise, just set $\alpha = 1$ and $\beta = 0$, or vice versa.)
$\Rightarrow \|x\|_1 > 0$ and $\|y\|_1 > 0$.

Consider $\vec{v} = \|y\|_1 \cdot \vec{x} - \|x\|_1 \cdot \vec{y}$

$$\sum_i (\|y\|_1 \cdot x - \|x\|_1 y)_i = \|y\|_1 \sum_i x_i - \|x\|_1 \sum_i y_i$$
$$= 0$$

$\vec{v} \neq 0$ since $\vec{x}$ and $\vec{y}$ are linearly independent.
$\therefore$ Cancellation means it must have mixed-sign entries. □

Theorem: Let M be any matrix that is:
1. Stochastic: $\sum_i M_{ij} = 1 \quad \forall i$
2. Entry-wise positive: $M_{ij} > 0 \quad \forall i,j$

Then, up to scaling, M has a unique e-value 1 e-vector, ie., $\dim N(M-I) = 1$.

Proof:
If M has two lin. indep. e-vectors with e-value 1, then the combination

$$\alpha\vec{x} + \beta\vec{y}$$

from Lemma 2 is still an e-value 1 e-vector, contradicting Lemma 1. ✓ □

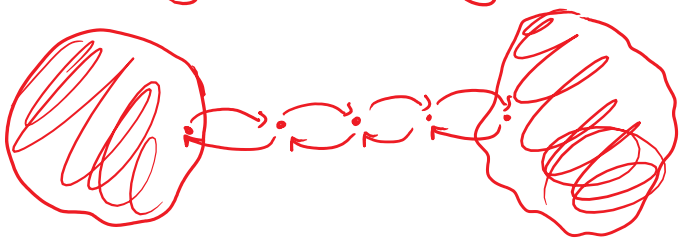
Exercise 6: Prove that if the graph is "strongly connected," ie., there is a directed path from any vertex to any other, then $\dim N(A-I) = 1$.
(Homework exercise.)

Note: By mixing the link matrix A with J,

$$M = (1-\alpha)A + \alpha J,$$

we actually solve two problems:
① Now importance score is unique
(essentially, because everything is connected)
② For nearly disconnected graphs, eg.,

② For nearly disconnected graphs, eg.,



it is much easier/faster to find the eigenvector for M than for A....

(essentially because the distribution converges faster to the equilibrium distribution)

THE "POWER METHOD" FOR FINDING THE LARGEST EIGENVALUE EIGENVECTOR

Let A be a diagonalizable matrix (for simplicity) with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.
Sort them so $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$.

① Start with a "generic" vector $\vec{x}_0$

② Repeat for $t = 1, 2, 3, \ldots$

Let $\vec{x}_t = A\vec{x}_{t-1}$.

Analysis:

Expand $\vec{x}_0$ in the basis of eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$:

$$\vec{x}_0 = \sum_{j=1}^{n} c_j \cdot \vec{v}_j$$

$$\Rightarrow A\vec{x}_0 = \sum_j c_j \lambda_j \vec{v}_j$$

$$\Rightarrow A^t x_0 = \sum_j c_j \lambda_j^t \vec{v}_j$$

↑ this is dominated by the largest λ_j's

$$= \lambda_1^t \cdot \left(c_1 \vec{v}_1 + \sum_{j>1} c_j \left(\frac{\lambda_j}{\lambda_1}\right)^t \vec{v}_j \right)$$

Provided $c_1 > 0$ and $|\lambda_2| < |\lambda_1|$, this will converge exponentially quickly to a multiple of $\vec{v}_1$.

It converges faster if $|\lambda_2/\lambda_1|$ is smaller.

Typically, one sets

$$\vec{x}_t = \frac{A\vec{x}_{t-1}}{\|A\vec{x}_{t-1}\|} = \frac{A^t \vec{x}_0}{\|A^t \vec{x}_0\|}$$

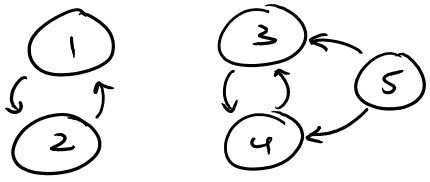
← this exact renormalization is not important; you just want to avoid overflowing the registers

to keep things from blowing up.

If the initial vector $\vec{x}_0$ is chosen randomly, it is unlikely that $c_1 = 0$.

What happens if $\lambda_1 = \lambda_2$?

Example:



from before:

```
>> A = [
0 1 0 0   0;
1 0 0 0   0;
0 0 0 1 1/2;
0 0 1 0 1/2;
0 0 0 0   0
];
n = 5;
J = ones(n,n)/n;
alpha = 0.15;
M = (1-alpha) * A + alpha * J;
x = null(M-eye(n));
x = x / sum(x)

x =
```

$$\begin{pmatrix} 1/5 \\ 1/5 \\ 57/200 \\ 57/200 \\ 3/100 \end{pmatrix}$$

in Mathematica:

```
In[54]:= x // N        ← the ev-1 evector
Out[54]= {0.2, 0.2, 0.285, 0.285, 0.03}

In[13]:= X = {{.24, .31, .08, .18, .19}};
         For[t = 1, t ≤ 50, t++,
           X = Append[X, M.X[[t]]];
          ];

In[55]:= X // MatrixForm
Out[55]//MatrixForm=
```

t=0	0.24	0.31	0.08	0.18	0.19
1	0.2935	0.234	0.26375	0.17875	0.03
2	0.2289	0.279475	0.194688	0.266938	0.03
3	0.267554	0.224565	0.269647	0.208234	0.03
4	0.22088	0.257421	0.219749	0.27195	0.03
5	0.248808	0.217748	0.273907	0.229537	0.03
6	0.215086	0.241486	0.237856	0.275571	0.03
⋮	0.235263	0.212823	0.276986	0.244928	0.03
	0.2109	0.229974	0.250939	0.278188	0.03
	0.225478	0.209265	0.27921	0.256048	0.03
	0.207875	0.221656	0.260391	0.280078	0.03
	0.218408	0.206694	0.280816	0.264082	0.03
	0.20569	0.215647	0.26722	0.281444	0.03
	0.2133	0.204836	0.281977	0.269887	0.03
	0.204111	0.211305	0.272154	0.282431	0.03
	0.209609	0.203494	0.282816	0.274081	0.03
	0.20297	0.208168	0.275719	0.283144	0.03
	0.206942	0.202525	0.283422	0.277111	0.03
	0.202146	0.205901	0.278294	0.283659	0.03
	0.205016	0.201824	0.28386	0.2793	0.03
	0.20155	0.204264	0.280155	0.284031	0.03
	0.203624	0.201318	0.284176	0.280882	0.03

(since $\|M\vec{p}\|_1 = \|\vec{p}\|_1 = 1$, there was no need)

since $\|v^{(t)}\|_1 = \|p\|_1 = 1$, there was no need to renormalize at each step

0.205016	0.201824	0.28386	0.2793	0.03
0.20155	0.204264	0.280155	0.284031	0.03
0.203624	0.201318	0.284176	0.280882	0.03
0.20112	0.20308	0.2815	0.2843	0.03
0.202618	0.200952	0.284405	0.282025	0.03
0.200809	0.202226	0.282471	0.284494	0.03
0.201892	0.200688	0.28457	0.28285	0.03
0.200585	0.201608	0.283173	0.284635	0.03
0.201367	0.200497	0.284689	0.283447	0.03
0.200422	0.201162	0.28368	0.284736	0.03
0.200988	0.200359	0.284776	0.283878	0.03
0.200305	0.200839	0.284046	0.284809	0.03
0.200713	0.200259	0.284838	0.284189	0.03
0.200221	0.200606	0.284311	0.284862	0.03
0.200515	0.200187	0.284883	0.284414	0.03
0.200159	0.200438	0.284502	0.2849	0.03
0.200372	0.200135	0.284915	0.284577	0.03

Here are the errors:

```
Table[Plus @@ Abs /@ (X[[t + 1]] - x), {t, 0, 50}]
Table[%[[t + 1]] / %[[t]], {t, 1, 50}]
```

l_1 errors from exact eigenvector

```
{0.62, 0.255, 0.21675, 0.184237, 0.156602, 0.133112, 0.113145, 0.0961731,
 0.0817472, 0.0694851, 0.0590623, 0.050203, 0.0426725, 0.0362716, 0.0308309,
 0.0262063, 0.0222753, 0.018934, 0.0160939, 0.0136798, 0.0116279, 0.00988368,
 0.00840113, 0.00714096, 0.00606982, 0.00515934, 0.00438544, 0.00372763,
 0.00316848, 0.00269321, 0.00228923, 0.00194584, 0.00165397, 0.00140587,
 0.00119499, 0.00101574, 0.000863381, 0.000733874, 0.000623793, 0.000530224,
 0.00045069, 0.000383087, 0.000325624, 0.00027678, 0.000235263, 0.000199974,
 0.000169978, 0.000144481, 0.000122809, 0.000104388, 0.0000887294}

{0.41129, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85,
 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85,
 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85,
 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85}
```

↑ errors drop exponentially quickly

$$\text{error}(t) \approx 0.85 \times \text{error}(t-1).$$

Note: Since the web is sparse, most entries of A will be 0, so it is very fast to compute

$$\vec{x}_{t+1} = M\vec{x}_t = (1-\alpha)A\vec{x}_t + \alpha \underbrace{J\vec{x}_t}_{\|\vec{x}_t\|_1 \left(\frac{1}{n}\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\right)}$$

PROOF THAT IT CONVERGES QUICKLY

Intuition:

Let $\vec{\pi}$ be the stationary distribution, $A\vec{\pi} = \vec{\pi}$.

Let $\vec{v}$ be any probability distribution.

$$\vec{v} = \vec{\pi} + (\vec{v} - \vec{\pi}) \quad \{\text{error from } \vec{\pi}$$

The intuition is that if $\vec{v} \neq \vec{\pi}$, then $\vec{v} - \vec{\pi}$ has both >0 and <0 entries (since $\sum_i (v-\pi)_i = \sum_i v_i - \sum_i \pi_i = 1-1 = 0$).
Therefore there will be cancellation in the error term when apply M:

$$M\vec{v} = M\vec{\pi} + M(\vec{v}-\vec{\pi})$$
$$= \vec{\pi} + M(\vec{v}-\vec{\pi})$$

Proposition: Let M be any stochastic matrix with $M_{ij} > 0$ for all i,j. Let $\vec{\pi}$ be the unique probability distn for which $M\vec{\pi} = \vec{\pi}$. (We proved earlier that such $\vec{\pi}$ exists and is unique.) Let $\vec{v}$ be any probability distribution. Then

$$\| M\vec{v} - \vec{\pi} \|_1 \leq (1-m) \| v - \pi \|_1$$

where

$$m = n \times \min_{i,j} M_{ij}$$

Note: In our application $\min_{i,j} M_{ij} = \frac{\alpha}{n}$, so $m = \alpha = 0.15$.
$\Rightarrow$ error drops by $\geq 15\%$ each step (as we saw)

Corollary: $\vec{\pi} = \lim_{t \to \infty} M^t \vec{v}$ (no matter what $\vec{v}$ is)

Proof:
Split M into two pieces,

$$M = m \cdot J + (M - mJ)$$

($J = \frac{1}{n}\begin{pmatrix} 1 & 1 & \cdots \\ 1 & 1 & \cdots \\ \vdots & & \end{pmatrix}$)

Observe: $(M-mJ)_{ij} = M_{ij} - \min_{k,\ell} M_{k,\ell} \geq 0 \ \forall i,j$

(In our application, $m = \alpha$, so this is exactly $M = \alpha J + (1-\alpha)A$.)

$$\| Mv - \pi \|_1 = \| M(v-\pi) \|_1 \quad \text{since } M\pi = \pi$$
$$\leq \| mJ(v-\pi) \|_1 + \| (M-mJ)(v-\pi) \|_1 \quad \text{triangle inequality } (\|a+b\|_1 \leq \|a\|_1 + \|b\|_1)$$

$mJ(v-\pi) = 0$ since $Jv = J\pi$ = uniform distn

$$= \| (M-mJ)(v-\pi) \|_1$$
$$= \sum_i \left| ((M-mJ)(v-\pi))_i \right|$$

← this is where the cancellation occurs!!!

} from here on, nothing really

$$= \sum_i \left| \sum_j (M - mJ)_{ij} (v_j - \pi_j) \right|$$

$$\leq \sum_{ij} (M - mJ)_{ij} \cdot |v_j - \pi_j| \quad \text{triangle inequality}$$

$$= \sum_j |v_j - \pi_j| \cdot \underbrace{\sum_i (M - mJ)_{ij}}_{=}$$

$$\underbrace{\sum_i M_{ij}}_{=1} - m \underbrace{\sum_i J_{ij}}_{=1} = 1 - m$$

$$= (1-m) \cdot \|v - \pi\|_1 \quad \checkmark$$

□

Question: How should α be chosen?

Larger α $\Rightarrow$ faster convergence

Smaller α $\Rightarrow$ PageRank score depends more on the graph.

Observe: Not all M we get will be diagonalizable, eg.,

$$M = (1-\alpha) \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 \\ 1 & 1/2 & 0 \end{pmatrix} + \alpha \cdot \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

is not.