April 16, 2018

CS 361: Probability & Statistics

Markov chains

Sequences

We will be studying models for sequences in this chapter

By the end we will be able to give a principled answer to the following problem, among other things

Give the best completion of this sentence
I had a glass of red wine with my grilled ________

Because we will learn to develop principled models of certain simple kinds of sequential data — sequences of words in this case

Markov Chains

Example

Consider this problem. You choose to flip a fair coin until you see two heads in a row at which point you will stop flipping.

What is the probability that you will flip the coin exactly twice and stop?

$$P(\{2 \text{ flips}\}) = P(\{HH\}) = P(H)P(H) = 1/4$$

What is the probability that you flip the coin exactly three times?

$$P(\{THH\}) = 1/8$$

How about 4 times?

$$P(\{HTHH, TTHH\}) = 2/16$$

Calculating the probability that we stop after N flips is going to be pretty laborious, so let's explore some ways of thinking about this problem

Recurrence relation

Consider this problem. You choose to flip a fair coin until you see two heads in a row at which point you will stop flipping. What is P(N) the probability that you stop after exactly N flips?

Write Σ_N for a string of length *N* that ends in HH

Then for N > 2, either the string starts with T and $\Sigma_N = T\Sigma_{N-1}$

Or it starts with HT and $\Sigma_N = HT\Sigma_{N-2}$

Base cases:

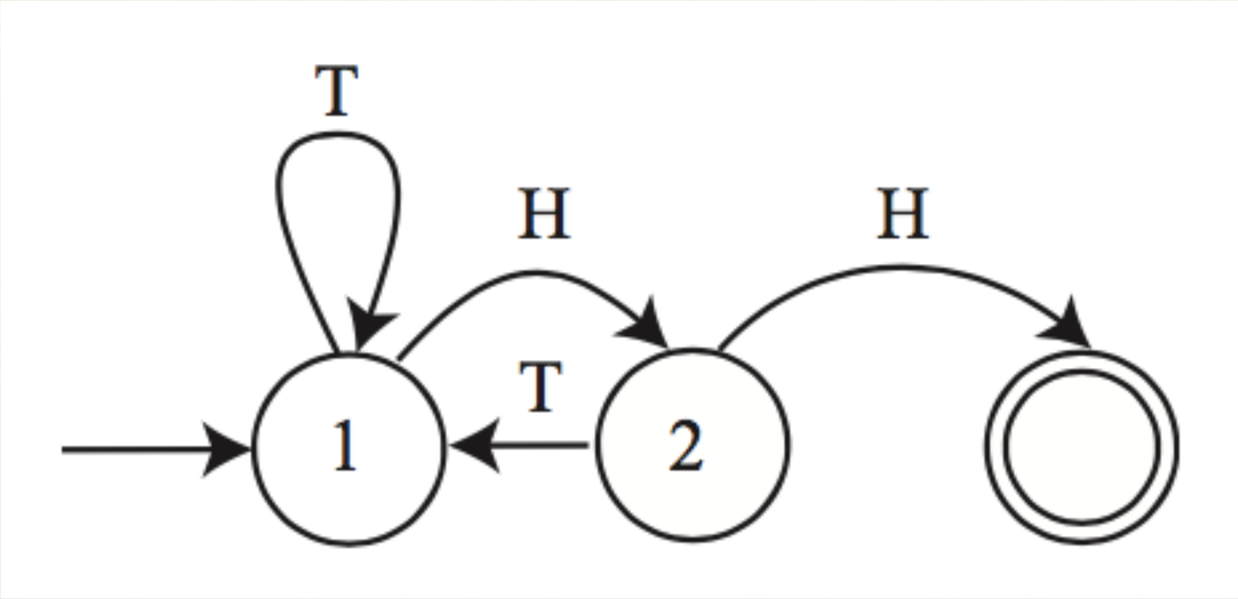
P(1) = 0   P(2) = 1/4

$$\begin{aligned} P(N) &= P(T)P(N-1) + P(HT)P(N-2) \\ &= (1/2)P(N-1) + (1/4)P(N-2) \end{aligned}$$

Finite state diagrams

Useful to model a problem like this in terms of thinking about states and transitions

Start state indicated by an arrow

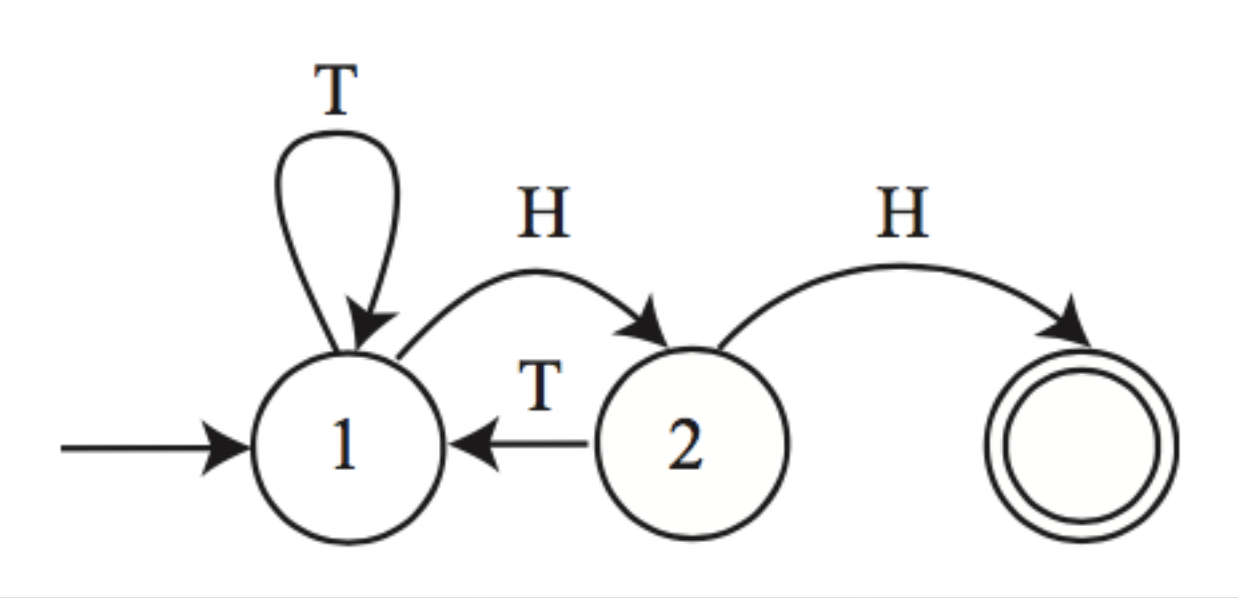


An end state indicated by an extra circle

Each node represents a state

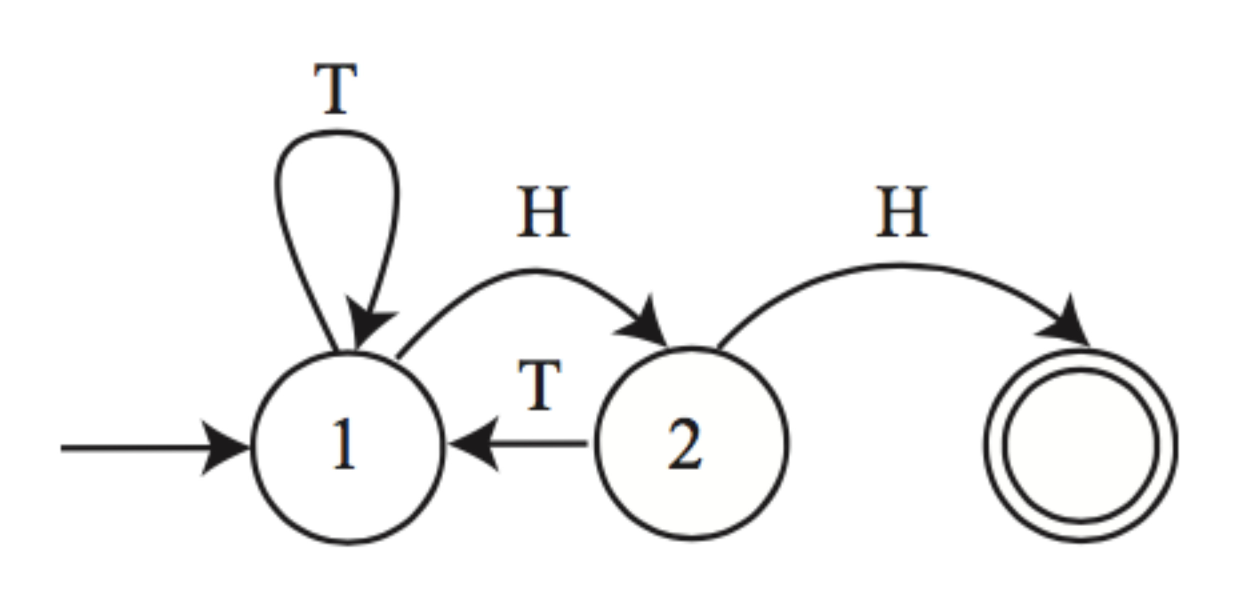
Something happens, we observe, then transition to a new state by following the appropriate edge on the diagram

Finite state diagrams



Many problems admit a formulation in terms of being describable as a random sequence of states where we can ask questions like “what is the probability of being in a given state after some number of time steps?”

Finite state machine



Get from state 1 to end in exactly N moves

If N < 2, it can't be done

If N=2 take edge H and then edge H

Else

Take edge T and then get from 1 to end in N-1 moves

Or take edge H, then T, and then get from 1 to end in N-2 moves

$$\begin{aligned} P(N) &= P(T)P(N-1) + P(HT)P(N-2) \\ &= (1/2)P(N-1) + (1/4)P(N-2) \end{aligned}$$

Motivation 2: gambler's ruin

Suppose you're playing a game of chance and you win with probability p and lose with probability 1-p. You bet $1 in every round of play, if you win, you get your bet back plus $1, if you lose, you lose your $1

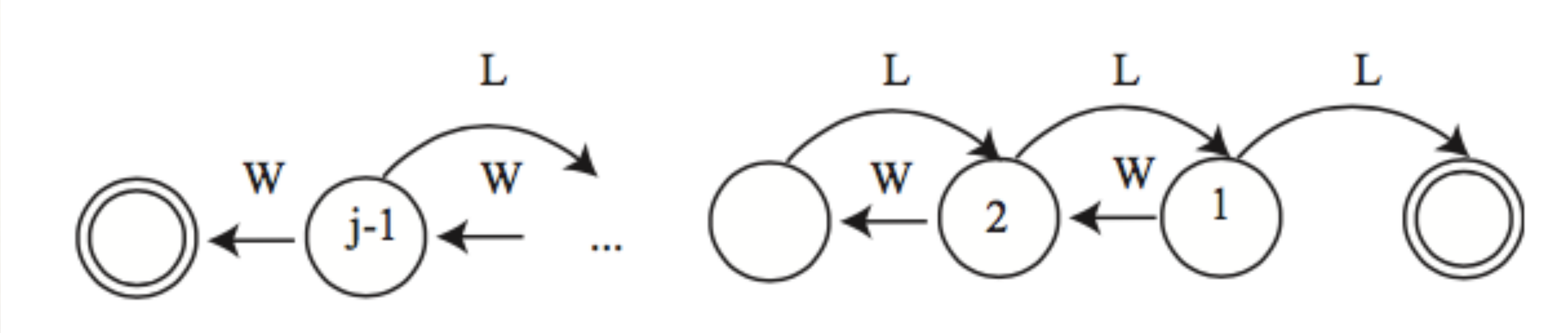
Assume you have $s to start. You'll keep playing until you've lost all your money or you've accumulated $j

Write p_s to represent the probability that you leave with nothing ("ruined") given that you start with $s

What is p_0? p_0 =1

What is p_j? p_j = 0

Gambler's ruin



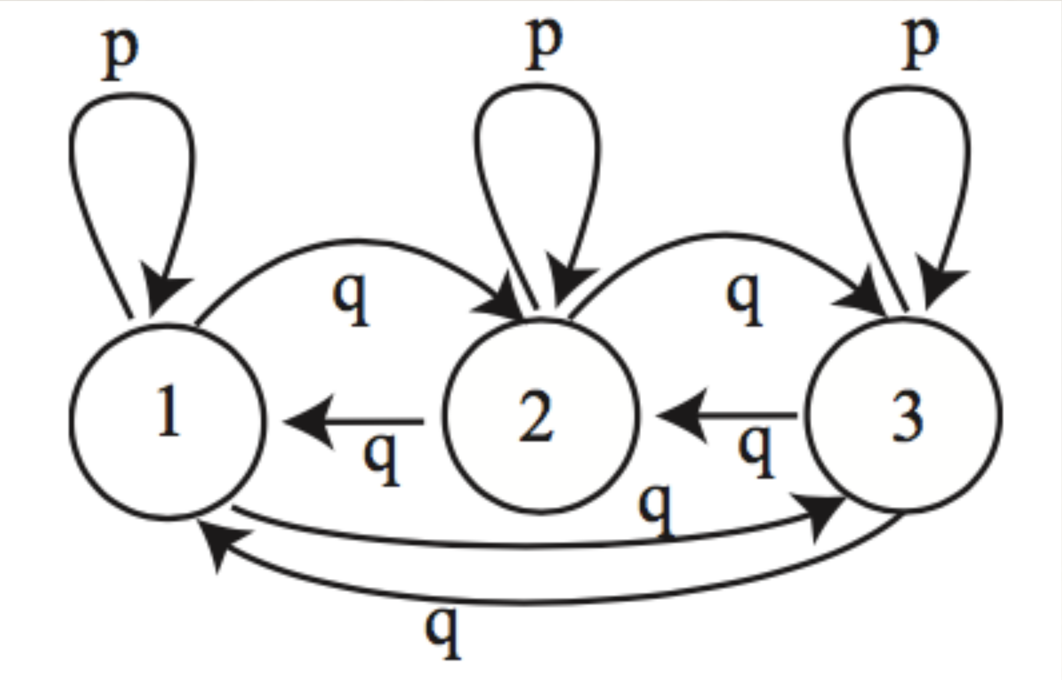
Write a recurrence relation for p_s

$$p_s = pp_{s+1} + (1-p)p_{s-1}$$

Motivation 3: viruses

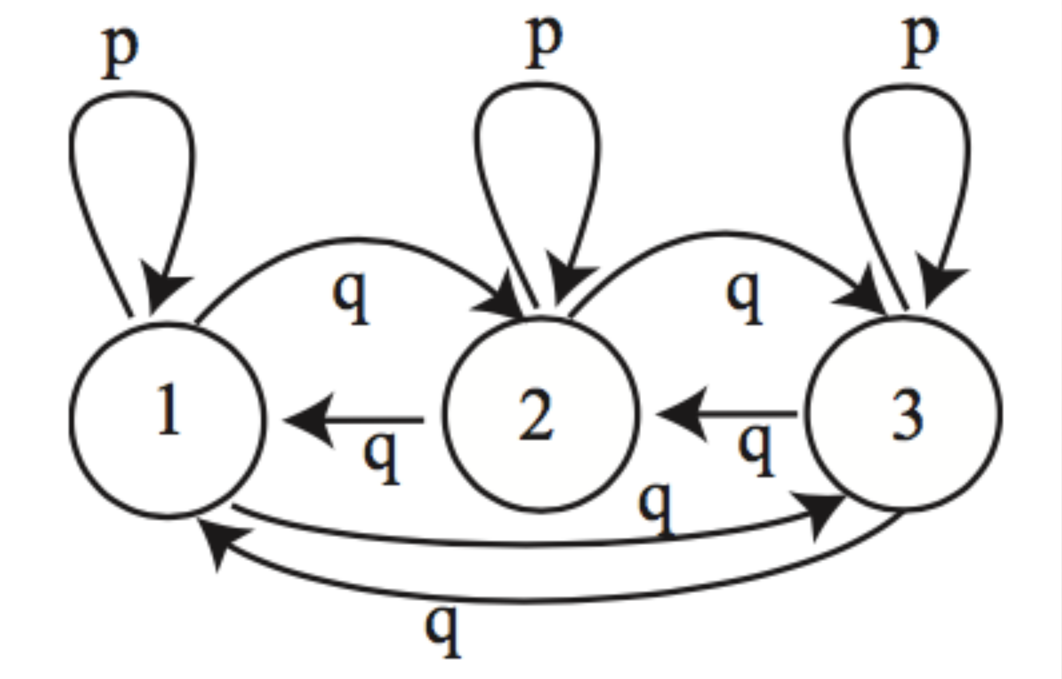
These kinds of finite state problems don't need to have end states

Consider a virus that can exist in one of 3 strains. At the end of the year it mutates with probability *a* and if it does it changes to one of the other strains uniformly at random. Then with p = (1-a) and q=a/2 this state diagram is appropriate



Note that the edges are now transition probabilities

Markov Chains



Picture a bug being placed on one of the nodes of this weighted directed graph. Its initial placement is determined by some probability distribution. Then, it walks from node to node according to the transition probabilities indicated in the arrows.

As it walks, the states that it visits (state 1 through state k) are a sequence of random variables. The random variable X_n=j if the bug was on state *j* at time step *n*

Markov chains

If we have such a setup—a distribution for the initial state, a set of transition probabilities for each state, and a sequence of random variables that take values indicating the state of the process at time *n*—and one additional constraint is met we have a Markov chain

The additional constraint is that the probability that the process is in state *j* at time *n* depends only on the state that the process is in at time *n-1*

$$P(X_n = j | \text{all previous states visited}) = P(X_n = j | X_{n-1})$$

Markov chains

Any model built by putting probabilities on the transitions of a finite state diagram is a Markov Chain but this isn't the only way to build or represent them

Another representation is to use a matrix to encode the transition probabilities from state *i* to state *j*

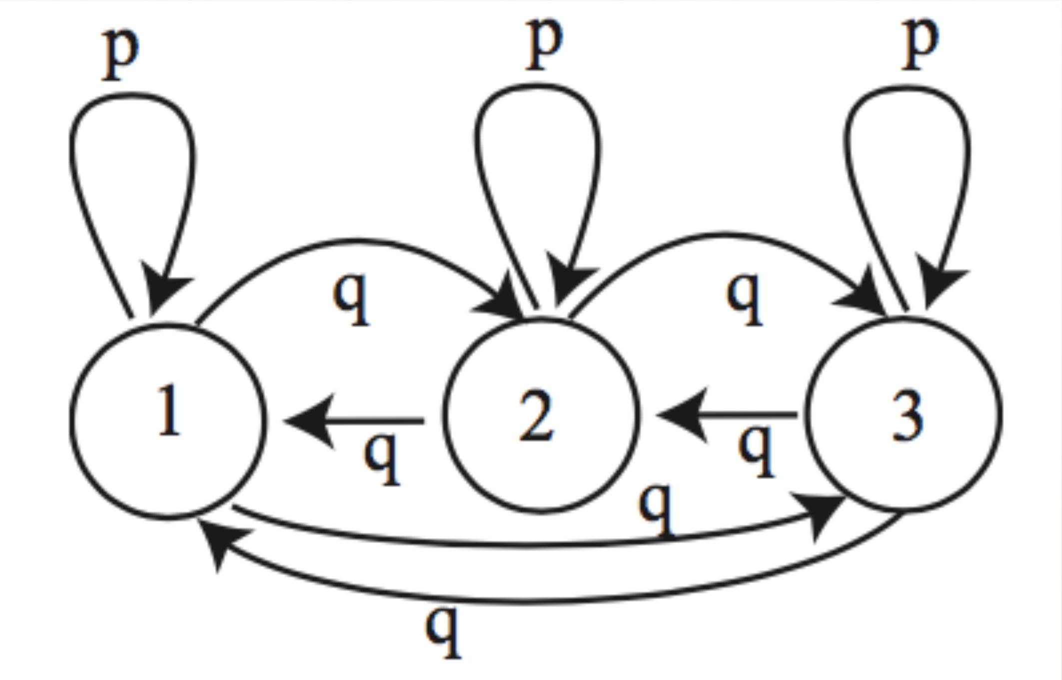
Define a matrix *P* such that

$$p_{ij} = P(X_n = j | X_{n-1} = i)$$

Note that this matrix will satisfy $p_{ij} \geq 0$ and $\sum_j p_{ij} = 1$

Example

Write the transition matrix for the virus example with a=0.2



p = 1-a

q = a/2

Recall:

$$p_{ij} = P(X_n = j | X_{n-1} = i)$$

$$P = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

Markov chains

Write π for the *k* dimensional row vector corresponding to the probability distribution of the initial state, i.e. at time t=0

So for example, when we write π_i it means $P(X_0 = i)$

Then if we want to calculate a probability distribution for the state at time 1, we have

$$
\begin{aligned}
P(X_1 = j) &= \sum_i P(X_1 = j, X_0 = i) && \text{law of total probability} \\
&= \sum_i P(X_1 = j | X_0 = i) P(X_0 = i) && \text{Bayes rule} \\
&= \sum_i p_{ij} \pi_i . && \text{Transition matrix and initial distribution}
\end{aligned}
$$

Markov chains

If we write $\mathbf{p}^{(n)}$ to represent the probability distribution over states at time n then what we figured out on the last slide is that

$$\mathbf{p}_j^{(1)} = \sum_i p_{ij}\pi_i \qquad \text{or} \qquad \mathbf{p}^{(1)} = \pi\mathcal{P}$$

Now consider time 2

$$\begin{aligned} P(X_2 = j) &= \sum_k P(X_2 = j, X_1 = k) \\ &= \sum_k P(X_2 = j | X_1 = k) P(X_1 = k) \\ &= \sum_k P(X_2 = j | X_1 = k) \mathbf{p}_k^{(1)} \\ &= \sum_k p_{kj} \sum_i p_{ik}\pi_i \end{aligned}$$

I.e. we have

$$\mathbf{p}^{(2)} = \pi P^2$$

Markov chains

And in general for Markov chains with transition probabilities given by P and initial state given by π we have

$$\mathbf{p}^{(n)} = \pi P^n$$

Example

With our earlier example for viruses, for a=0.2, we can compute the distribution of states after two state transitions, given that the virus starts in strain 1 as

$$\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}^2 = \begin{bmatrix} 0.66 & 0.17 & 0.17 \end{bmatrix}$$

We also have

$$\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}^{20} = \begin{bmatrix} 0.3339 & 0.3331 & 0.3331 \end{bmatrix}$$

Over time, the process seems to "forget" which state it starts in

Stationary distribution

This forgetting is not atypical. As long as it is possible to reach any state in a Markov chain from any other state, the chain is called **irreducible**. And irreducible Markov chains exhibit the following

For any initial state distribution there is a unique vector *s* called the **stationary distribution** of the Markov chain, given by

$$\lim_{n \to \infty} \pi \mathcal{P}^{(n)} = \mathbf{s}.$$

PageRank

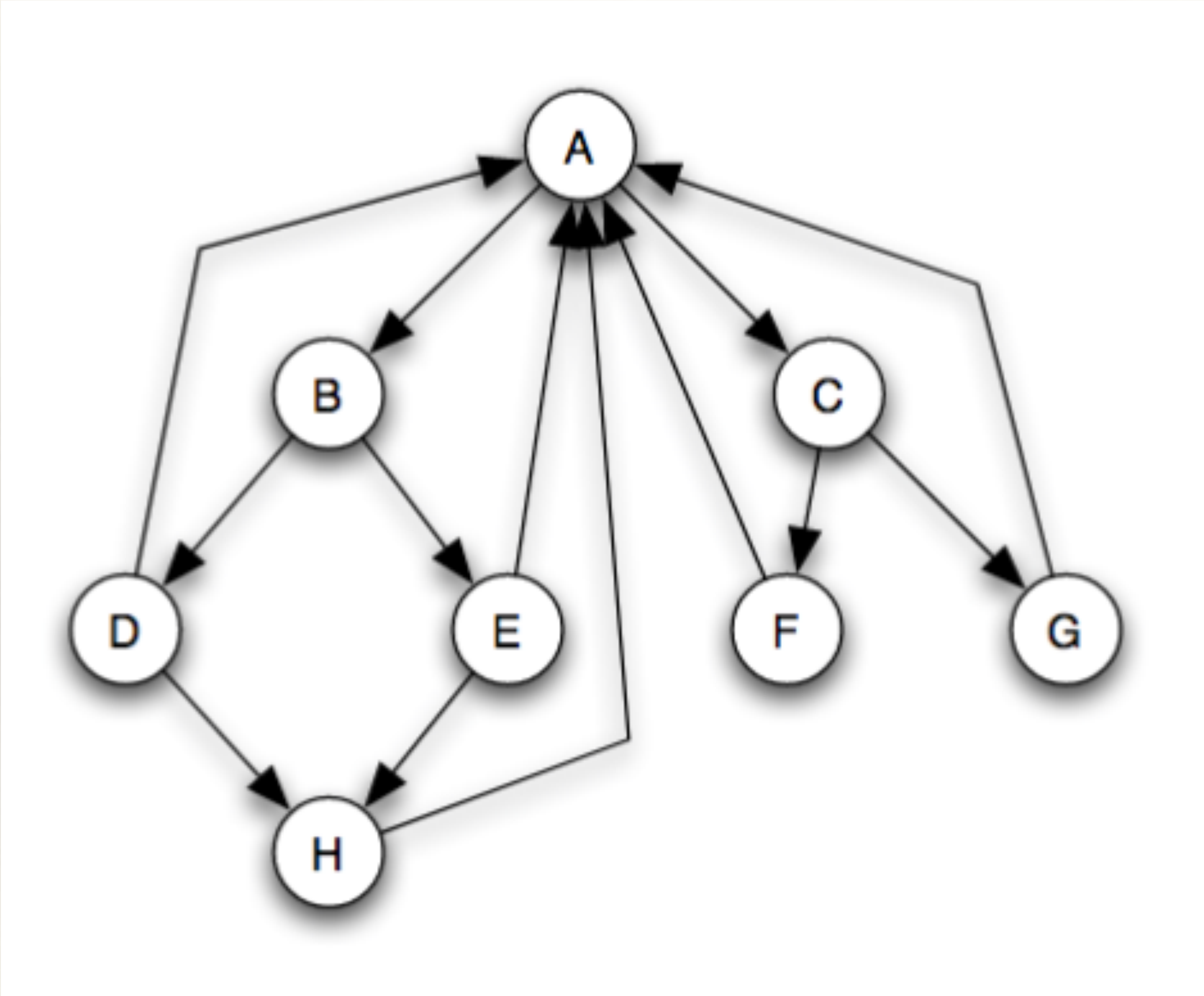
Linking and importance

The goal of a ranking system is to sort a set of objects by their importance

The web, academic publications, blog posts, etc form a network where the links or citations from one node to another can be viewed as an endorsement of the importance of the linked node

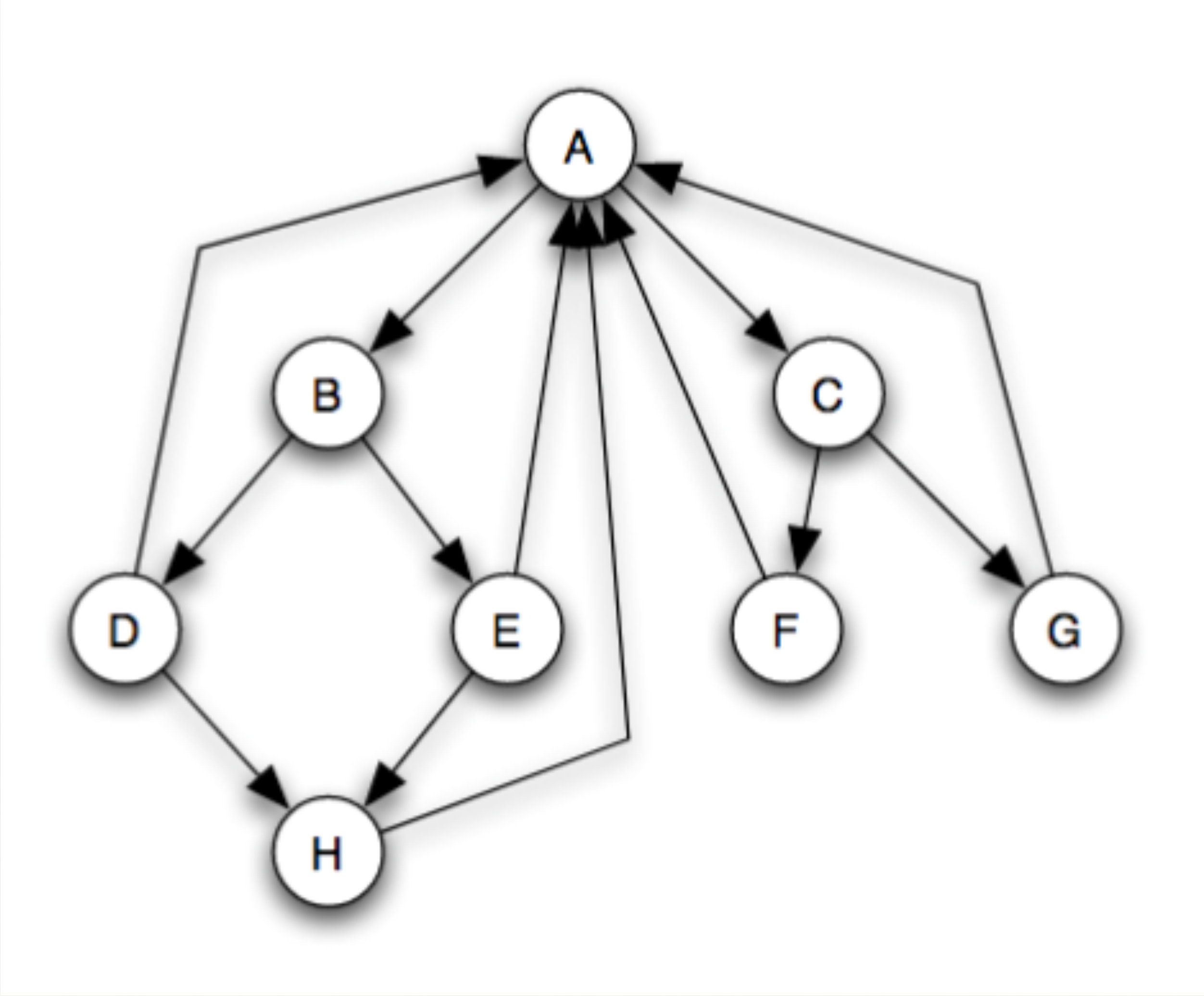
The key principle behind PageRank is recursive: a webpage is important if important pages link to it

The web as a graph



We will use the structure of this graph to calculate an importance score for each page

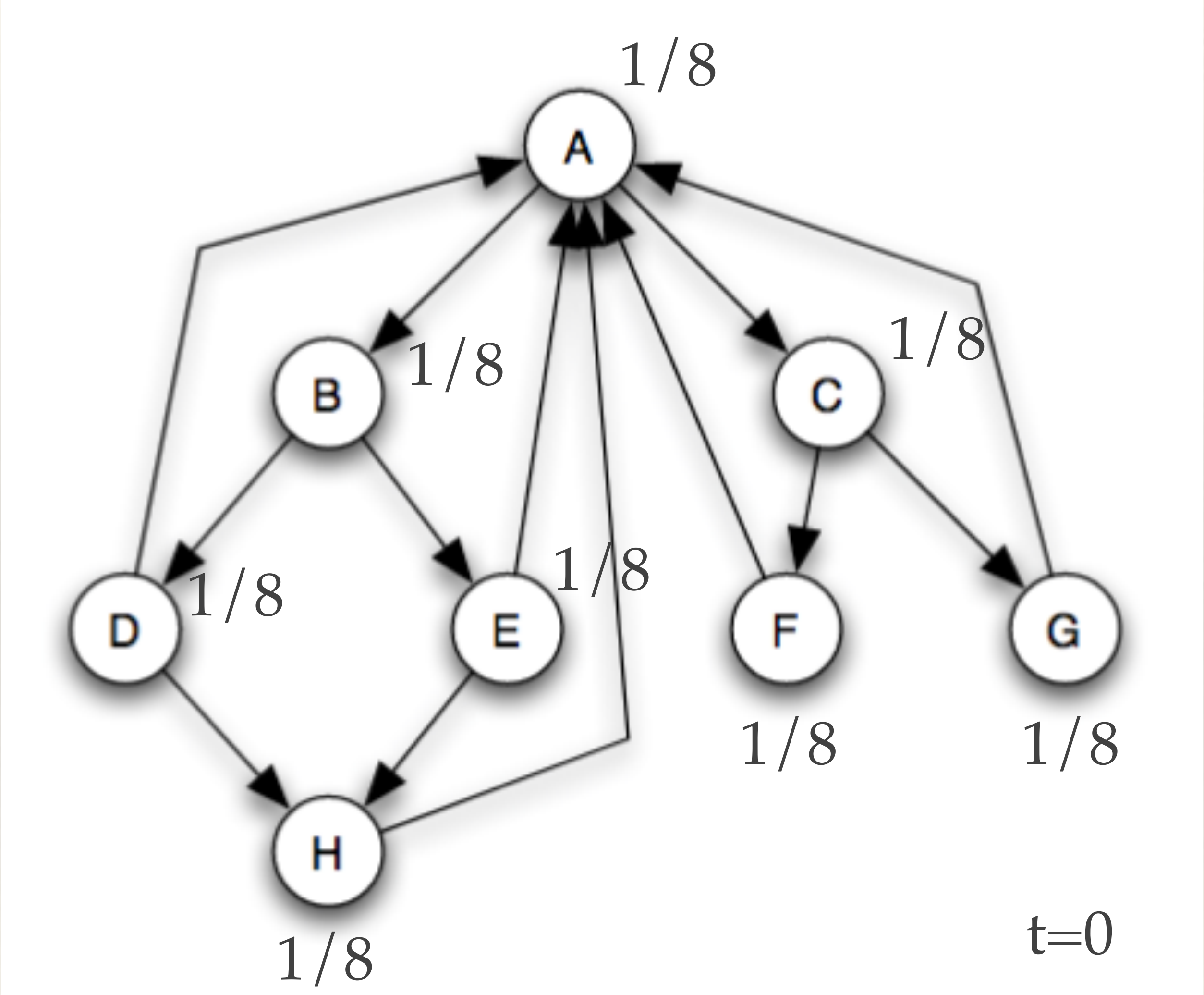
PageRank



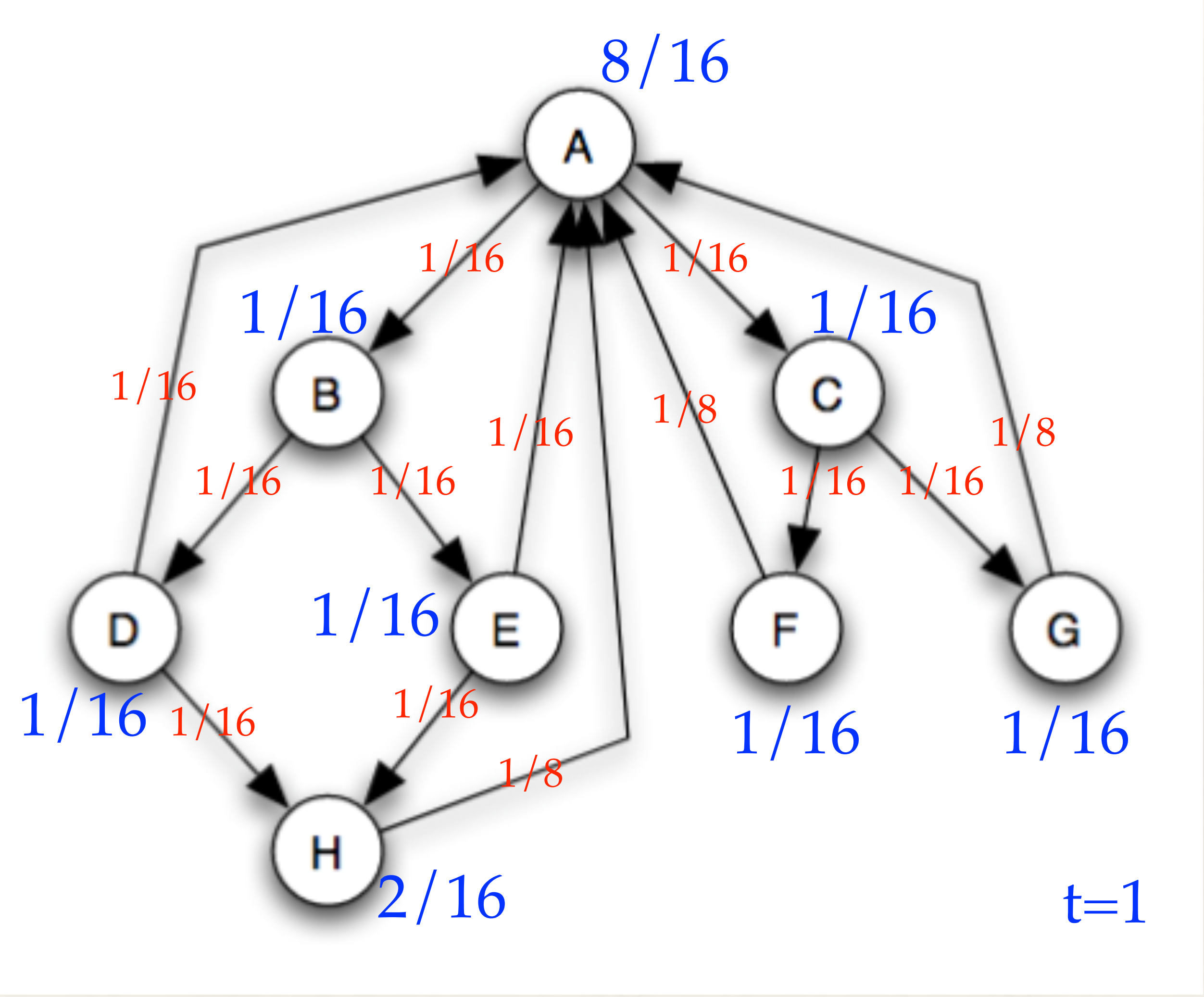
Initialize: Each node starts with 1/n importance

Update: And at each step, propagate importance evenly across outgoing links

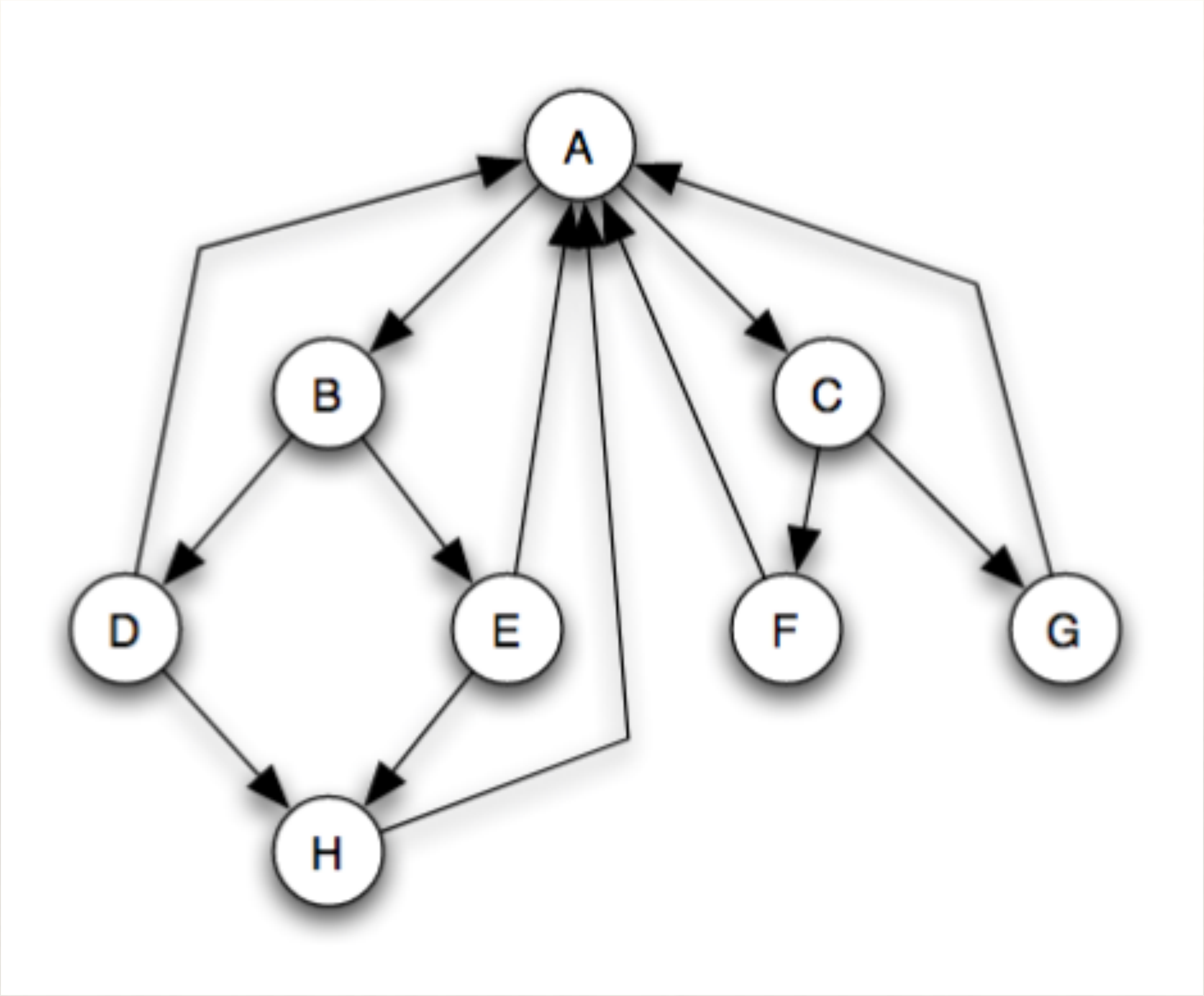
PageRank



PageRank


8/16
A
1/16
1/16
1/16
1/16
B
C
1/16
1/16
1/8
1/8
1/16
1/16
1/16
1/16
D
1/16
E
F
G
1/16
1/16
1/16
1/16
1/16
1/8
H
2/16
t=1

PageRank

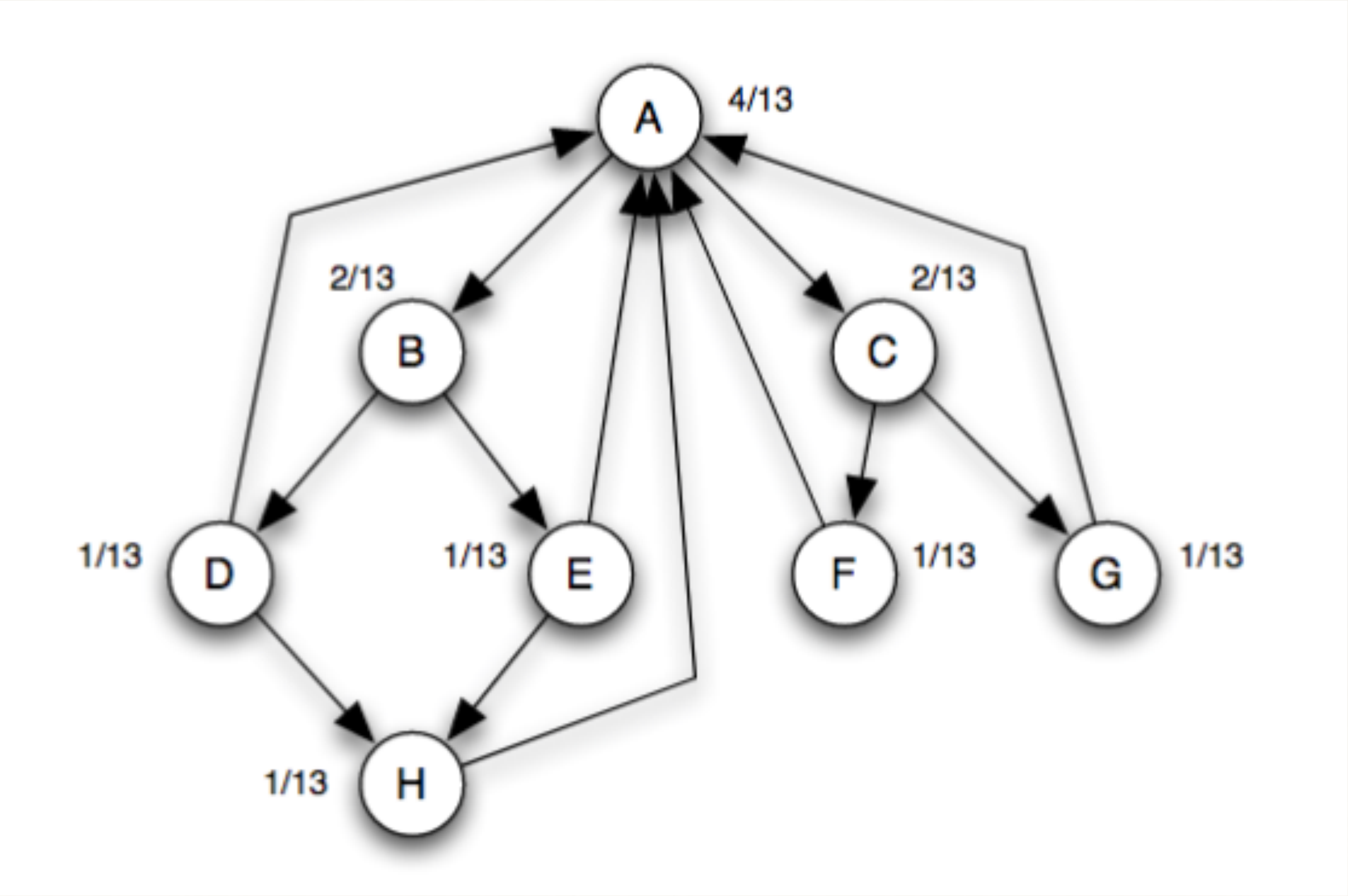


Step	A	B	C	D	E	F	G	H
1	1/2	1/16	1/16	1/16	1/16	1/16	1/16	1/8
2	3/16	1/4	1/4	1/32	1/32	1/32	1/32	1/16

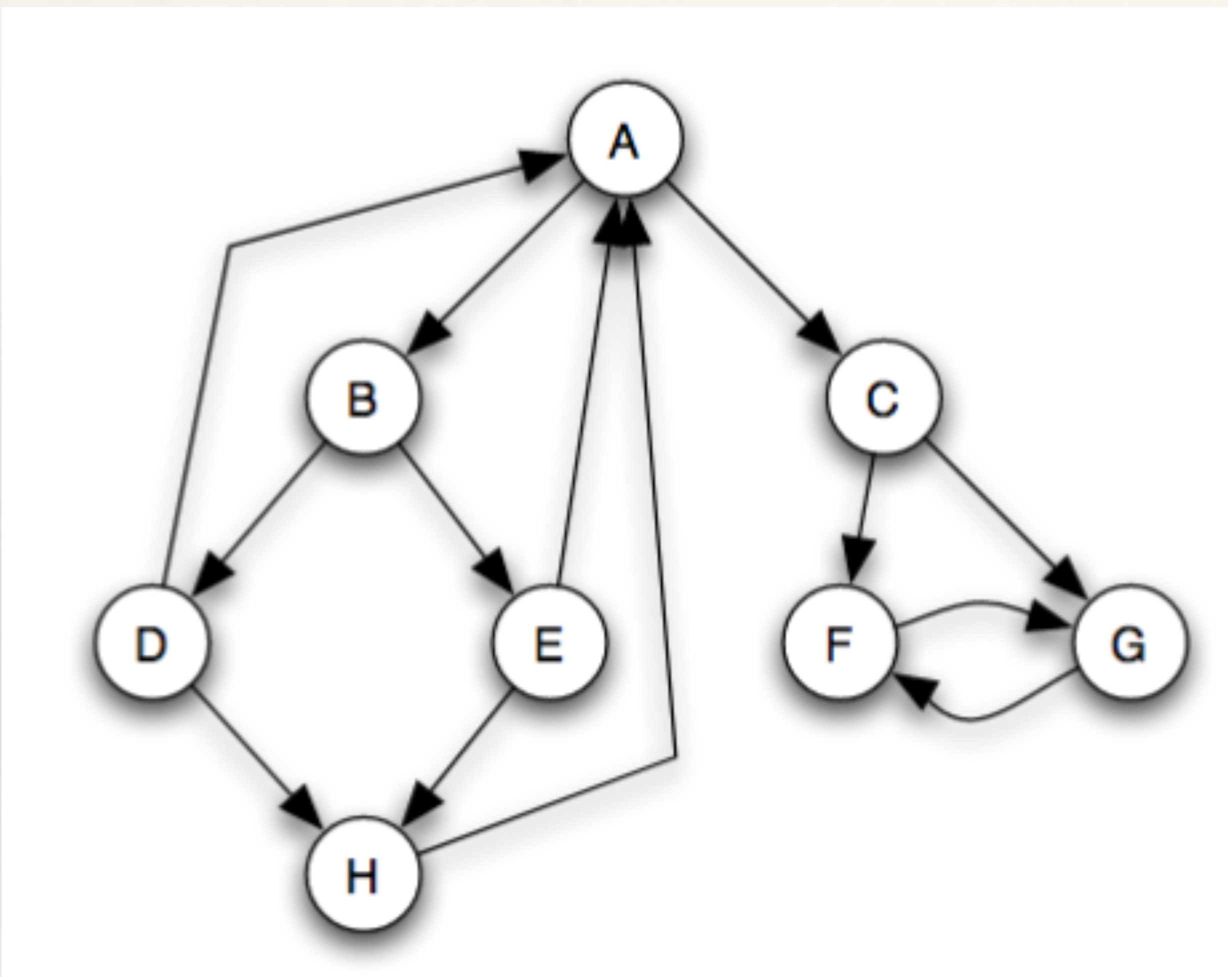
Note how at every step the total PageRank for all pages sums to 1

Equilibrium

Equilibrium is reached when the difference in the PageRank at every node changes very little between step k-1 and step k



What happens here?



All of the PageRank gets trapped in nodes F and G

We fix this by doing the PageRank update as usual, but afterwards, scale down the PageRank at each node by a factor s and then add (1-s)/n PageRank to every node

Teleporting some of the PageRank. In practice we usually have s=0.8 or 0.9

Probabilistic interpretation

1/(number of outlinks from page i) if link i -> j exists, 0 otherwise

$$P(X_k = j) = \sum_i P(X_k = j | X_{k-1} = i) P(X_{k-1} = i)$$

Amount of PageRank in page j after k steps

Amount of PageRank in page i after k-1 steps

So if we define

$$\pi = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix}$$

$$p_{ij} = \begin{cases} 1/\deg(i) & (i,j) \in G \\ 0 & \text{otherwise} \end{cases}$$

We have a Markov chain with $\mathbf{p}^{(k)} = \pi P^k$

Probabilistic interpretation

This justifies our interpretation of PageRank vector as being the proportion of time a random web surfer would spend at each page

Teleportation in this interpretation captures the fact that you can always type in a new address in the address bar

With teleportation we have

$$p_{ij} = \begin{cases} s/\deg(i) + (1-s)/n & (i,j) \in G \\ (1-s)/n & \text{otherwise} \end{cases}$$

Probabilistic interpretation

The way we have defined things gives us a Markov chain

$$\pi = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix} \qquad p_{ij} = \begin{cases} s/\deg(i) + (1-s)/n & (i,j) \in G \\ (1-s)/n & \text{otherwise} \end{cases}$$

Furthermore it gives an irreducible Markov chain since there is a path from every state to every other state. This means the Markov chain has a stationary distribution **v**

$$\lim_{n \to \infty} \pi P^{(n)} = \mathbf{v}$$

And entry *i* of this column vector gives the PageRank of web page *i*