PageRank

Jeff M. Phillips

April 11, 2018
Final Report

At most 4 pages/student. Don’t cram in too much!

- Succinct title (and names)
- Problem definition and motivation.
- Explain your Data.
- **key idea**
- What did you do (which techniques, an implementation, a comparison, an extension)
- What did you learn? Artifacts (charts, plots, examples, math) and Intuition (in words, did it work?)
Page Rank key technology inside Google Search.

Search Engine?

Inverted index

Big challenges

- Ranking pages for key words
Rank web page on keyword?

→ pie

n-grams on text on page
+ Jaccard ↗ 3
+ Minhashing

Cosine similarity

$\text{cosine similarity } U_{\text{pie}} = \langle 0, 0, 0, \ldots, 0 \rangle$

down page $\langle 1, 2, 4, \ldots, 9, 11, 0 \rangle$

What can go wrong?

Is many copies of "pie"
Add more context.

search [ pie ]

\( \nu_{\text{pie}} = (0,0,0,1,0,0,1), 7, 1 \),

what can break this?

recipe

7 delicious pies

copy entire popular webpages
How did search engines know about pages?

→ Crawlers: goes to webpage, follows links, + records page.

- hyperlinks
  - `<a href="" up ref">text</a>`
  - a very informative words on page
  - put into how rep for page

page how vec
  - $(u_1, u_2, \ldots, u_n)$

and hyperlinks $(u_1, u_2, \ldots, u_n)$
  - $(\bar{u_1, \bar{u_2}, \ldots, \bar{u_n}})$
Index

Pages: hand-curated list of links

1. Google
2. Youtube
3. Facebook
4. Baidu
5. Wikipedia
6. Reddit
7. Yahoo!
8. Google India
9. Tencent AQ
10. Amazon
Idea #1. Pages are important if linked to by other important pages.

Idea #2. How likely a random surfer would find this page.

Markov Chain ← model of web graph. ergodic?
Anatomy of Web

IN

Strongly Connected Component

Tendrils

OUT

ANATOMY of WEB

Lonely

Tendrils

IN

OUT

Tubes

Disconnected

California
Anatomy of Web

not ergodic!
Teleportation (Taxation)

Idea:

15% steps, jump to random page.

\[ g^* = P^n g_0 \]

\[ g_{t+1} = \left( (1 - \beta)P + \beta \Phi \right) g_t \]

\[ P_{\text{ergodic}} \]

\[ g^* = (P)^n g_0 \]

Random Page
\[ \text{eig}(M) = \{ \lambda_1, \lambda_2, \ldots \} \]

\[ M = V L V^T \]

\[ \sqrt{M} = V \sqrt{\Sigma} \]

\[ \sqrt{2} = \begin{pmatrix} \sqrt{\lambda_1} \\ \sqrt{\lambda_2} \end{pmatrix} \]