Lecture 9 - PageRank

Thursday, February 8, 2024 9:05 AM

Joke: What professional football team used the power of graph theory to 'tour' their way to the 1960 and 1961 AFL championship?

A: Houston Eulers

(Oilers)

PageRank

"A modern classic"

- Gittens

PageRank: a centrality algorithms

↳ importance

BITD: a lot of internet,

but no good way

to find anything

Seach wasn't well established

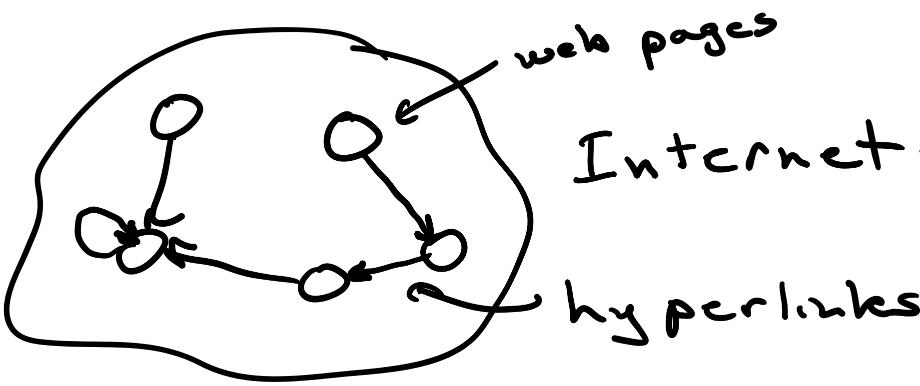
→ keyward matching was

generally a prime technique

generally a prime technique
Yahoo, Altavista

Google: we'll incorporate "trust" in our searches

Trust: reliable site, what you're actually looking for

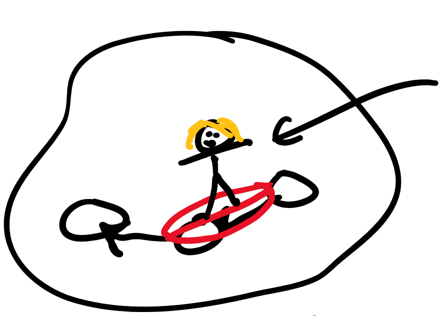


Internet = big ol' graph

"Trust" → you're trusted if other "trusted sites" link to you

---

Random surfer model (walk)



Internet

random surfer
↳ performing an infinite random walk

PageRank: probability that our random surfer is at some vertex at some point in time on an infinite walk

Issues: what about $d^+(v) = 0$ ← sink
$d^-(v) = 0$ ↑ source

↳ randomly jump from sinks to some other vertex

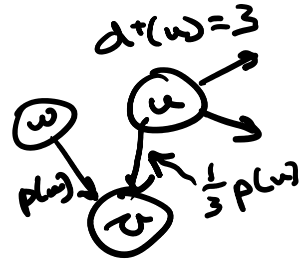
Graph Algorithmic Model

↳ vertex centric computation over some iterations

PageRank

initialize $p(v) = \dfrac{1}{|V(G)|}$ for all $v \in V(G)$

we iterate $p(v) = \sum\limits_{u \in N^-(v)} \dfrac{p(u)}{d^+(u)}$



For source / sinks:

For source/sinks:

sinks: add edges to all $v \in V(G)$

(surfer randomly jumps from sink)

Q: what if there are sources but no sinks?

A: we'll also randomly jump after each step on the random walk with same probability p (damping factor)

Linear Algebraic Model

(Graphs = Matrices)

Consider our adjacency matrix A

Consider the <u>diagonal degree matrix</u> D

$$D_{ii} = \sum_j A_{ij}$$

$$D_{ij} = 0 \quad \forall i \neq j$$

We also have the transitional probability matrix M^T (fix notation next year) we also have the transitional probability matrix M^T (fix notation next year)

$M^T = D^{-1}A$ ← define transitional probs, following out edges

$M = (D^{-1}A)^T$ ← in edge transition probabilities

Q: How can we use this to compute PageRanks?

$$p_0(v) = \frac{1}{|V(G)|} \quad \forall v \in V(G)$$

↑ PR on iteration zero

$p_{i+1} = M p_i$ ← vector of all pageRanks at iteration i

eventually, after n iterations ($n \to \infty$)

$$\underbrace{p_\infty = M p_\infty}$$

aka $\|p_{i+1} - p_i\| < \epsilon$

recall for some matrix A

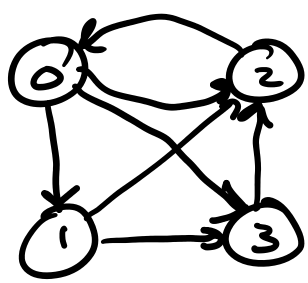
$$\underbrace{Ax = \lambda x}$$

vector, scalar

eigenvector, eigenvalue

Takeaway: PageRanks are just the eigenvector of transition probability matrix with eigenvalue of $\lambda = 1$

Example PR calculation



$$A = \begin{array}{c|cccc} & 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 2 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 1 & 0 \end{array}$$

$$D = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad D^{-1} = \begin{bmatrix} 1/3 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$M^T = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad M = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$

$$p_0 = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} \qquad p_1 = M p_0 = \begin{bmatrix} 1/4 \\ 1/12 \\ 1/12 + \frac{1}{8} + \frac{1}{4} = \frac{11}{24} \\ 5/24 \end{bmatrix}$$

$\frac{1}{4}$ $\qquad\qquad$ $\frac{5}{24}$

Competition Networks

vertices = competitors

edges = competitions

Orient: add a direction to on an undirected edge

For PageRank on a competition network:

- we can orient each edge to point to the victors
- PR "flows" to winners

↳ gives us a ranking of competitors and allows us to make predictions on future matches

Modifications:

Can accounting for margins of victory

↳ by weighting edges in many ways (margin of victory,

ways (margin of victory,
quality of competitor)

Higher weighted edges get
a high proportion of PR

Today's example:

We have data for the
2023-2024 NFL season

Let's predict the outcome
of the Super Bowl