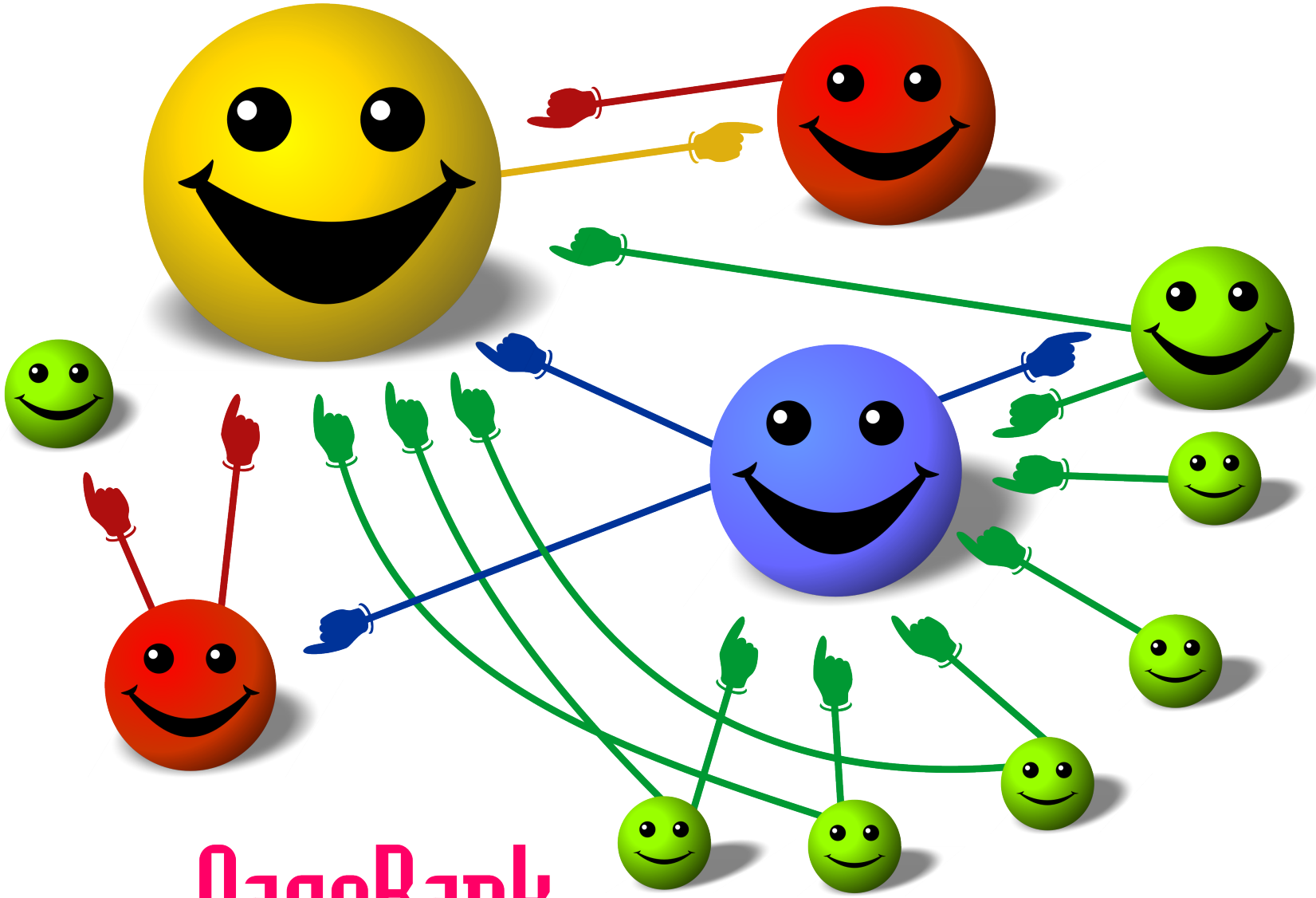
PageRank

Brian Doherty

April 6, 2022


PageRank

Introduction

"PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites." - Google

- Outputs a probability distribution representing the probability that someone randomly clicking on links, who will eventually stop, will arrive at any particular page
- Iterative process
- Links from a page to itself are ignored
- Multiple links from one page to another are treated as a single link
- PageRank transferred from a page to the targets of its links upon the next iteration is divided equally among all its links
- If the surfer arrives at a page with no links to other pages, they pick another page at random and continue surfing again

Simplified algorithm

The PageRank PR of page u is given by

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}. \tag{1}$$

- B_u = set of all pages linking to page u
- $L(v)$ = number of links from page v

Damping factor

The probability at any step that the surfer will continue clicking is called the damping factor d. With this factor, and assuming the surfer switches to a random page if they stop clicking, the equation becomes

$$PR(u) = \frac{1-d}{N} + d \sum_{v \in B_u} \frac{PR(v)}{L(v)}. \tag{2}$$

- N = total number of pages

Frame Title

Application to colleges and universities

- Econ Job Market placement data
- Assistant professor placements
- Create a directed graph where the nodes (pages) are the schools and the edges (links) are the placements
- Which way should the edges point? How does one school link to another?
- In the original application of the algorithm to links, a link was thought of as an endorsement of the page being linked to
- When we observe one school hiring a candidate from another school, this is effectively a revealed endorsement of the candidate's school
- In our model, the hiring school "links" to the candidate's school
- Confusing because the candidate goes to the hiring school

Alternative conceptualization

What does it mean to be a good school?

Alternative conceptualization

What does it mean to be a good school? Good schools place candidates at good schools.

Comparison to reverse directed search model

3 differences

Tier 1 of reverse directed search model (21 schools):
Stanford; Harvard; Yale; Berkeley; Princeton; Chicago; Columbia; Northwestern; Michigan; NYU; LSE; UCLA; Penn; Maryland; Duke; U of T; BU; Minnesota; **Wisconsin (PageRank 26)**; **Cornell (PageRank 22)**; **Ohio State (PageRank 38)**

Top 21 schools according to PageRank:
Stanford; Harvard; Yale; Berkeley; Princeton; Chicago; Columbia; **MIT (Tier 2)**; Northwestern; Michigan; NYU; LSE; UCLA; Penn; Maryland; Duke; **Brown (Tier 2)**; U of T; BU; **Bonn (Tier 2)**; Minnesota

Comparison to another ranking system (RePEc)

8 differences

Top 21 economics departments according to RePEc:
Harvard; MIT; Berkeley; Chicago; **Paris School of Economics (82)**; Princeton; Stanford; **Oxford (24)**; **Barcelona School of Economics (205)**; Columbia; **Toulouse School of Economics (25)**; Yale; NYU; Brown; Penn; BU; **Dartmouth (387)**; **UC San Diego (69)**; Northwestern; **University College London (27)**; **USC (53)**;

Top 21 schools according to PageRank:
Stanford; Harvard; Yale; Berkeley; Princeton; Chicago; Columbia; MIT; Northwestern; **Michigan (24)**; NYU; **LSE (23)**; **UCLA (22)**; Penn; **Maryland (45)**; **Duke (33)**; Brown; **U of T (37)**; BU; **Bonn (95)**; **Minnesota (62)**

University of British Columbia

- PageRank
  - UBC – 35
  - U of T – 18
- Reverse directed search model
  - UBC – Tier 2
  - U of T – Tier 1
- RePEc
  - UBC – 32
  - U of T – 37