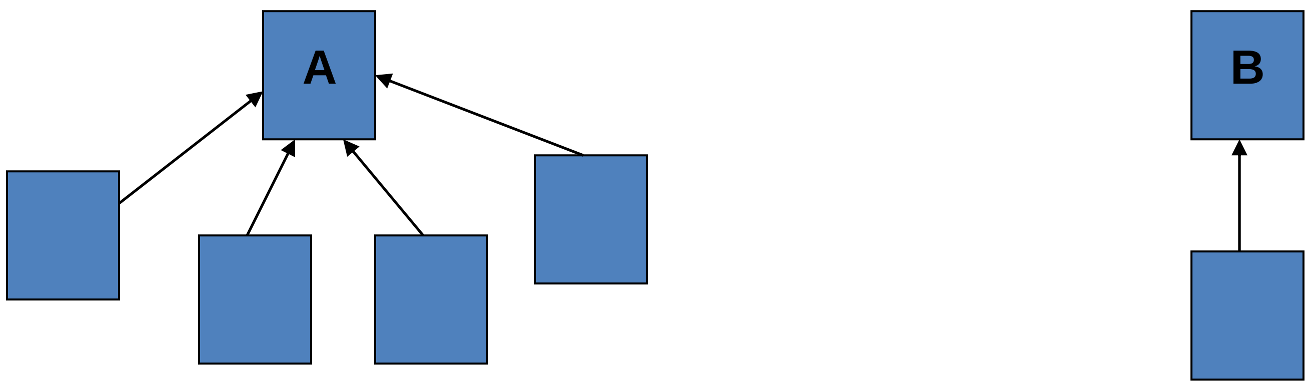
Page Rank:
Google’s Billion Dollar Solution

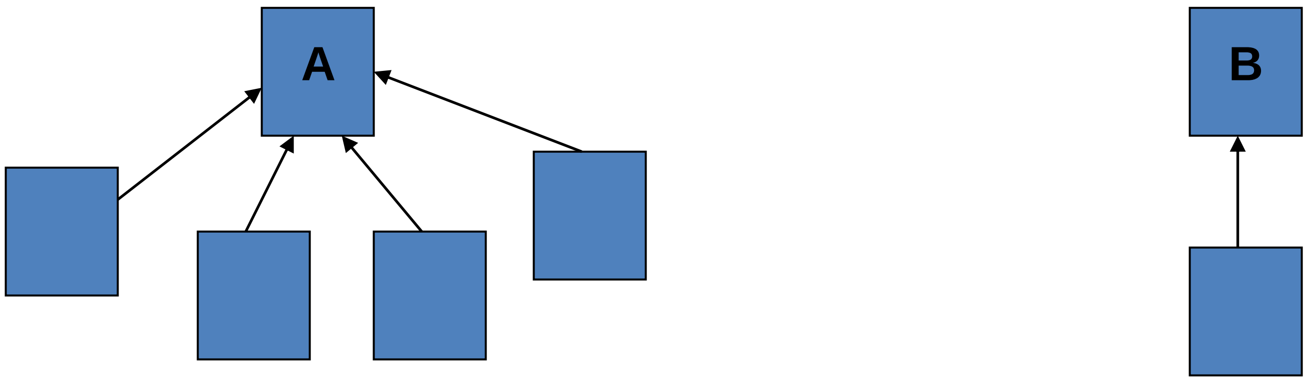
Problem Statement

What do you mean by PageRank?

Which is the more popular page below, A or B?



PageRank is NOT a simple citation index



What if the links to A were from unpopular pages, and the one link to B was from www.yahoo.com ?

Intuitively ***PageRank*** is analogous to popularity

- The web as a graph: each page is a **vertex**, each hyperlink a directed **edge**.

- A page is popular if a few very popular pages point (via hyperlinks) to it.

- A page could be popular if many not-necessarily popular pages point (via hyperlinks) to it.

NOTE: While PageRank is an important part of Google's search results, it is not the sole means used to rank pages.

How do you realize the intuitive thought of page rank?

Realizing PageRank

- Assume that we have a unit of importance to distribute to all nodes.
  - Initially each node gets $\frac{1}{n}$ amount of rank.
- Each node distributes the importance value they have to their neighbors.
- The importance value of each node is the sum of the importance fractions it collects from its neighbors.

$$w_v = \sum_{u \to v} \frac{1}{d_{out}(u)} w_u$$

w_v: the PageRank value of node v

Recursive definition

Example

$w_1 = 1/3\ w_4 + 1/2\ w_5$

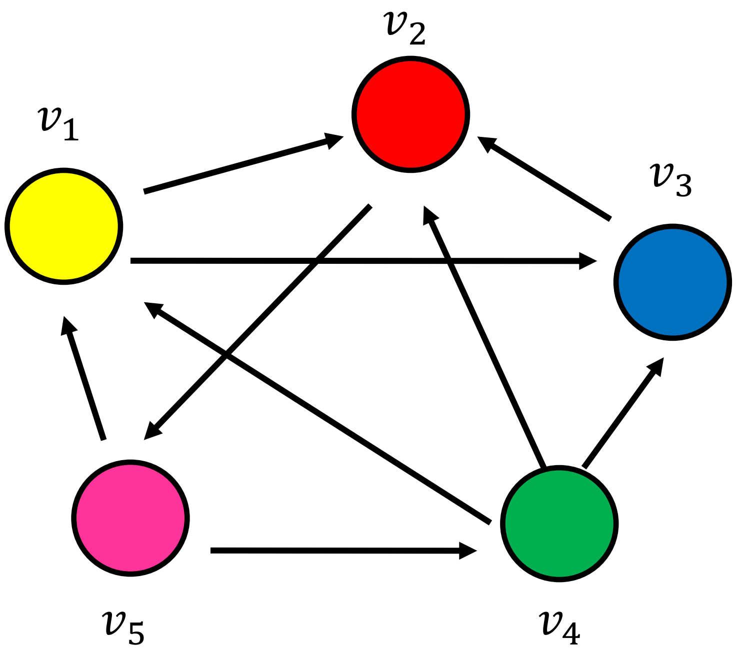
$w_2 = 1/2\ w_1 + w_3 + 1/3\ w_4$

$w_3 = 1/2\ w_1 + 1/3\ w_4$

$w_4 = 1/2\ w_5$

$w_5 = w_2$

$$w_v = \sum_{u \to v} \frac{1}{d_{out}(u)} w_u$$



Computing PageRank weights

- A simple way to compute the weights is by iteratively updating the weights.
- PageRank Algorithm

> Initialize all PageRank weights to $\frac{1}{n}$
>
> Repeat:
>
> $$w_v = \sum_{u \to v} \frac{1}{d_{out}(u)} w_u$$
>
> Until the weights do not change

- This process converges.

Example

$$w_v = \sum_{u \to v} \frac{1}{d_{out}(u)} w_u$$

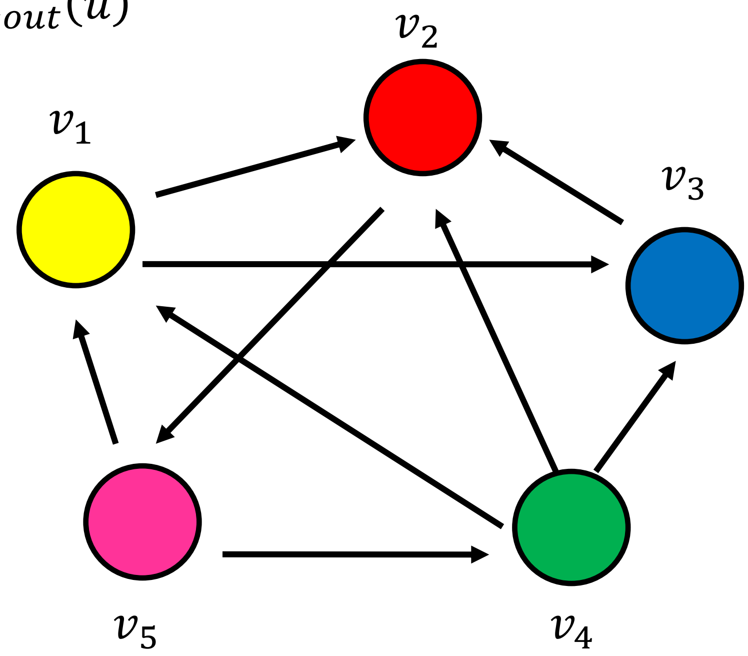
$w_1 = 1/3\ w_4 + 1/2\ w_5$

$w_2 = 1/2\ w_1 + w_3 + 1/3\ w_4$

$w_3 = 1/2\ w_1 + 1/3\ w_4$

$w_4 = 1/2\ w_5$

$w_5 = w_2$



	w_1	w_2	w_3	w_4	w_5
t=0	0.2	0.2	0.2	0.2	0.2
t=1	0.16	0.36	0.16	0.1	0.2
t=2	0.13	0.28	0.11	0.1	0.36
t=3	0.22	0.22	0.1	0.18	0.28
t=4	0.2	0.27	0.17	0.14	0.22

Think of the weight as a fluid: there is constant amount of it in the graph, but it moves around until it stabilizes.

Example

$$w_v = \sum_{u \to v} \frac{1}{d_{out}(u)} w_u$$

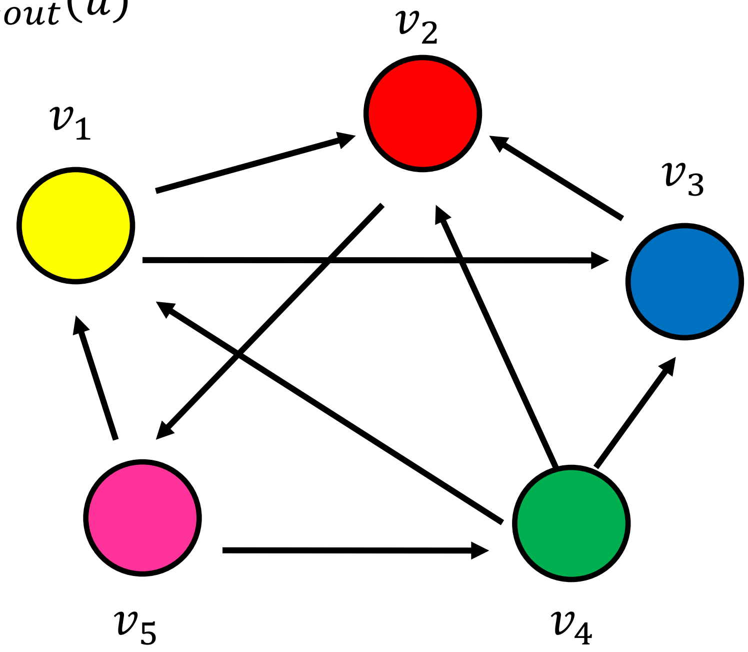
$w_1 = 1/3\ w_4 + 1/2\ w_5$

$w_2 = 1/2\ w_1 + w_3 + 1/3\ w_4$

$w_3 = 1/2\ w_1 + 1/3\ w_4$

$w_4 = 1/2\ w_5$

$w_5 = w_2$



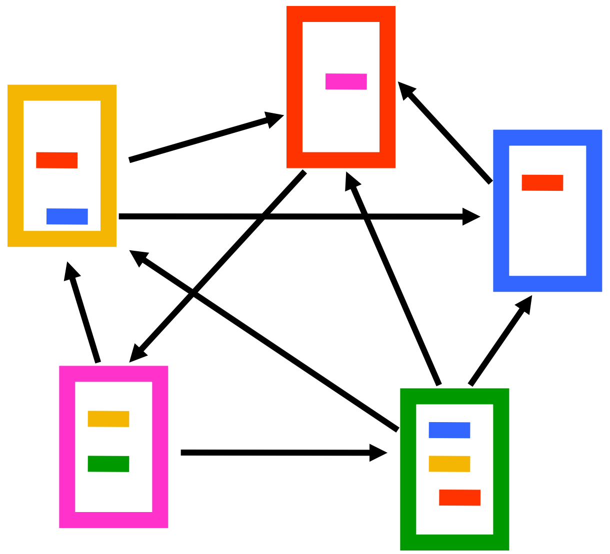
	w_1	w_2	w_3	w_4	w_5
t=25	0.18	0.27	0.13	0.13	0.27

Think of the weight as a fluid: there is constant amount of it in the graph, but it moves around until it stabilizes.

PageRank algorithm [BP98]

- The Random Surfer model
  - pick a page at random.
  - with probability 1- α jump to a random page.
  - with probability α follow a random outgoing link.
- Rank according to the stationary distribution
- $$PR(p) = \alpha \sum_{q \to p} \frac{PR(q)}{|Out(q)|} + (1-\alpha)\frac{1}{n}$$

$\alpha = 0.85$ in most cases



1. Red Page.
2. Purple Page.
3. Yellow Page.
4. Blue Page.
5. Green Page.