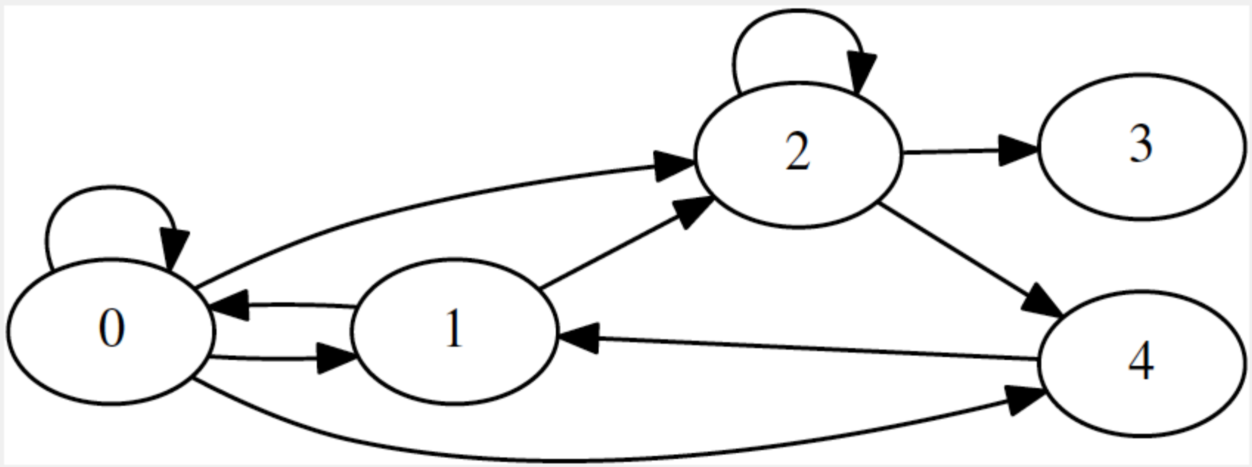
Graphs and Markov chains

Graphs as matrices

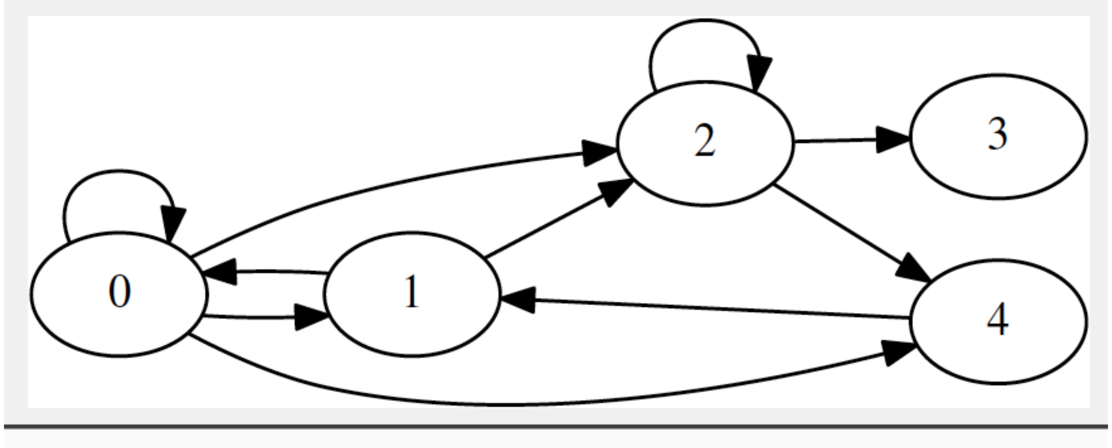
How could this (directed) graph be written as a matrix?



If there is an edge (arrow) from node i to node j, then $A_{ji} = 1$ (otherwise zero)

	0	1	2	3	4
0	1	1	0	0	0
1	1	0	0	0	1
2	1	1	1	0	0
3	0	0	1	0	0
4	1	0	1	0	0

Adjacency Matrix



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

Matrix-vector multiplication:

$$\boldsymbol{b} = \boldsymbol{A}\,\boldsymbol{x} = x_1\,\boldsymbol{A}[:,1] + x_2\,\mathbf{A}[:,2] + \cdots + x_j\,\mathbf{A}[:,j] + \cdots + x_n\,\mathbf{A}[:,n]$$

Contain all the nodes that are reachable from node $\boldsymbol{j}$

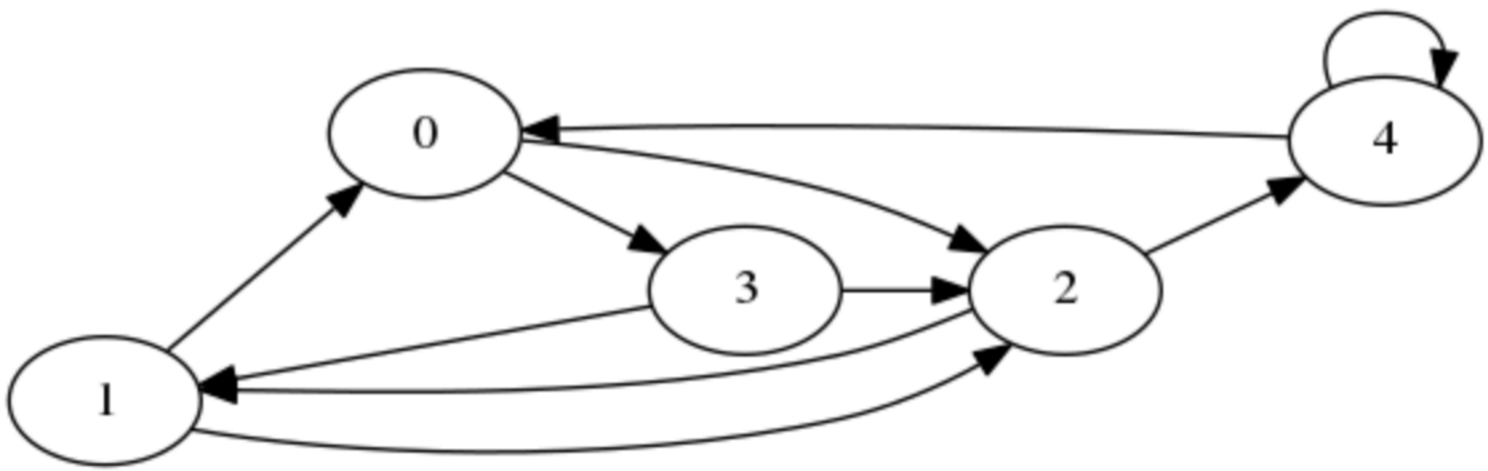
Hence, if we multiply $\boldsymbol{A}$ by the $\boldsymbol{u}_i$ unit vector, we get a vector that indicates all the nodes that are reachable by node $\boldsymbol{i}$. For example,

$$\boldsymbol{A}\,\boldsymbol{u}_2 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Iclicker question

Adjacency Matrix

For the graph given below, which is the adjacency matrix, assuming that all edge weights are 1?



A)

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

B)

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

C)

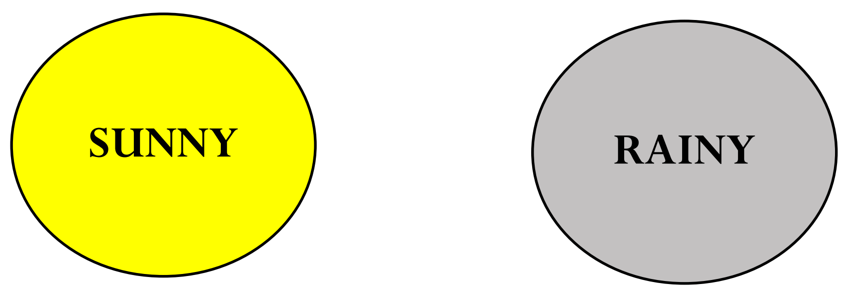
$$\begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

D)

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Using graphs to represent the transition from one state to the next

After collecting data about the weather for many years, you observed that the chance of a rainy day occurring after a rainy day is 50% and that the chance of a rainy day after a sunny day is 10%.

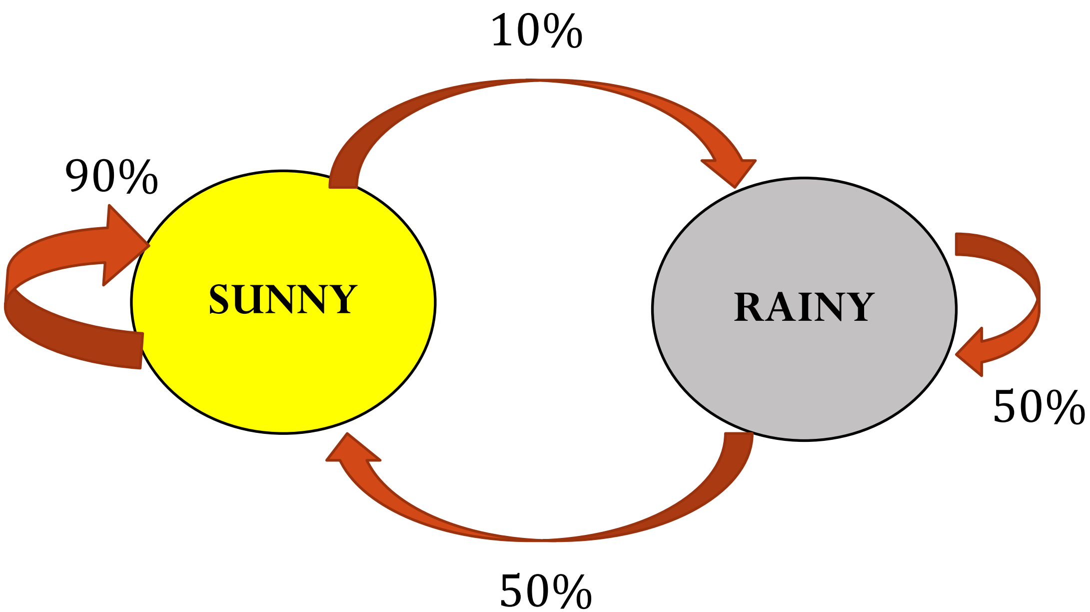


The graph can be represented as an adjacency matrix, where the edge weights are the probabilities of weather conditions (transition matrix)

	Sunny	Rainy
Sunny		
Rainy		

Transition (or Markov) matrices

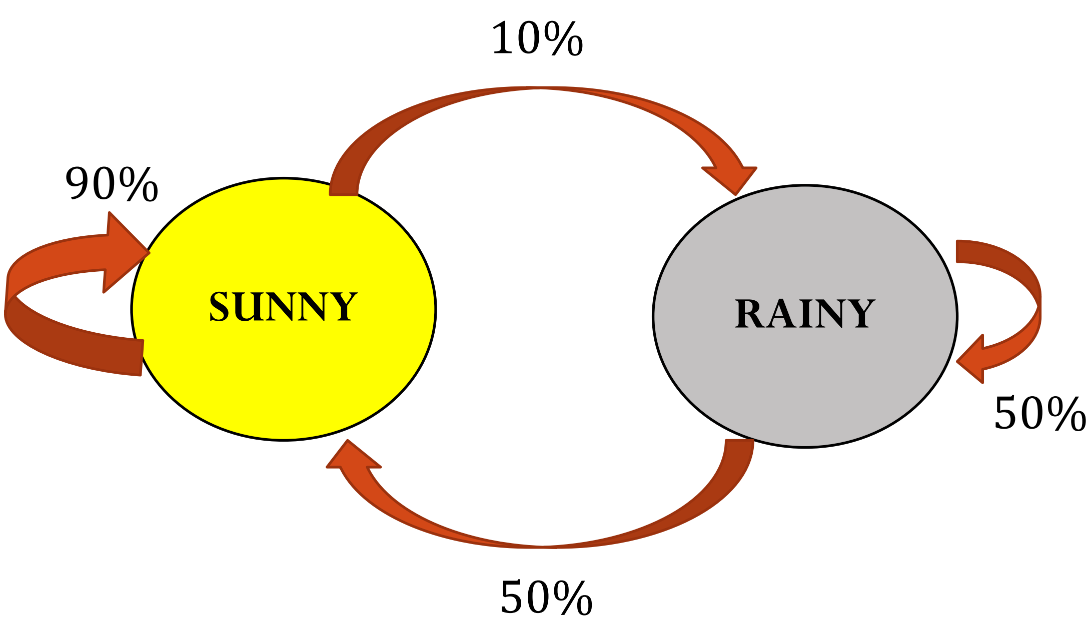
- Note that **only** the most recent state matters to determine the probability of the next state (in this example, the weather predictions for tomorrow will only depend on the weather conditions of today) – memoryless process!
- This is called the **Markov property**, and the model is called a **Markov chain**



	Sunny	Rainy
Sunny	0.9	0.5
Rainy	0.1	0.5

Transition (or Markov) matrices

- The transition matrix describe the transitions of a Markov chain. Each entry is a non-negative real number representing a probability.
- *(I,J)* entry of the transition matrix has the probability of transitioning from state *J* to state *I*.
- Columns add up to one.



	Sunny	Rainy
Sunny	0.9	0.5
Rainy	0.1	0.5

Iclicker question

http://setosa.io/ev/markov-chains/

The weather today is sunny. What is the probability of a sunny day on Saturday?

A) 81%
B) 86%
C) 90%
D) 95%

Demo "Weather predictions"

What if I want to know the probability of days that are sunny in the long run?

What if I want to know the probability of days that are sunny in the long run?

- Initial guess for weather condition on day 1: $\boldsymbol{x}_0$
- Use the transition matrix to obtain the weather probability on the following days:

$$\boldsymbol{x}_1 = \boldsymbol{A}\,\boldsymbol{x}_0 \qquad \boldsymbol{x}_2 = \boldsymbol{A}\,\boldsymbol{x}_1 \qquad \cdots \qquad \boldsymbol{x}_5 = \boldsymbol{A}\,\boldsymbol{x}_4$$

- Predictions for the weather on more distant days are increasingly inaccurate.
- What does this look like? Power iteration method!
- Power iteration method converges to ***steady-state*** vector, that gives the weather probabilities in the long-run.

$$\boldsymbol{x}^* = \boldsymbol{A}\,\boldsymbol{x}^*$$

  $\boldsymbol{x}^*$ is the eigenvector corresponding to eigenvalue $\lambda = 1$
- This "long-run equilibrium state" is reached regardless of the current state.

How can we show that the largest eigenvalue of the Markov Matrix is one?

If $\boldsymbol{A}$ is a Markov Matrix (only positive entries and the **columns sum to one**), we know that 1 is an eigenvalue for $\boldsymbol{A}$, since $\boldsymbol{e} = [1,1,..,1]$ is an eigenvector associated with 1.

$$\boldsymbol{A}\boldsymbol{e} = \boldsymbol{A}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} = \begin{bmatrix}\boldsymbol{A}[0,:]\cdot\boldsymbol{e}\\\boldsymbol{A}[1,:]\cdot\boldsymbol{e}\\\vdots\\\boldsymbol{A}[n-1,:]\cdot\boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\sum_{j=0}^{n-1}\boldsymbol{A}[0,j]\\\sum_{j=0}^{n-1}\boldsymbol{A}[1,j]\\\vdots\\\sum_{j=0}^{n-1}\boldsymbol{A}[n-1,j]\end{bmatrix} = \begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}$$

We still need to show that all the eigenvalues satisfy $|\lambda| \leq 1$, if we denote (λ, x) an eigenpair of the matrix $\boldsymbol{A}$, such that

$$|\lambda| = \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$$

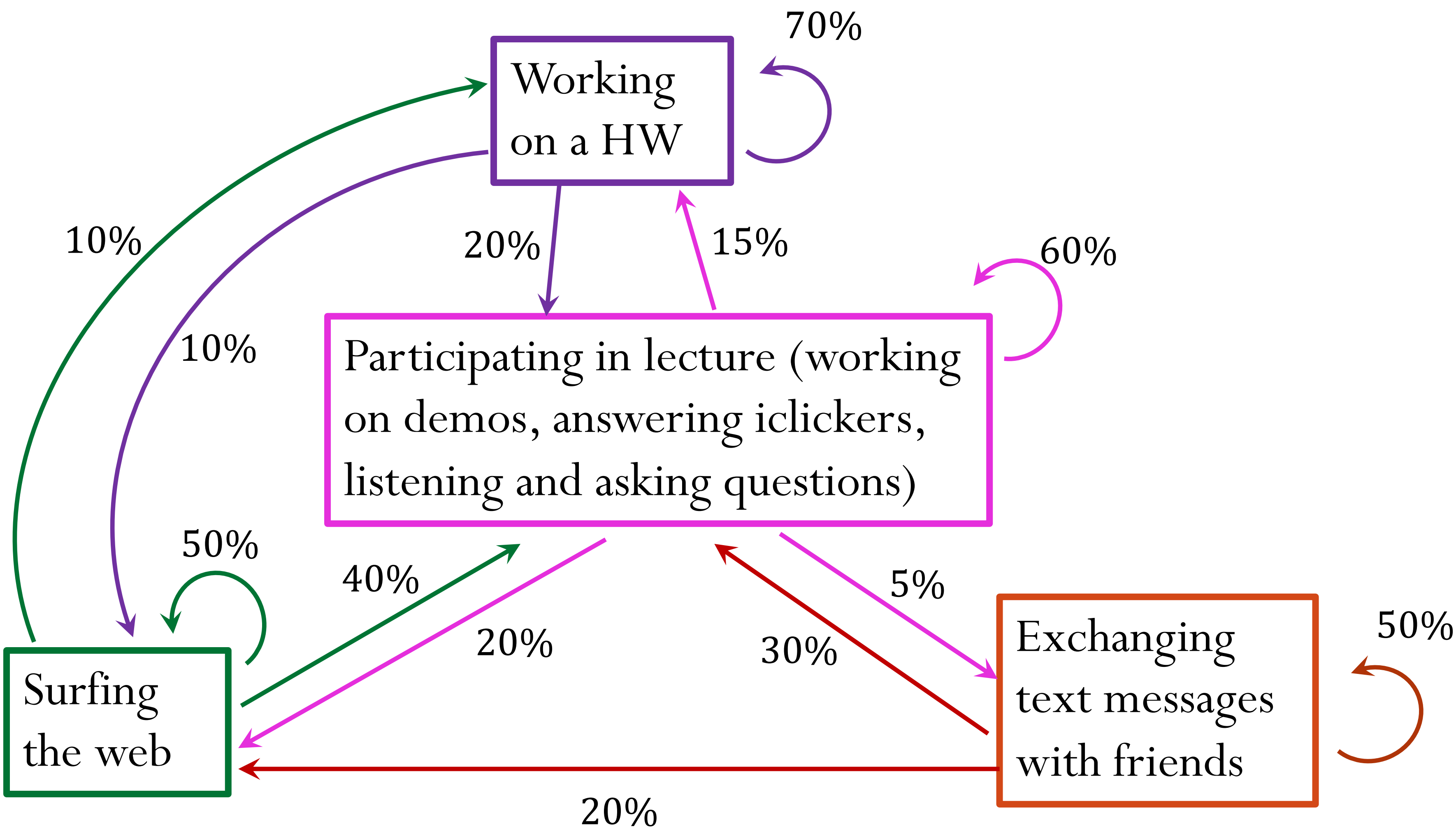
We will use the induced matrix norm definition:

$$\|\boldsymbol{A}\| = \max_{\|\boldsymbol{x}\|\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$$

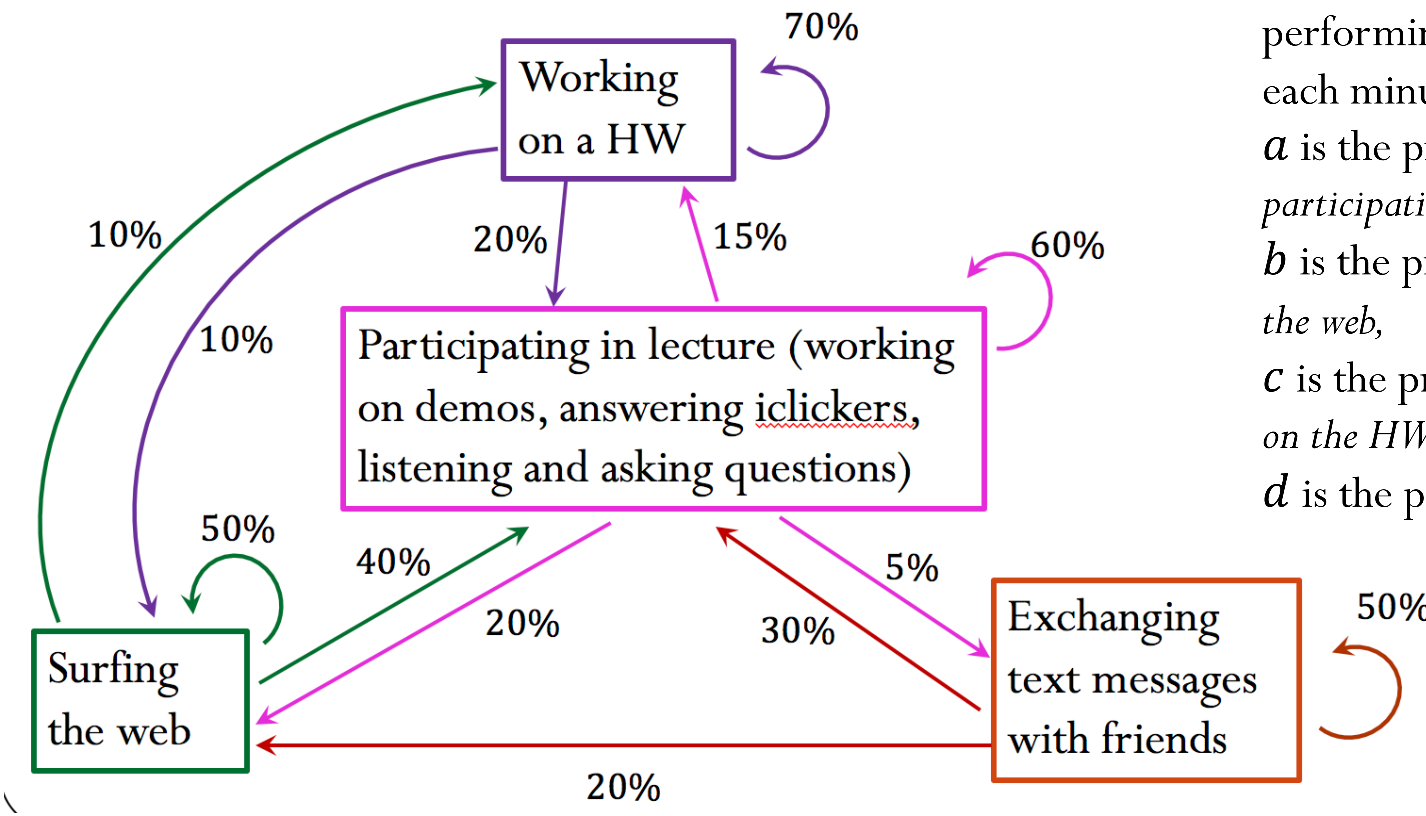
to write $|\lambda| \leq \|\boldsymbol{A}\|$. Since $\|\boldsymbol{A}\|_1 = 1$, the we have $|\lambda| \leq 1$

Another example…

Consider the following graph of states. Suppose this is a model of the behavior of a student at each minute of a lecture. ☺



Student in-class activity



$\boldsymbol{x} = (a, b, c, d)$ contains the probabilities of a student performing each activity at each minute of the class:

a is the probability of *participating in lect*ure,

b is the probability of *surfing the web*,

c is the probability of *working on the HW*,

d is the probability of *texting*.

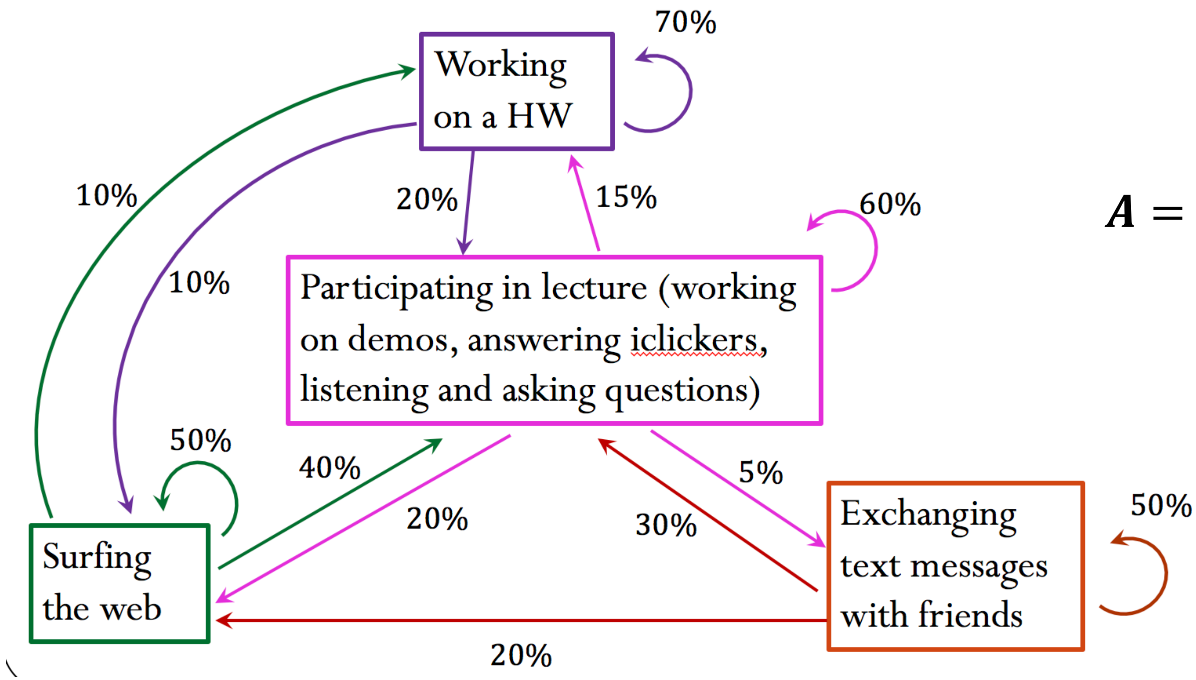
1) If the initial state is $\boldsymbol{x}_0 = (0.8, 0.1, 0.0, 0.1)$, what is the probability that the student will be *working on the HW* after one minute of the class (time step $k = 1$) ?
2) What is the probability that the student will be *surfing the web* at after 5 minutes?
3) What is the steady-state vector for this problem?
4) Would your answer change if you were to start with a different initial guess?

Student in-class activity

1) After 5 minutes, which of the following activities will have higher probability (if initial state is given by is $\boldsymbol{x}_0 = (0.8,0.1,0.0,0.1)$)?
   A. Surfing the web
   B. Working on HW
   C. Texting

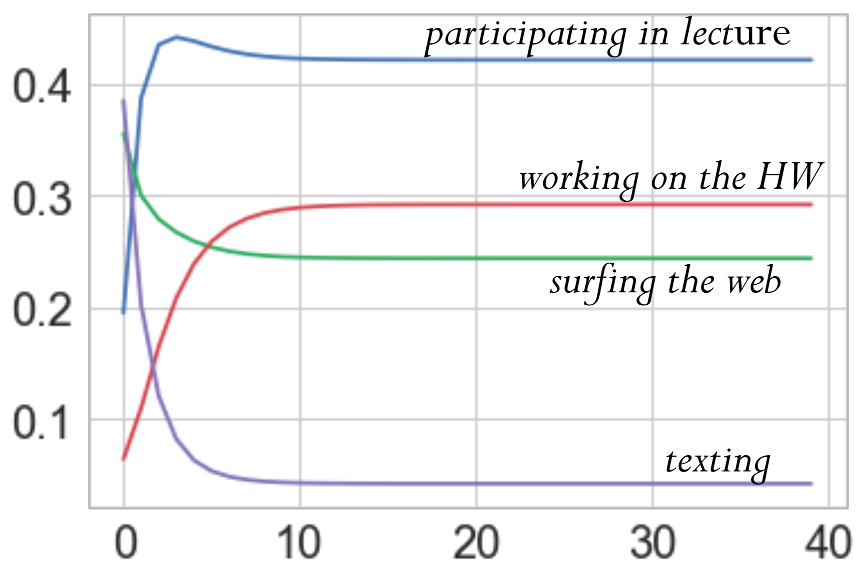
2) Could your answer above change if starting from a different initial state?
   A. YES
   B. NO

Demo "Student-Activities-During-Lecture"

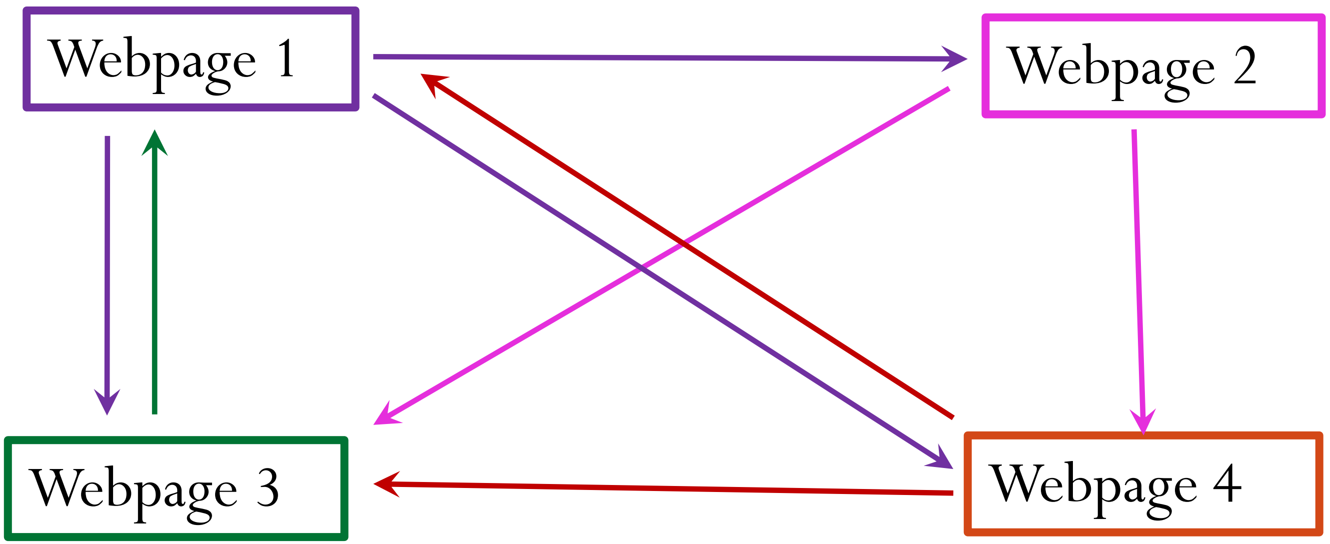


$A =$

	Lect	Web	HW	Text
Lec	0.6	0.4	0.2	0.3
Web	0.2	0.5	0.1	0.2
HW	0.15	0.1	0.7	0.0
Text	0.05	0.0	0.0	0.5



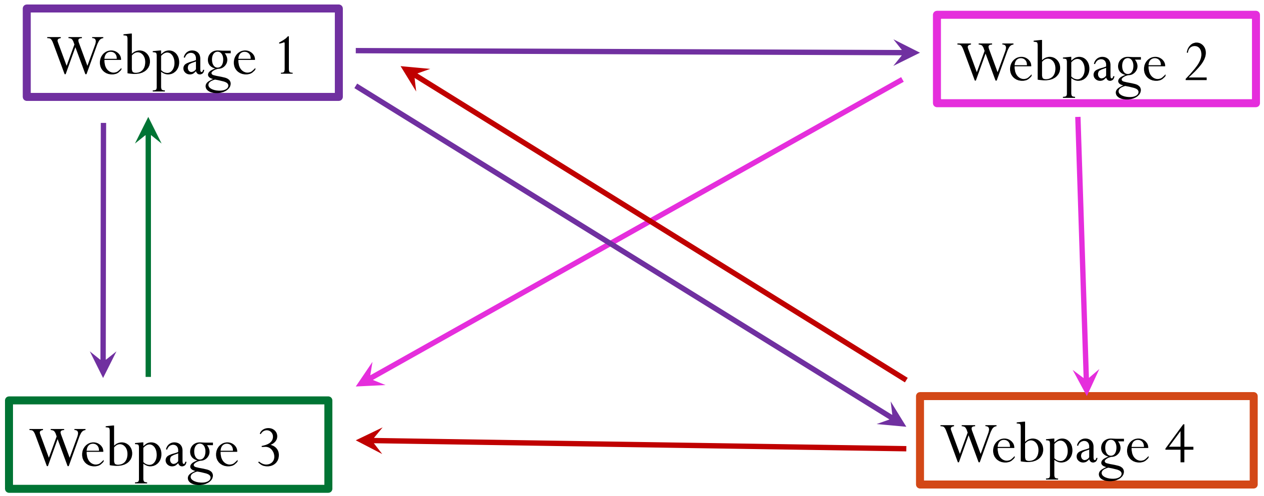
Page Rank



Problem: Consider n linked webpages (above we have $n = 4$). Rank them.

- A link to a page increases the perceived ***importance*** of a webpage
- We can represent the ***importance*** of each webpage k with the scalar x_k

Page Rank



A possible way to rank webpages…

- x_k is the number of links to page k (incoming links)
- $x_1 = 2, x_2 = 1, x_3 = 3, x_4 = 2$
- Issue: when looking at the links to webpage 1, the link from webpage 3 will have the same weight as the link from webpage 4. Therefore, links from important pages like "The NY Times" will have the same weight as other less important pages, such as "News-Gazette".

Page Rank

Another way… Let's think of Page Rank as an stochastic process.

http://infolab.stanford.edu/~backrub/google.html

"PageRank can be thought of as a model of user behavior. We assume there is a random surfer who is given a web page at random and keeps clicking on links, never hitting "back"…"

So the importance of a web page can be determined by the probability of a random user to end up on that page.

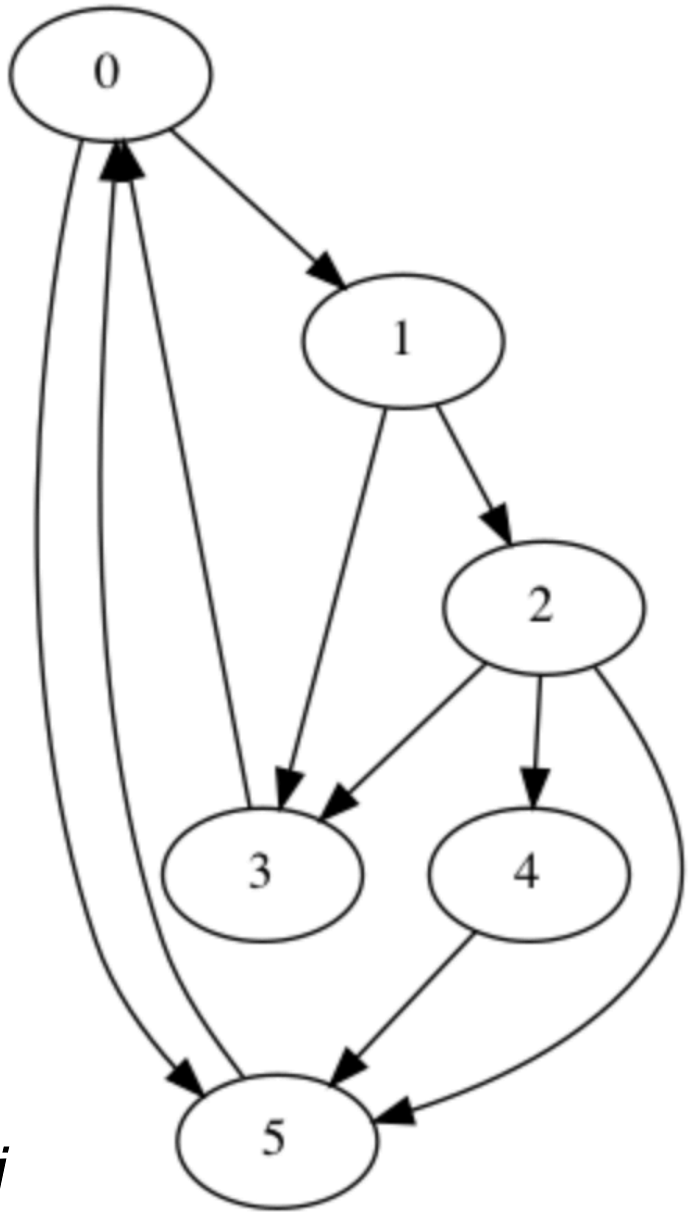
Page Rank

Let us write this graph problem (representing webpage links) as a matrix (adjacency matrix).

	0	1	2	3	4	5
0	0			1		1
1	1	0				
2		1	0			
3		1	1	0		
4			1		0	
5	1		1		1	0

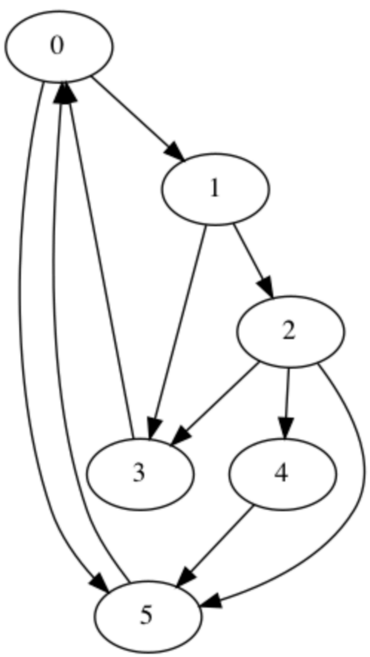
2	2	3	1	1	1

Number of outgoing links for each webpage *j*



Page Rank

- The influence of each page is split **evenly** between the pages it links to (i.e., equal weights for each outgoing link)
- Therefore, we should divide each row entry by the total column sum



	0	1	2	3	4	5
0	0			1		1
1	1	0				
2		1	0			
3		1	1	0		
4			1		0	
5	1		1		1	0

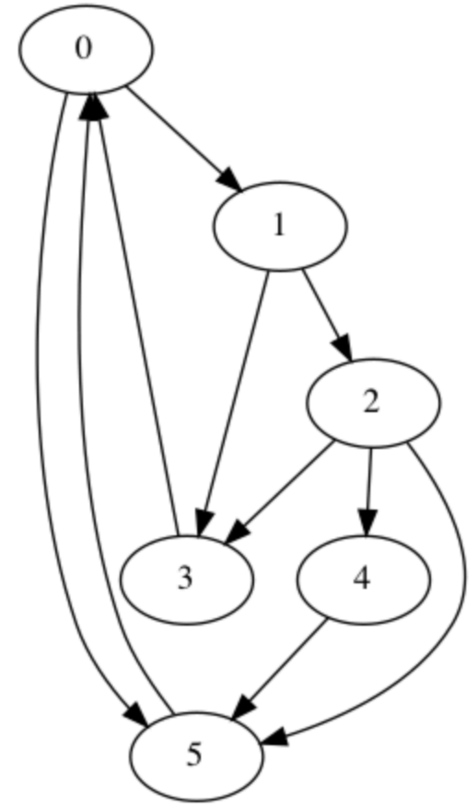
	0	1	2	3	4	5
0	0			1.0		1.0
1	0.5	0				
2		0.5	0			
3		0.5	0.33	0		
4			0.33		0	
5	0.5		0.33		1.0	0

Page Rank

Note that the sum of each column is equal to 1. This is the Markov matrix!

$\boldsymbol{A} =$

0			1.0		1.0
0.5	0				
	0.5	0			
	0.5	0.33	0		
		0.33		0	
0.5		0.33		1.0	0



We want to know the probability of a user to end up in each one of the above 6 webpages, when starting at random from one of them.

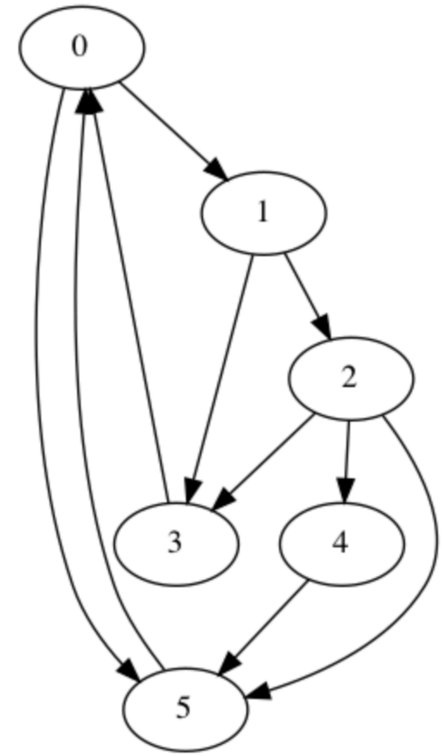
Suppose that we start with the following probability at time step 0:

$$\boldsymbol{x}_0 = (0.1, 0.2, 0.1, 0.3, 0.1, 0.2)$$

What is the probability that the user will be at "webpage 3" at time step 1?

Page Rank

$$A = \begin{pmatrix} 0 & 0 & 0 & 1.0 & 0 & 1.0 \\ 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.33 & 0 & 0 & 0 \\ 0 & 0 & 0.33 & 0 & 0 & 0 \\ 0.5 & 0 & 0.33 & 0 & 1.0 & 0 \end{pmatrix} \quad \boldsymbol{x}_0 = \begin{pmatrix} 0.1 \\ 0.2 \\ 0.1 \\ 0.3 \\ 0.1 \\ 0.2 \end{pmatrix}$$



$$\boldsymbol{x}_1 = \boldsymbol{A}\,\boldsymbol{x}_0 = \begin{pmatrix} 0.5 \\ 0.05 \\ 0.1 \\ 0.133 \\ 0.033 \\ 0.184 \end{pmatrix}$$

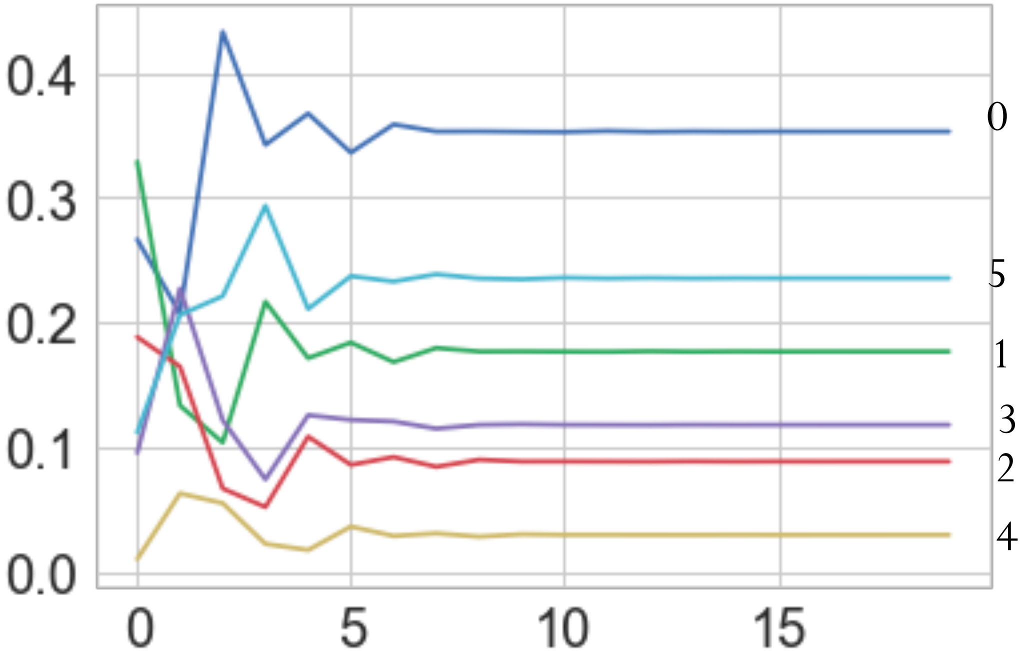
The user will have a probability of about 13% to be at "webpage 3" at time step 1.

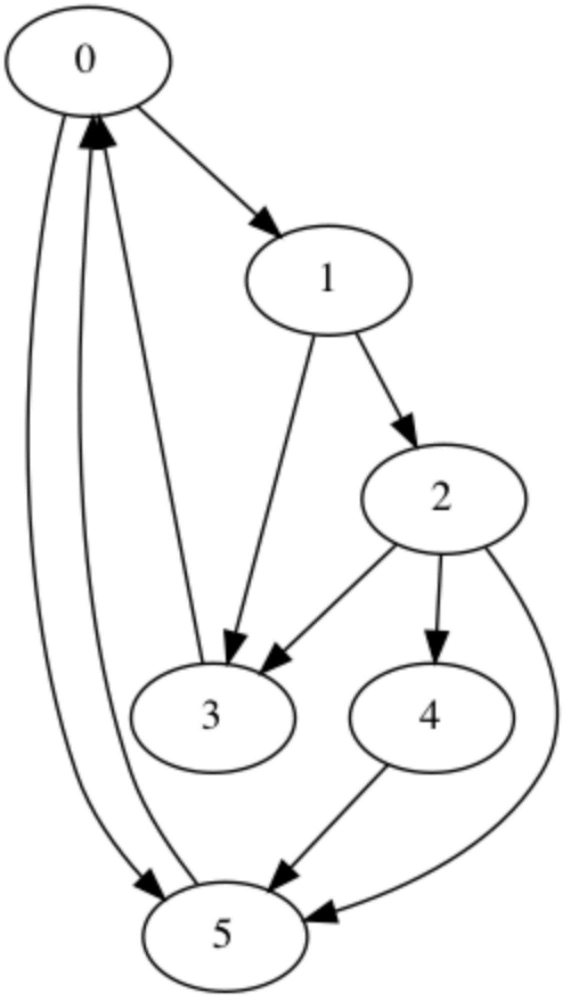
At steady-state, what is the most likely page the user will end up at, when starting from a random page?

Perform $\boldsymbol{x}_n = \boldsymbol{A}\,\boldsymbol{x}_{n-1}$ until convergence!

Page Rank

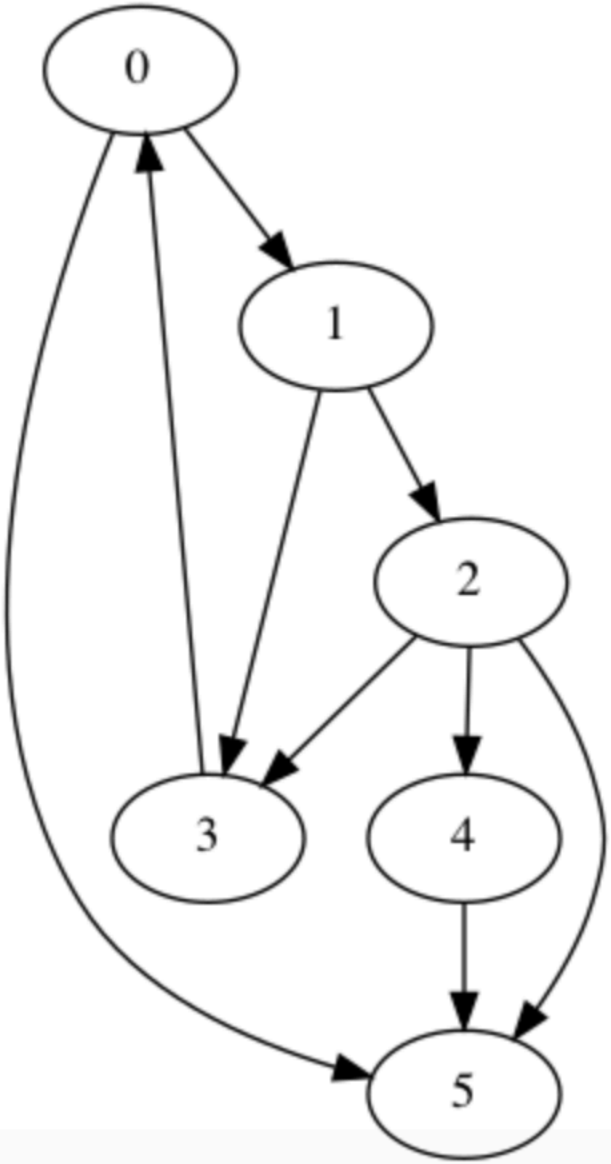
The plot below shows the probabilities of a user ending up at each webpage for each time step.





The most "important" page is the one with the highest probability. Hence, the ranking for these 6 webpages would be (starting from the most important): Webpages 0,5,1,3,2,4

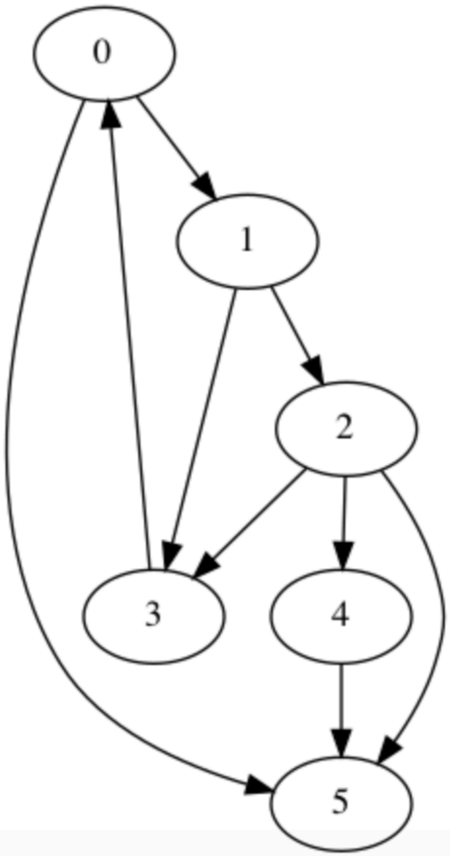
What if we now remove the link from webpage 5 to webpage 0?



	0	1	2	3	4	5
0	0			1		
1	1	0				
2		1	0			
3		1	1	0		
4			1		0	
5	1		1		1	0

Note that we can no longer divide the entries of the last column by the total column sum, which in this case is zero (no outgoing links).

Approach: Since a random user will not stay on the same webpage forever, we can assume that all the other webpages have the same probability to be linked from "webpage 5".



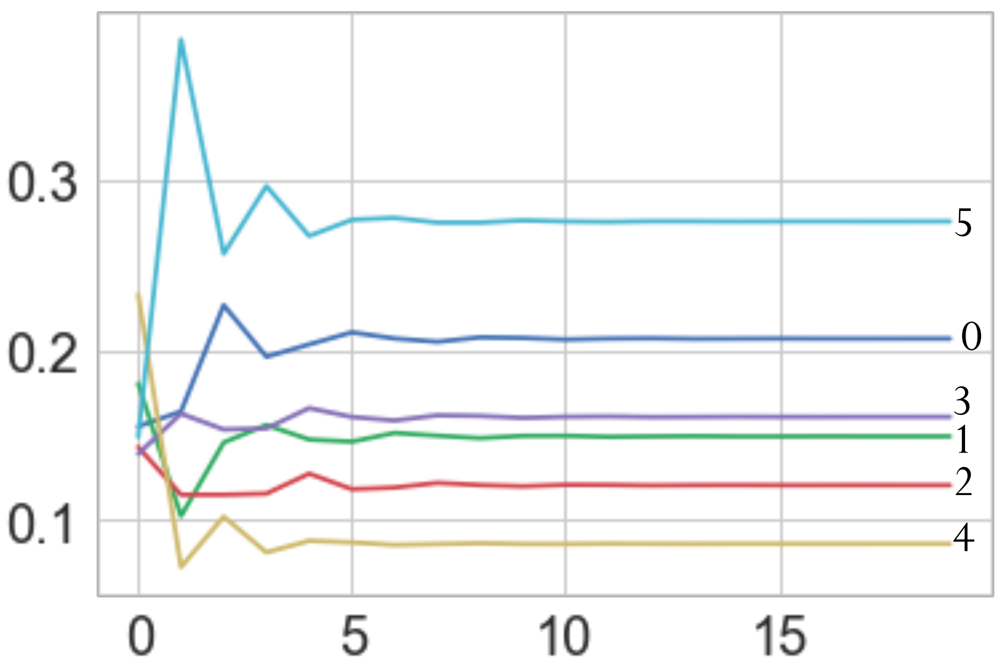
	0	1	2	3	4	5
0	0			1		
1	1	0				
2		1	0			
3		1	1	0		
4			1		0	
5	1		1		1	0

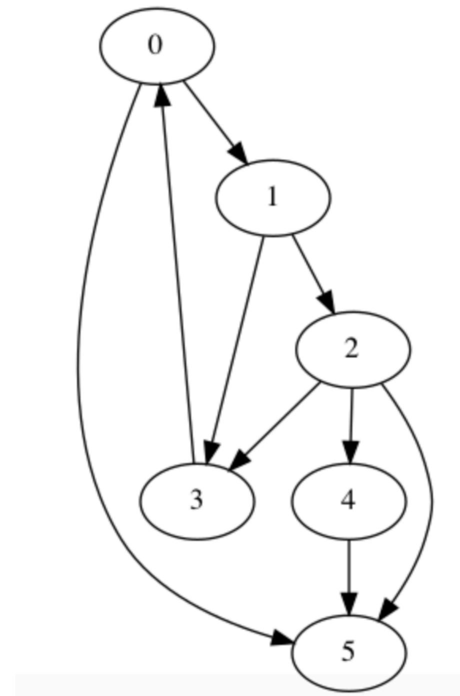
	0	1	2	3	4	5
0	0			1.0		0.166
1	0.5	0				0.166
2		0.5	0			0.166
3		0.5	0.33	0		0.166
4			0.33		0	0.166
5	0.5		0.33		1.0	0.166

Page Rank

$$A = \begin{pmatrix} 0 & 0 & 0 & 1.0 & 0 & 0.166 \\ 0.5 & 0 & 0 & 0 & 0 & 0.166 \\ 0 & 0.5 & 0 & 0 & 0 & 0.166 \\ 0 & 0.5 & 0.33 & 0 & 0 & 0.166 \\ 0 & 0 & 0.33 & 0 & 0 & 0.166 \\ 0.5 & 0 & 0.33 & 0 & 1.0 & 0.166 \end{pmatrix}$$

The plot below shows the probabilities of a user ending up at each webpage for each time step.

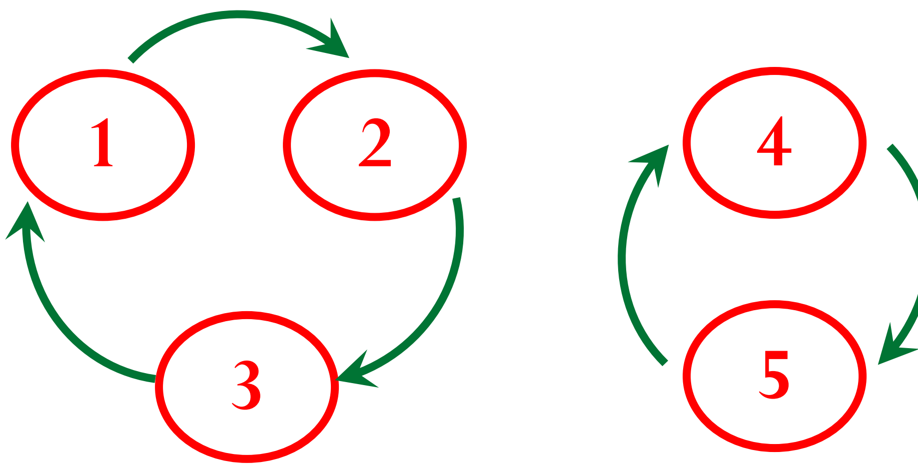




The most "important" page is the one with the highest probability. Hence, the ranking for these 6 webpages would be (starting from the most important): Webpages 5,0,3,1,2,4

Page Rank

One remaining issue: the Markov matrix does not guarantee a unique solution



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

Matrix A has two eigenvectors corresponding to the same eigenvalue 1

$$\boldsymbol{x}^* = \begin{pmatrix} 0.33 \\ 0.33 \\ 0.33 \\ 0 \\ 0 \end{pmatrix} \qquad \boldsymbol{x}^* = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

Perron-Frobenius theorem (CIRCA 1910): If $\boldsymbol{A}$ is a Markov matrix with all positive entries, then M has unique steady-state vector $\boldsymbol{x}^*$.

Page Rank

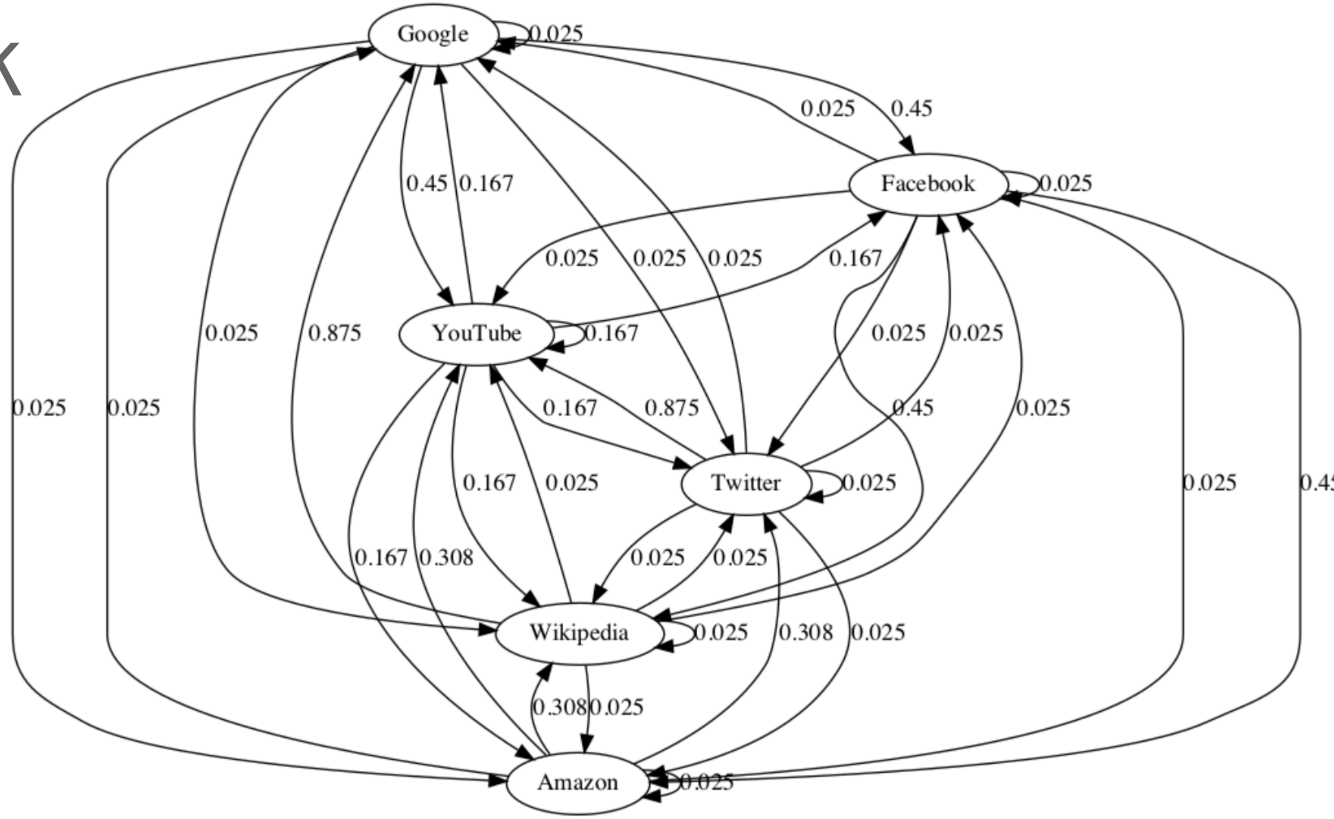
Brin-Page (1990s) proposed: "PageRank can be thought of as a model of user behavior. We assume there is a random surfer who is given a web page at random and keeps clicking on links, never hitting "back", but eventually gets bored and starts on another random page."

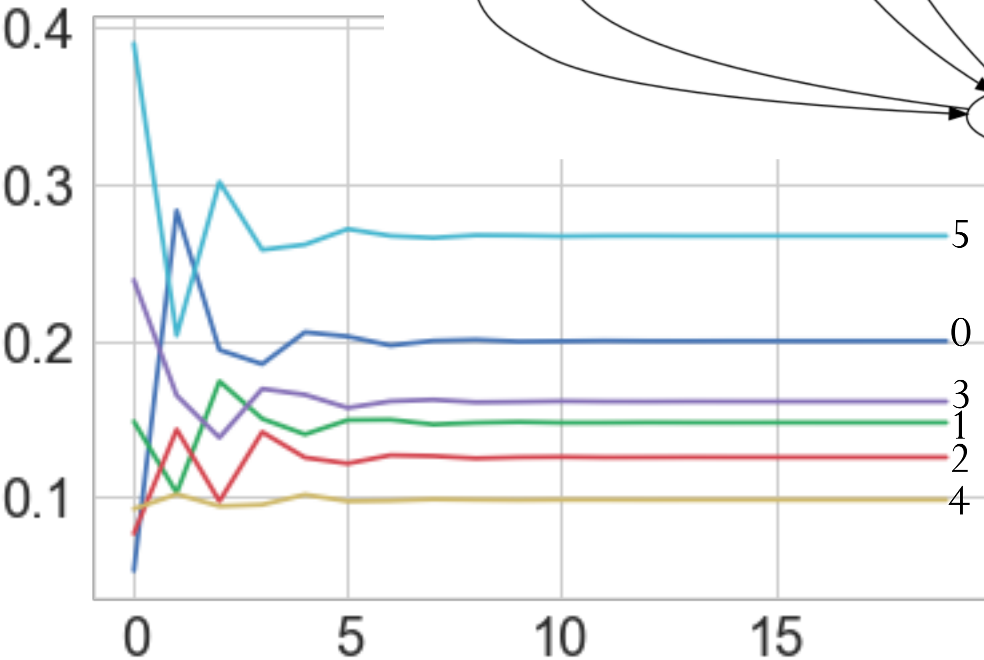
$$\boldsymbol{M} = 0.85\,\boldsymbol{A} + \frac{0.15}{n}\boldsymbol{1}$$

So a surfer clicks on a link on the current page with probability 0.85 and opens a random page with probability 0.15.

This model makes all entries of **M** greater than zero, and guarantees a unique solution.

Page Rank





$$\boldsymbol{M} = 0.85\,\boldsymbol{A} + \frac{0.15}{n}\mathbb{1}$$

Iclicker question

For the Page Rank problem, we have to compute

$$\boldsymbol{M} = 0.85\,\boldsymbol{A} + \frac{0.15}{n}\mathbb{1}$$

And then perform a matrix-vector multiplications $\boldsymbol{x}_n = \boldsymbol{M}\,\boldsymbol{x}_{n-1}$

What is the cost of the matrix-vector multiplication $\boldsymbol{b} = \mathbb{1}\,\boldsymbol{x}_{n-1}$?

A) $O(1)$
B) $O(n)$
C) $O(n^2)$
D) $O(n^3)$