PageRank devised by Brin & Page in 1998
is an algorithm to determine the "importance" of webpages

V = webpages
E = directed edges for hyperlinks

for page $x \in V$,
let $O(x) = \{y : x \to y \in E\}$ = outgoing edges from x
$I(x) = \{y : y \to x \in E\}$ = incoming edges from y

Let $\pi(x)$ = rank of page x
We are trying to define $\pi(x)$ in a sensible way.

<u>Idea</u>: citation counts —
more citations, more important
where a citation is a link to the page.

So set $\pi(x) = |I(x)|$ = # of links to x.

<u>But</u> if page y has only 1 outgoing link

& it's to x

that's more valuable than a link from a Page z with many outgoing links.

Thus, if page y has $|O(y)|$ outgoing links

then each page x with a link from y

gets $\frac{1}{|O(y)|}$ of a citation.

Therefore, set $\pi(x) = \sum_{y \in I(x)} \frac{1}{|O(y)|}$

<u>But</u> if an important page like Google has a link to x that's more valuable than a link to x from Eric Vigoda's webpage.

So define $\pi(x)$ recursively.

Page y has importance $\pi(y)$

So each link from y gets $\frac{\pi(y)}{|O(y)|}$

Therefore,

$$\pi(x) = \sum_{y \in I(x)} \frac{\pi(y)}{|O(y)|}$$

This is a recursive definition of π, how do we find it?

Think of the random walk on the graph $G=(V,E)$.

From a page $y \in V$,

we choose a random link on y & click it.

This is a Markov chain.

If $x \to y \in E$ then

$$P(x,y) = \Pr(X_{t+1} = y \mid X_t = x) = \frac{1}{|O(x)|}$$

What is the stationary distribution of this Markov chain defined by P?

This is the distribution π where

$$\pi = \pi P$$

Thus, $$\pi(x) = \sum_{y \in V} \pi(y) P(y,x)$$

$$P(y,x) = \begin{cases} \frac{1}{|O(y)|} & \text{if } y \to x \in E \\ 0 & \text{otherwise} \end{cases}$$

So $$\pi(x) = \sum_{y \in I(x)} \frac{\pi(y)}{|O(y)|}$$

This is exactly what we want.

So the ranking π is the stationary distribution of the random walk on the webgraph.

Is this π the <u>unique</u> (only) stationary distribution?

In other words is this Markov chain ergodic?

Need that G is strongly connected.

It probably isn't.

And some pages y have $O(y) = \emptyset$

No outgoing links.

then hit the "random" button.

Introduce "damping factor" α where $0 < \alpha \leq 1$.

(in practice apparently use $\alpha \approx .85$)

With probability α go to a random outgoing link from the current page y.

and with probability $1-\alpha$ go to a completely random page.

Thus, let $N = |V| = \#$ of webpages.

$$P(y,x) = \begin{cases} \frac{(1-\alpha)}{N} + \frac{1}{|O(y)|} & \text{if } y \to x \in E \\ \frac{(1-\alpha)}{N} & \text{otherwise} \end{cases}$$

This new MC is ergodic.

Thus it has a unique stationary distribution π.

How to find π?

Take previous π, compute P^t for big t,

then compute πP^t & reset π to this

Consider undirected $G=(V,E)$. Let $n=|V|$.

Look at random walk on G.

From a vertex $x \in V$,

choose a random neighbor y.

Suppose G is connected so this Markov chain is irreducible.

To ensure it's aperiodic, let's do the following.

States are vertices V.

From $X_t \in V$,

1) Choose a random neighbor y of X_t.

2) With probability $\frac{1}{2}$, set $X_{t+1} = y$

& with probability $\frac{1}{2}$, set $X_{t+1} = X_t$.

Let $d(x)$ = degree of vertex x.

Then the transition matrix P is

$$P(x,y) = \begin{cases} \frac{1}{2d(x)} & \text{if } (x,y) \in E \\ 0 & \text{otherwise} \end{cases}$$

$$\& \ P(x,x) = \frac{1}{2}.$$

What is the unique stationary distribution?

If G is d-regular,

so $d(x) = d$ for all $x \in V$.

Then P is symmetric $P(x,y) = P(y,x) = \frac{1}{2d}$ if $(x,y) \in E$

Thus π is uniform over V.

Otherwise,

Claim: $\pi(x) \propto d(x)$

$\uparrow$ proportional to

Let $m = |E|$.

Note, $\sum_{x \in V} d(x) = 2m$.

Thus, we are claiming that: $\pi(x) = \frac{d(x)}{2m}$

Proof:

$\pi(y) = \frac{d(y)}{2m}$ & we need to check that $\pi = \pi P$.

Checking entry y:

$$(\pi P)(y) = \sum_{x \in V} \pi(x) P(x,y)$$

$$= \sum_{x:(x,y) \in E} \pi(x) \frac{1}{d(x)}$$

$$= \sum_{x:(x,y) \in E} \frac{1}{2m} \quad \text{since } \pi(x) = \frac{d(x)}{2m}$$

$$= \frac{d(y)}{2m} = \pi(y) \quad \checkmark$$