Introduction to Information Retrieval CS 221
Donald J. Patterson

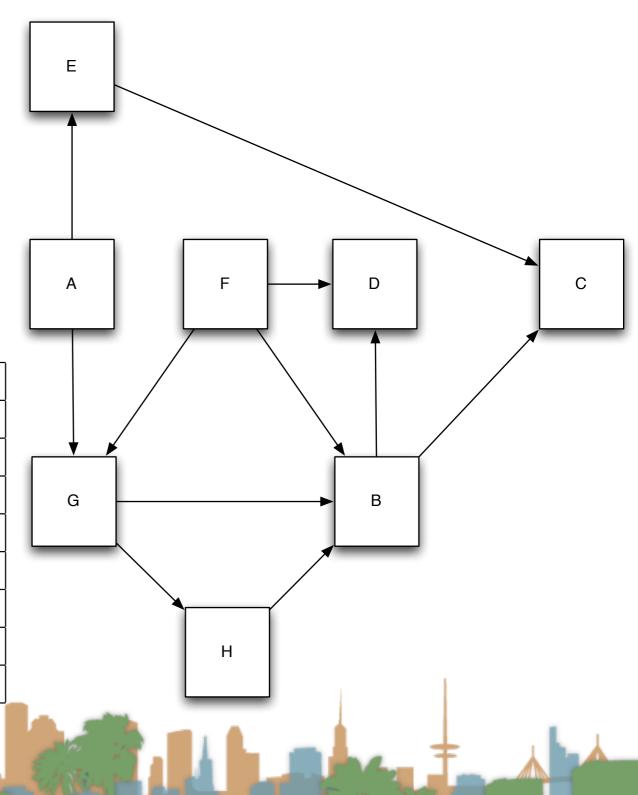
Content adapted from Hinrich Schütze http://www.informationretrieval.org



Markov Chains

- Example:
 - 8 states
 - (web pages or whatever)
 - 8 by 8 transition prob. matrix

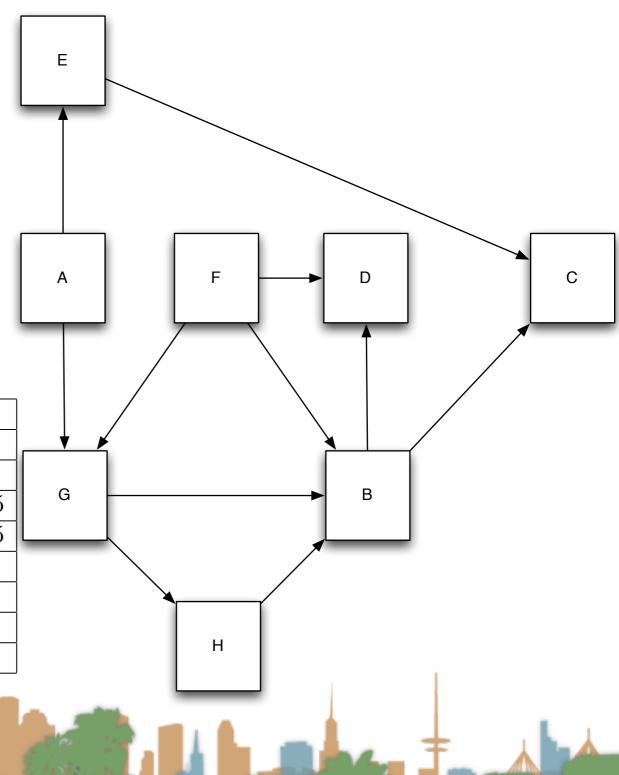
	A	B	C	D	$oxed{E}$	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0	0	0	0	0	0	0	0
\overline{D}	0	0	0	0	0	0	0	0
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0



Markov Chains

- Example:
 - 8 states
 - 8 by 8 transition prob. matrix
 - Handle Dead-Ends also

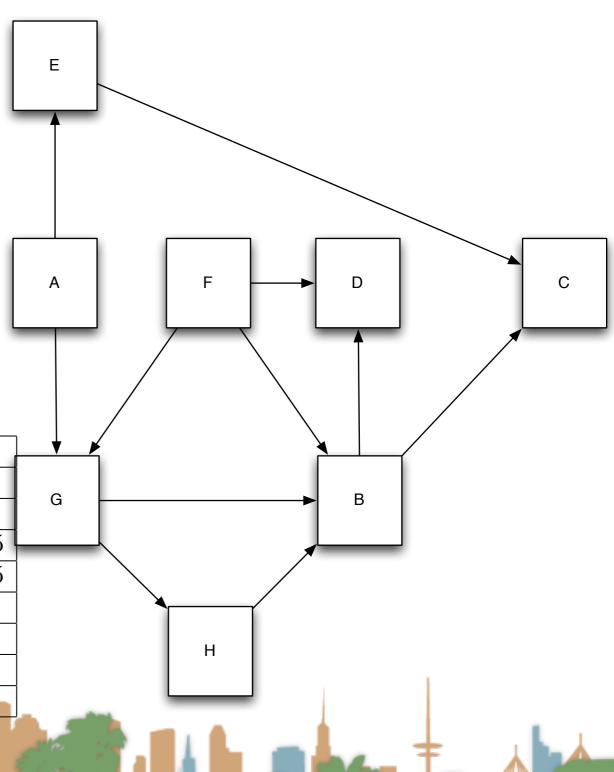
	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0



Markov Chains

- Example:
 - 8 states
 - 8 by 8 transition prob. matrix
 - Handle Dead-Ends also
 - Handle teleports

	$A \mid$	B	$\mid C \mid$	D	E	F	G	H
A	0.01	0.01	0.01	0.01	0.47	0.01	0.47	0.01
B	0.01	0.01	0.47	0.47	0.01	0.01	0.01	0.01
C	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
E	0.01	0.01	0.93	0.01	0.01	0.01	0.01	0.01
\overline{F}	0.01	0.32	0.01	0.32	0.01	0.01	0.32	0.01
G	0.01	0.47	0.01	0.01	0.01	0.01	0.01	0.47
H	0.01	0.93	0.01	0.01	0.01	0.01	0.01	0.01

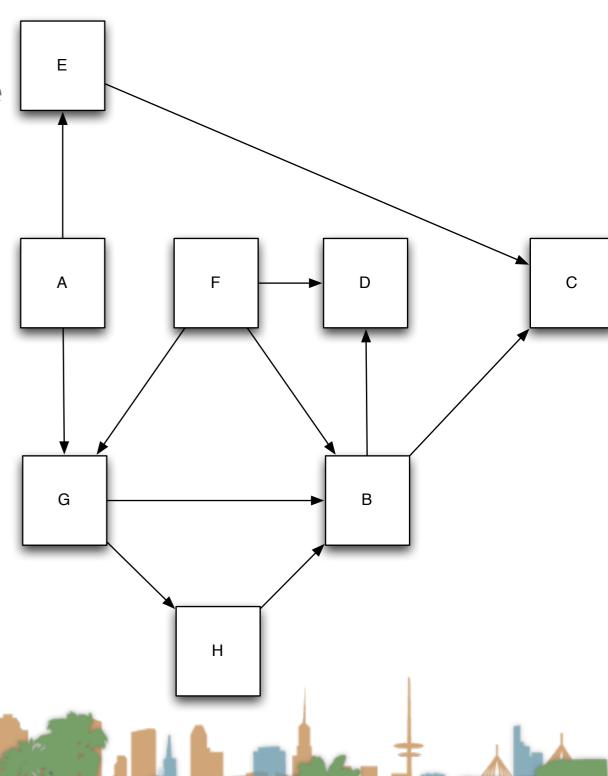


Markov Chain: The Game

You may be in one state at a time

 Every tick you move one step chosen randomly from the transition probability matrix

	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0

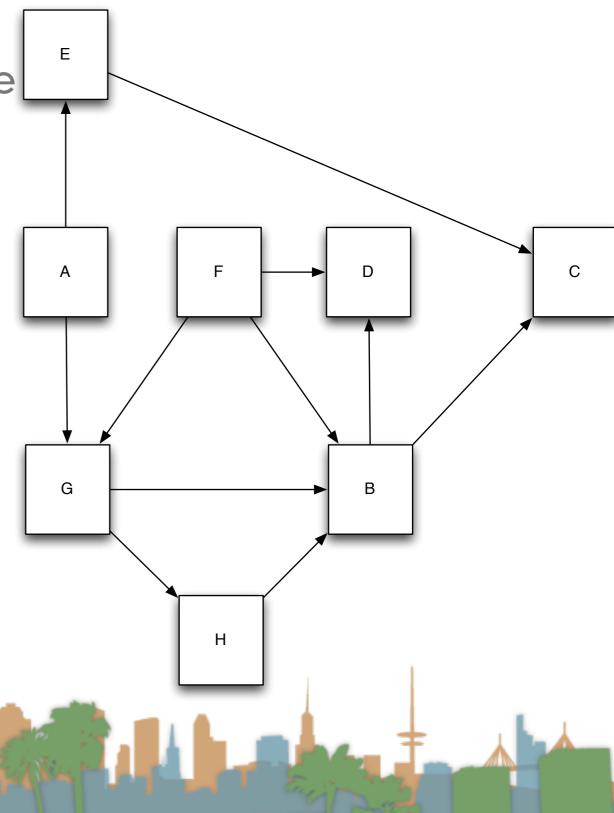


The Markov Property

 It doesn't matter where you came from.

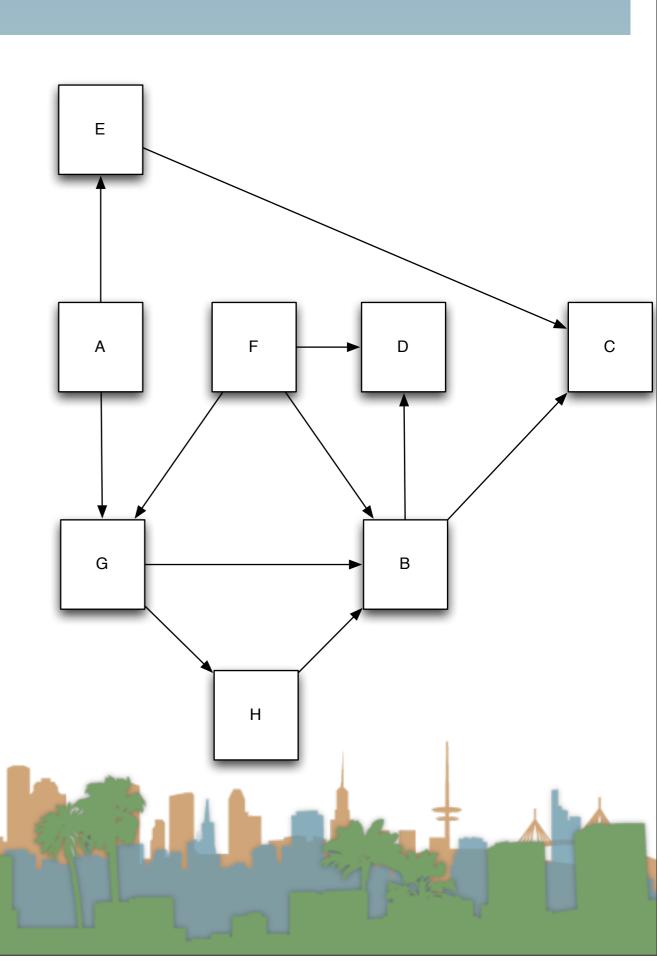
 All information that you need to take the next step comes from your current state and the transition probability matrix

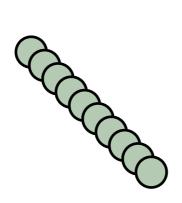
 History is irrelevant given your current state



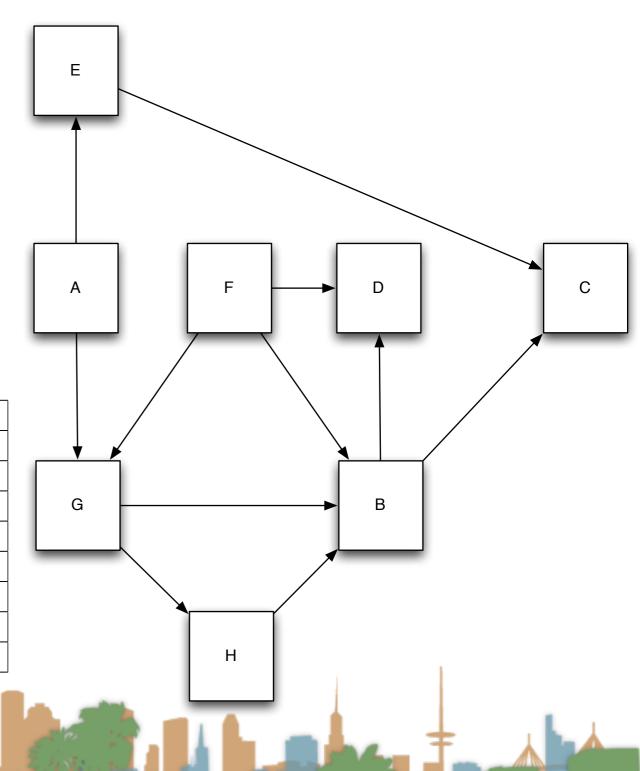
PageRank

- PageRank is the long term visit rate of a random walk on the graph.
- With teleports

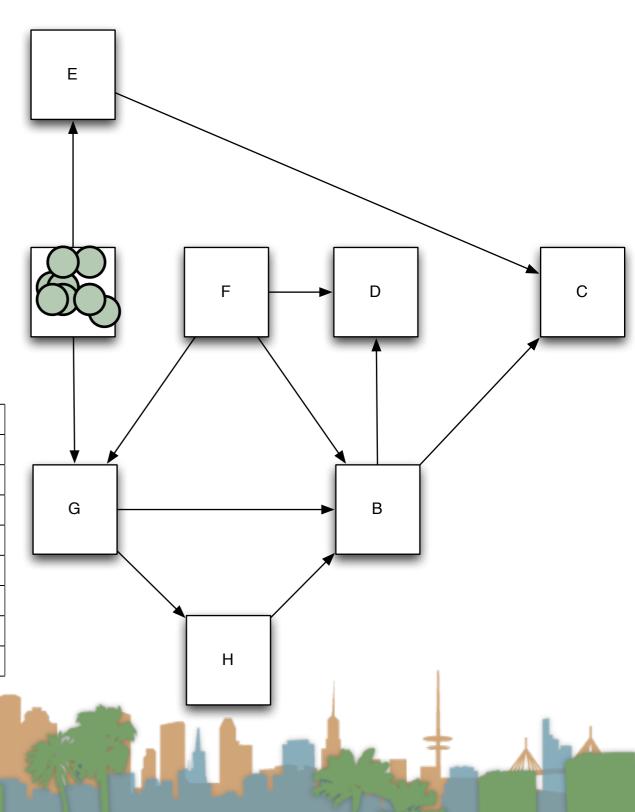




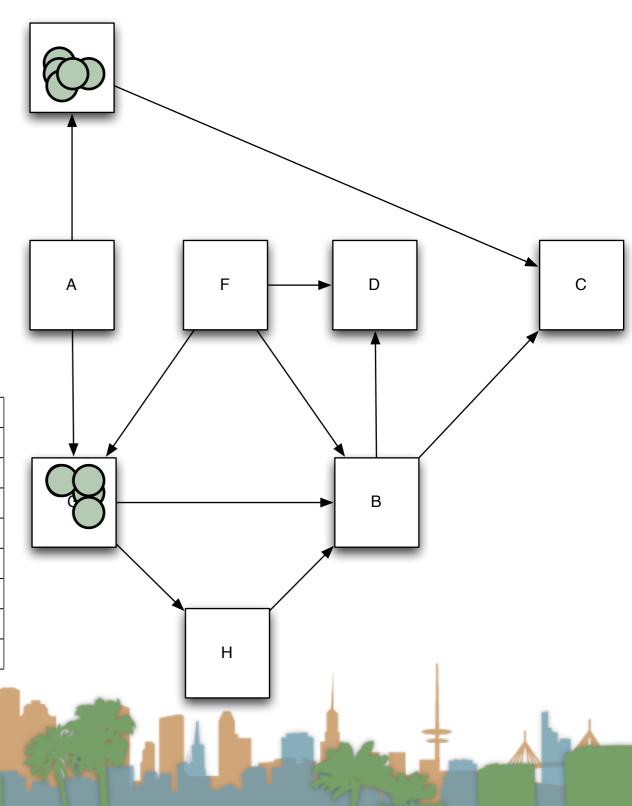
	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0



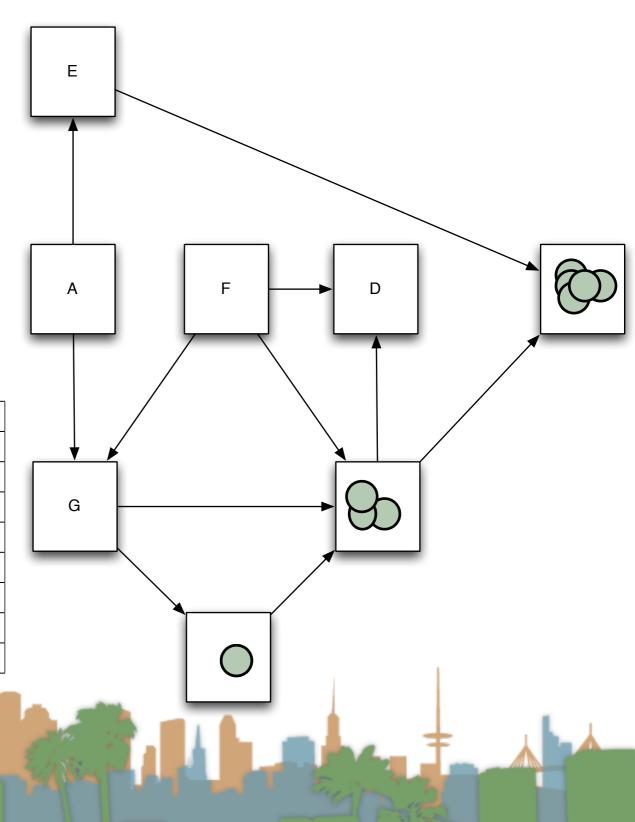
	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0



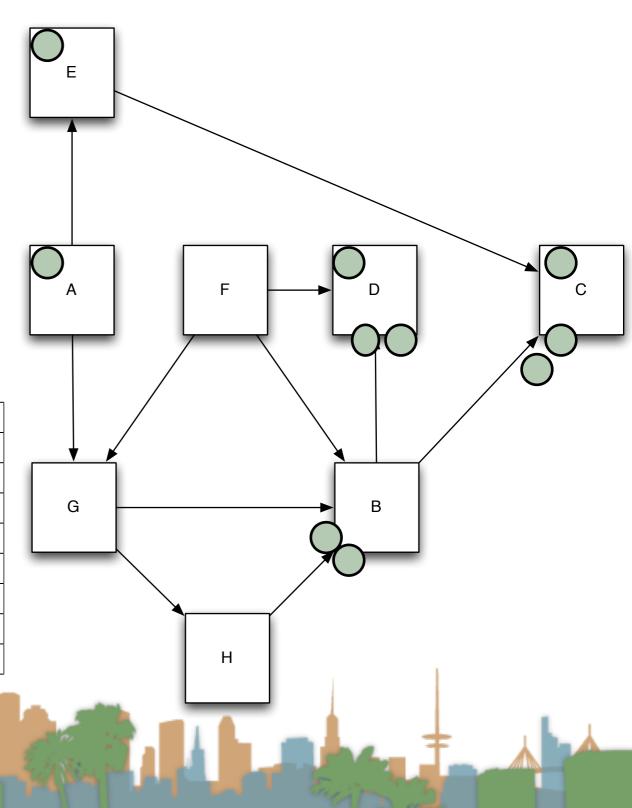
	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0



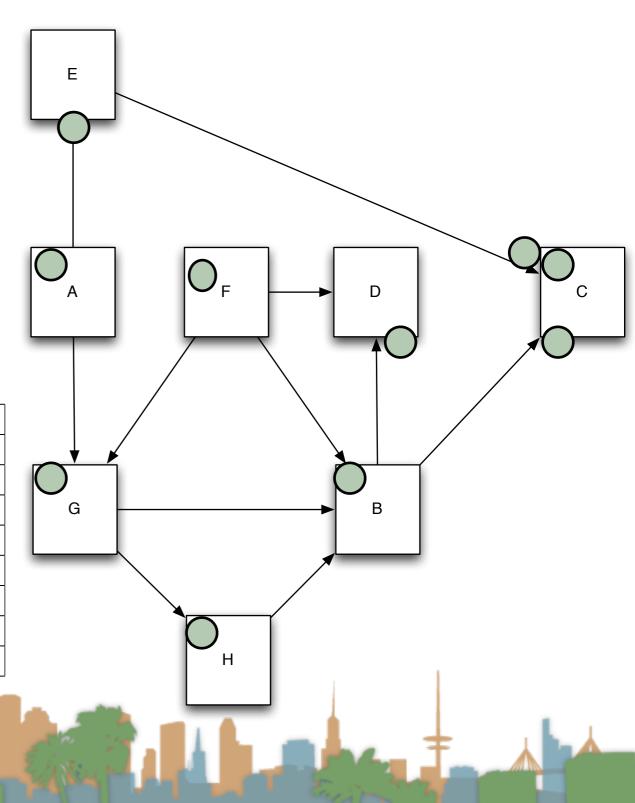
	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0



	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0



	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0



Long-Term visit rate

• A: 5%

• B: 21%

• C: 23%

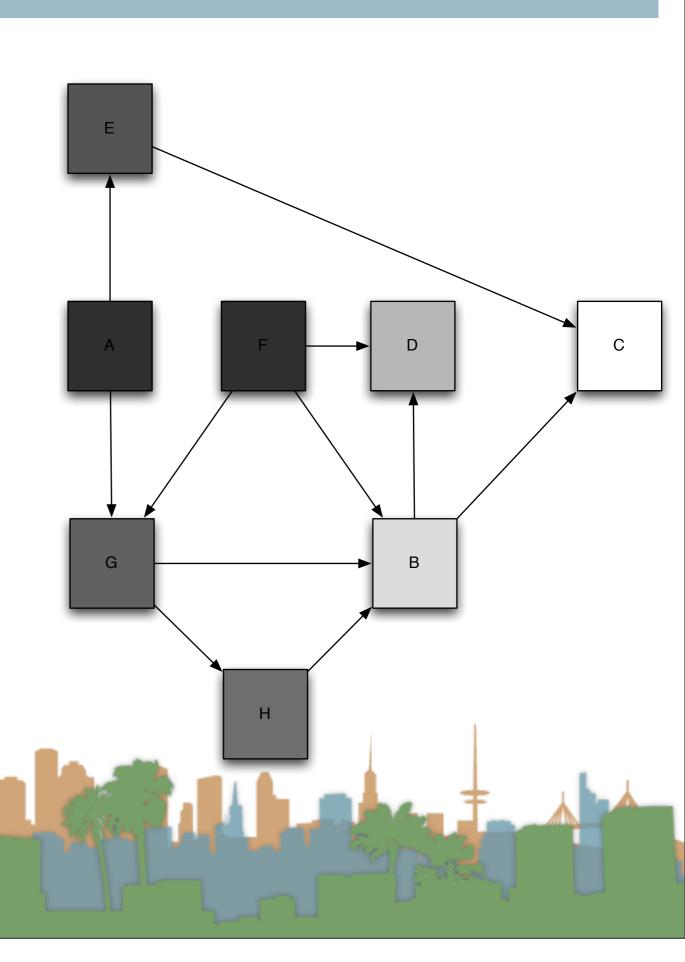
• D: 18%

• E: 8%

• F: 5%

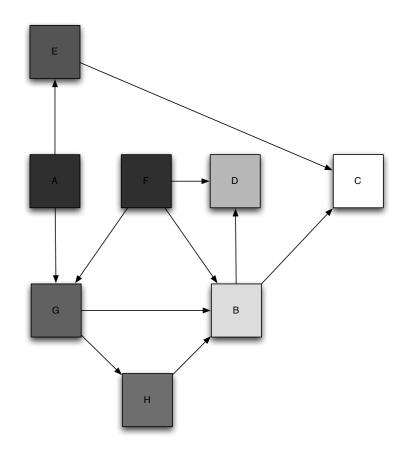
• G: 9%

• H: 10%



Some properties of Markov chains

- Ergodic:
 - All states can reach all states
 - What did we have to do to enable this for a web graph?
- Steady State Theorem:
 - Every ergodic markov chain has a steady state -> has a PageRank



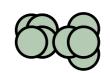


Calculating PageRank

Visual representation to math representation

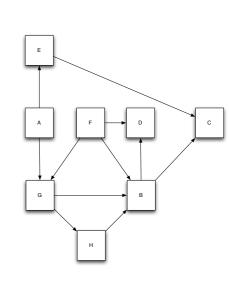
 $\vec{x_0}$

\overline{A}	B	C	D	$\mid E \mid$	$\mid F \mid$	G	H
1.0	0	0	0	0	0	0	0



P

	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0



Calculating PageRank

$$\vec{x_0}P = \vec{x_1}$$

A	B	C	D	$\mid E \mid$	F	G	H
1.0	0	0	0	0	0	0	0

	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0



Calculating PageRank

• Take one step is multiplying state vector

$$\vec{x_0}P = \vec{x_1}$$

times	transition	probability	matrix
-------	------------	-------------	--------

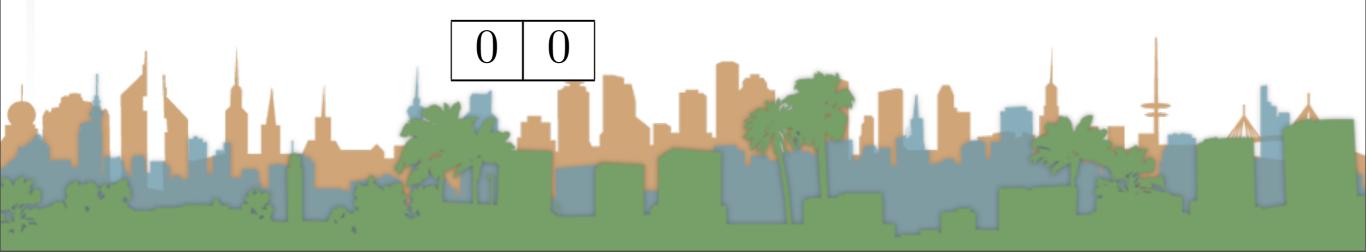
	B		A	B	C	D	E	F	G	H
0	\bigcirc	\overline{A}	0	0	0	0	0.5	0	0.5	0
		\overline{B}	0	0	0.5	0.5	0	0	0	0
0		\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
$\ \circ$	E	\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
-		\overline{E}	0	0	1.0	0	0	0	0	0
	F	\overline{F}	0	0.33	0	0.33	0	0	0.33	0
$\ $		\overline{G}	0	0.5	0	0	0	0	0	0.5
	1:	H	0	1.0	0	0	0	0	0	0

Calculating PageRank

Take one step is multiplying state vector

$$\vec{x_0}P = \vec{x_1}$$

	A	E	$\overline{\Lambda}$	C	D	E	F	G	H
\overline{A}	0	0		0	0	0.5	0	0.5	0
\overline{B}	0	$ \circ 0$		0.5	0.5	0	0	0	0
\overline{C}	0.125	0.1	25	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.1	25	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0		F	1.0	0	0	0	0	0
\overline{F}	0	0.	33	0	0.33	0	0	0.33	0
\overline{G}	0	<u>0.</u>	5	0	0	0	0	0	0.5
\overline{H}	0	Ъ	DEI	0	0	0	0	0	0

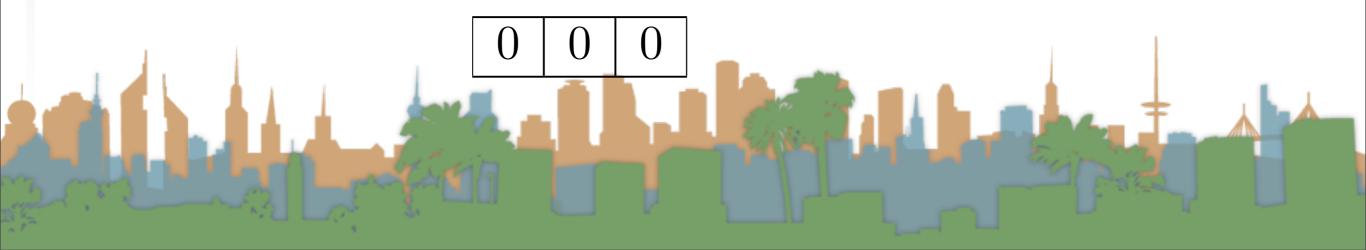


Calculating PageRank

Take one step is multiplying state vector

$$\vec{x_0}P = \vec{x_1}$$

	A	B		D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	165 C	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	100	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	(C)	1 0	0	0	0	0

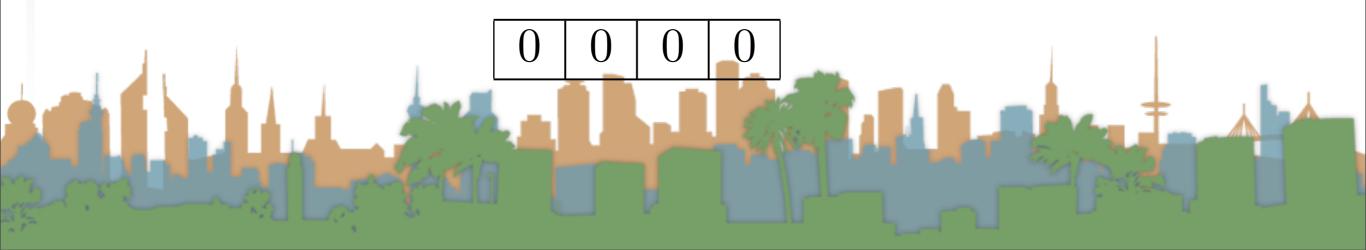


Calculating PageRank

Take one step is multiplying state vector

$$\vec{x_0}P = \vec{x_1}$$

	A	B	C	P	$\mid E \mid$	F	G	H
\overline{A}	0	0	0		0.5	0	0.5	0
\overline{B}	0	0	0.5	\parallel \mathfrak{B}	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	1 B	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
H	0	1.0	0	B 5	0	0	0	0



Calculating PageRank

• Take one step is multiplying state vector

$\vec{x_0}P$	$=\vec{x_1}$
--------------	--------------

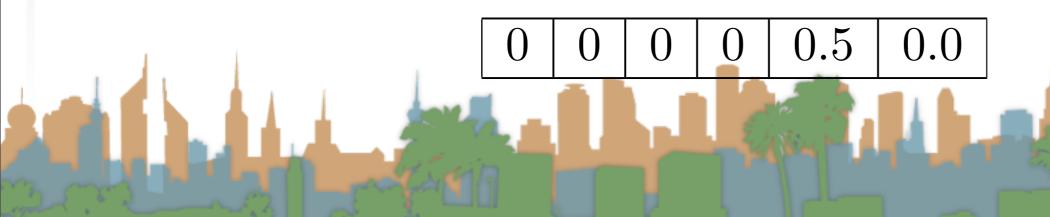
						-			
	A	B	C	D			F	G	H
\overline{A}	0	0	0	0	0	5	0	0.5	0
\overline{B}	0	0	0.5	0.5			0	0	0
\overline{C}	0.125	0.125	0.125	0.125	111	125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.	125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0		$\theta \mid \mathcal{F}$	0	0	0
\overline{F}	0	0.33	0	0.33		0	0	0.33	0
\overline{G}	0	0.5	0	0		0 4	0	0	0.5
\overline{H}	0	1.0	0	0		(A)	1 0	0	0

Calculating PageRank

• Take one step is multiplying state vector

	$\vec{x_0}P$	=	$\vec{x_1}$

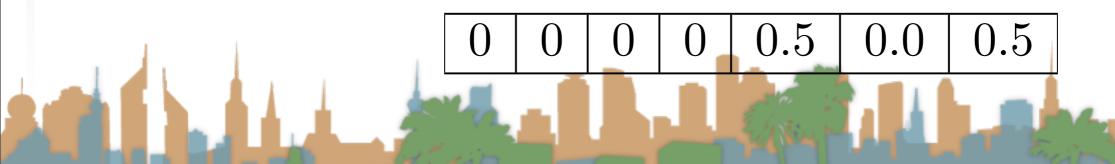
	A	B	C	D	E	$\ \cdot \ $		G	H
\overline{A}	0	0	0	0	0.5	$\left\ \cdot \right\ $		0.5	0
\overline{B}	0	0	0.5	0.5	0		9	0	0
C	0.125	0.125	0.125	0.125	0.125).125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125		0.725	0.125	0.125
\overline{E}	0	0	1.0	0	0		A 1	0	0
\overline{F}	0	0.33	0	0.33	0		0	0.33	0
\overline{G}	0	0.5	0	0	0		0 7	$\begin{array}{ccc} & 0 & \\ & & \end{array}$	0.5
\overline{H}	0	1.0	0	0	0		(A)	0	0



Calculating PageRank

X	701.0	D A	_	$\vec{x_1}$	
	0	B			

	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0		0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125		0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	(A)	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0 7	0.5
\overline{H}	0	1.0	0	0	0	0	(P) #	0



Calculating PageRank

Э	$\vec{c_0}P =$		
		0	\mathcal{B}
	G	1	
	0.5	19	$\int_{C} \int_{C} \int_{C$

	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
H	0	1.0	0	0	0	0	0	



Calculating PageRank

$$\vec{x_0}P = \vec{x_1}$$

\overline{A}	B	C	D	E	F	G	H	
1.0	0	0	0	0	0	0	0	

	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0

$$\vec{x_1} = \begin{bmatrix} 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \end{bmatrix}$$



Calculating PageRank

$$\vec{x_1}P = \vec{x_2}$$

0	0	0	0	0.5	0	0.5	0
			l .		l		

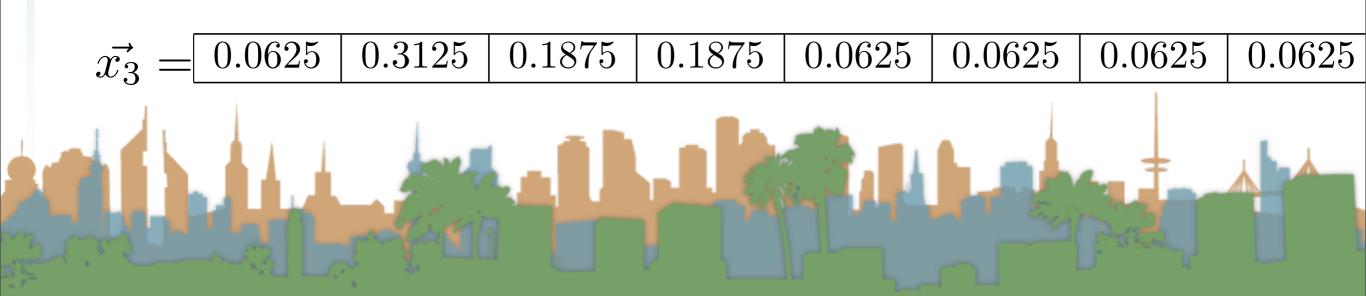
	A	B	C	D	E	F	G	H
\overline{A}	0	0	0	0	0.5	0	0.5	0
\overline{B}	0	0	0.5	0.5	0	0	0	0
\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
\overline{E}	0	0	1.0	0	0	0	0	0
\overline{F}	0	0.33	0	0.33	0	0	0.33	0
\overline{G}	0	0.5	0	0	0	0	0	0.5
\overline{H}	0	1.0	0	0	0	0	0	0



Calculating PageRank

$$\vec{x_1}P = \vec{x_2}$$

0 0.25 0.5 0 0.0 0 0.0 0.25		A	B	C	D	E	F	G	H
0 0.25 0.5 0 0.0 0 0.0 0.25	\overline{A}	0	0	0	0	0.5	0	0.5	0
	\overline{B}	0	0	0.5	0.5	0	0	0	0
	\overline{C}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
	\overline{D}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
	\overline{E}	0	0	1.0	0	0	0	0	0
	\overline{F}	0	0.33	0	0.33	0	0	0.33	0
	\overline{G}	0	0.5	0	0	0	0	0	0.5
	\overline{II}	0	1.0	0	O	n	n	0	\cap



Calculating PageRank

$$\vec{x_1}P = \vec{x_2}$$

$$\lim_{(n\to\infty)} x_n = PageRank$$



Long-Term visit rate

• A: 5%

• B: 21%

• C: 23%

• D: 18%

• E: 8%

• F: 5%

• G: 9%

• H: 10%

