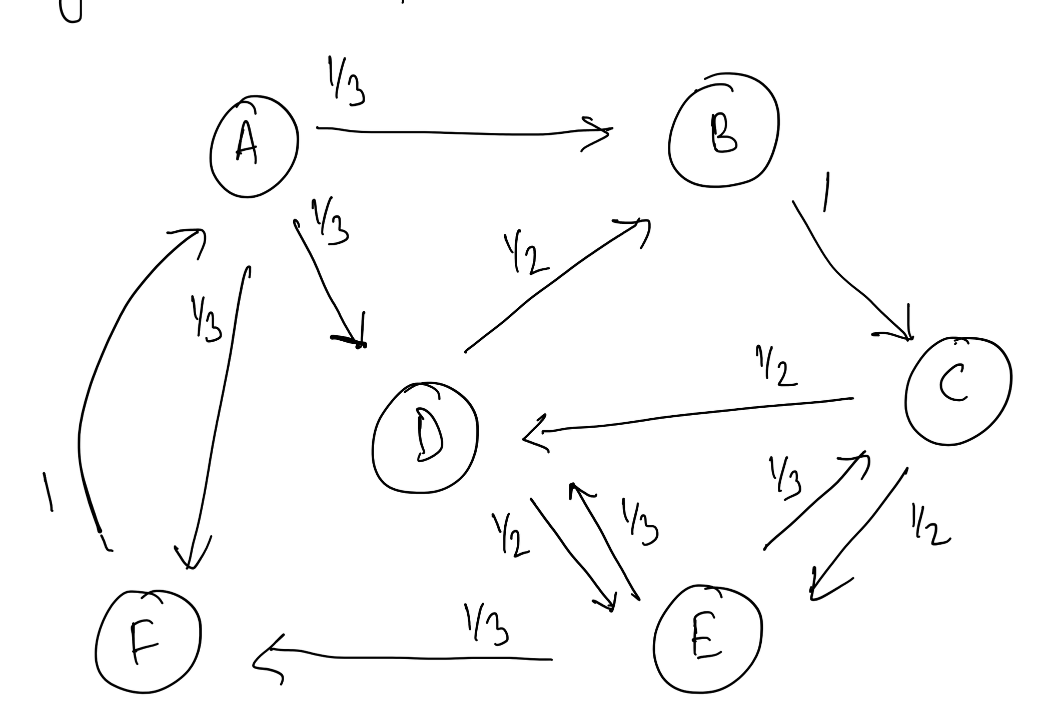
Page Rank Algorithm

Wednesday, 24 July 2024 10:54 PM

Developed by Larry Page + Sergey Brin in 1996 to determine relevance of webpages

Tiny model of the web:



Consider a user that when visiting a page, clicks a link chosen uniformly at random from the links on a page.
At any point of time, for any page P, what is the probability that the user is on page P?

$$\text{Let } H = \begin{array}{c|cccccc} & A & B & C & D & E & F \\ \hline A & & 1/3 & & 1/3 & & 1/3 \\ B & & & 1 & & & \\ C & & & & 1/2 & 1/2 & \\ D & & 1/2 & & & 1/2 & \\ E & & & 1/3 & 1/3 & & 1/3 \\ F & 1 & & & & & \end{array}$$

For pages P, P', $H_{P,P'}$ is the probability that the user goes to P', given that she is on page P.

Suppose $\pi(t)$ denotes the vector of probabilities of this user at time t. So $\pi_P(t)$ is prob. user is at page P at time t.
Suppose further $\pi(0) = \frac{1}{6}[1 \; ... \; 1]^T$, i.e., the user is at a uniformly random page at $t = 0$.

Then $\pi_P(1) = \sum_{P'} \pi_{P'}(0) H_{P',P}$

$$\Rightarrow \quad \pi(1)^T = \pi(0)^T H$$
$$\pi(2)^T = \pi(1)^T H = \pi(0)^T H^2$$
$$\dots \quad \pi(t)^T = \pi(0)^T H^t$$

(This is what is called a Markov chain)

Is there a steady-state probability distribution? Or will probability keep on changing at every step?

want: $\pi \in \mathbb{R}^n_+ : \pi^T = \pi^T H$
& $\pi^T \mathbb{1} = 1$

OR: $\boxed{\pi^T (I - H) = 0}$

If the Markov chain (in this case, H) is irreducible & aperiodic, then π exists & is unique.
(called the Fundamental Theorem of Markov Chains)

Ok, but will we ever reach π?
π is called a stationary distribution
Yes, if H is irreducible & aperiodic.
Then for any starting distribution $\pi(0)$,

$$\pi^T = \lim_{n \to \infty} \pi(0)^T H^n$$

(and we can even obtain bounds on the convergence)

Lemma 1: Let H be the transition matrix for a aperiodic, irreducible Markov Chain. The the largest eigenvalue $\lambda_1 = 1$.

Proof: By the Fundamental Theorem, we know that $\exists \pi$ s.t. $\pi^T H = \pi^T$, hence $\exists$ an eigenvalue with value 1. So we only need to show that all eigenvalues have value ≤ 1.

(try & prove this yourself)

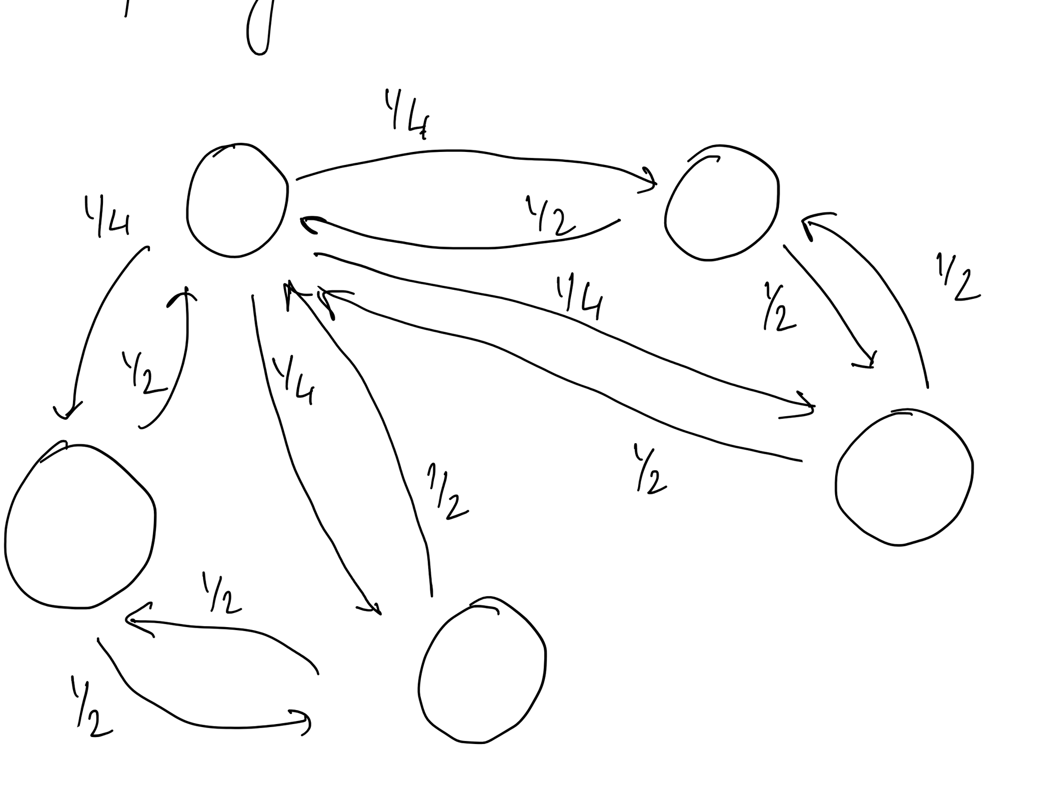
(I) Fact: Left eigenvalues = Right eigenvalues.
Hence we only need to show that if $\exists x, \lambda$ s.t. $Hx = \lambda x$, then $\lambda \leq 1$

(II) Consider the matrix $H^2 = H \times H$.
Can check that H^2 is also a transition matrix:
- all entries ≥ 0, &
- each row sums to 1

Thus so is H^k for all $k > 0$.

Now suppose $\exists x, \lambda$ s.t. $Hx = \lambda x$, $\lambda > 1$.
then $H^k x = \lambda^k x$
But each entry on the left is bounded, while $\lambda^k x$ grows unboundedly. $\boxtimes$

Exercise: Consider a Markov chain with bidirected edges, & from a vertex, an edge is chosen uniformly at random.



Then prove that the stationary distribution is:

$$\pi_v = \frac{\deg(v)}{2|E|}$$
