

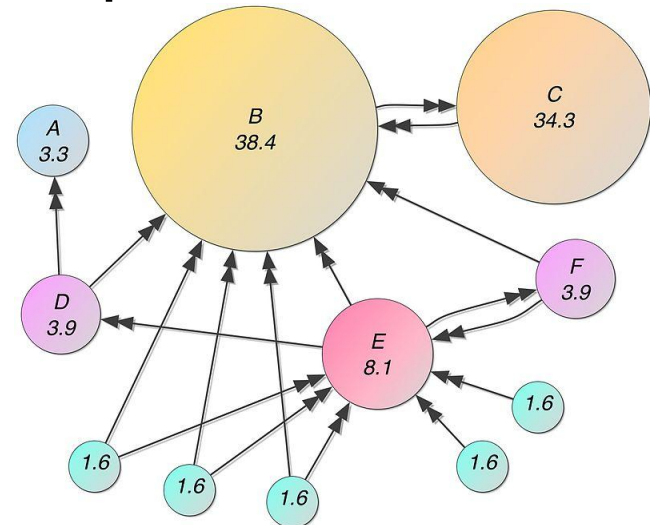
Introduction to PageRank

PageRank

- PageRank is the original famous algorithm used by the Google Search engine to rank vertexes (web pages) in a graph by order of importance
 - For the Google search engine
 - Vertexes are web pages in the World Wide Web,
 - Edges are hyperlinks among web pages
 - It assigns a numerical weighting (importance) to each node

PageRank

- It computes a likelihood that a person randomly clicking on links will arrive at any particular web page
- For a high PageRank, it is important to
 - Have many in-links
 - Be liked by relevant pages (pages characterized by a high PageRank)



PageRank

- Basic idea
 - Each link's vote is proportional to the importance of its source page p
 - If page p with importance $\text{PageRank}(p)$ has n out-links, each out-link gets $\text{PageRank}(p)/n$ votes
 - Page p 's importance is the sum of the votes on its in-links

PageRank: Simple recursive formulation

1. # Initialize each page's rank to 1.0
For each p in pages set $\text{PageRank}(p)$ to 1.0
2. Iterate for max iterations
 - a. Page p sends a contribution $\text{PageRank}(p)/\text{numOutLinks}(p)$ to its neighbors (the pages it links)
 - b. Update each page's rank $\text{PageRank}(p)$ to $\text{sum}(\text{received contributions})$
 - c. Go to step 2

PageRank with Random jumps

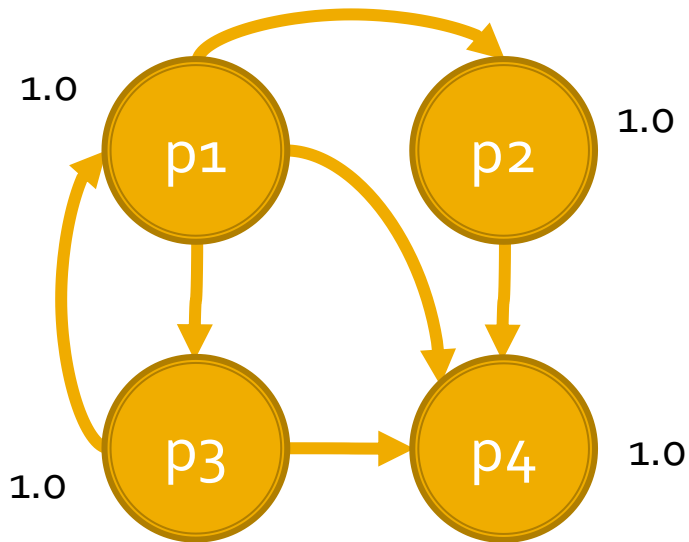
- The PageRank algorithm simulates the random walk of a user on the web
- At each step of the random walk, the random surfer has two options:
 - With probability $1-\alpha$, follow a link at random among the ones in the current page
 - With probability α , jump to a random page

PageRank with Random jumps

1. # Initialize each page's rank to 1.0
For each p in pages set $\text{PageRank}(p)$ to 1.0
2. Iterate for max iterations
 - a. Page p sends a contribution $\text{PageRank}(p)/\text{numOutLinks}(p)$ to its neighbors (the pages it links)
 - b. Update each page's rank $\text{PageRank}(p)$ to $\alpha + (1 - \alpha) * \text{sum}(\text{received contributions})$
 - c. Go to step 2

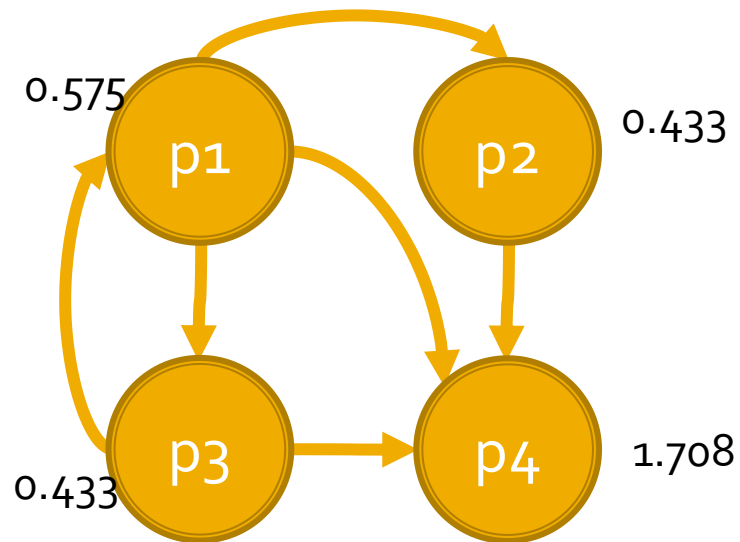
PageRank: Example

- $\alpha = 0.15$
- Initialization: $\text{PageRank}(p) = 1.0 \quad \forall p$



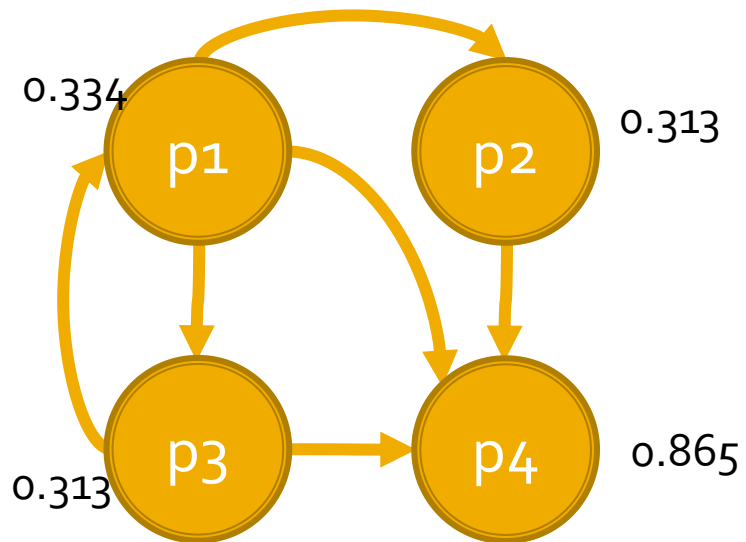
PageRank: Example

- Iteration #1



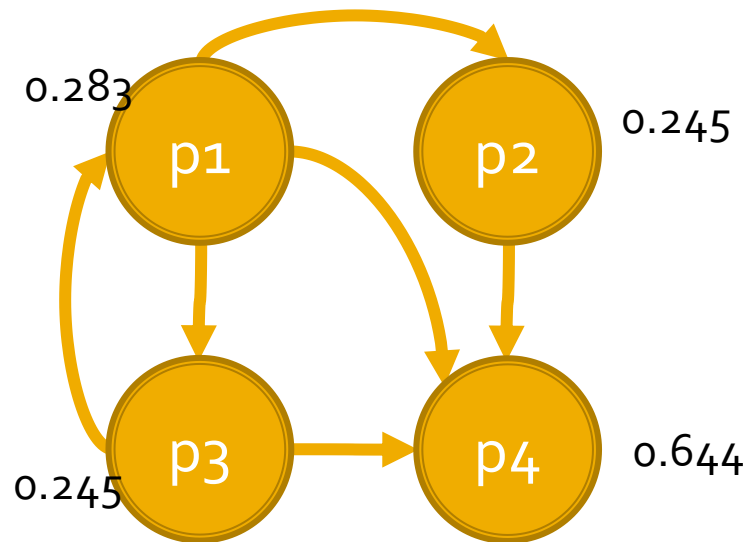
PageRank: Example

- Iteration #2



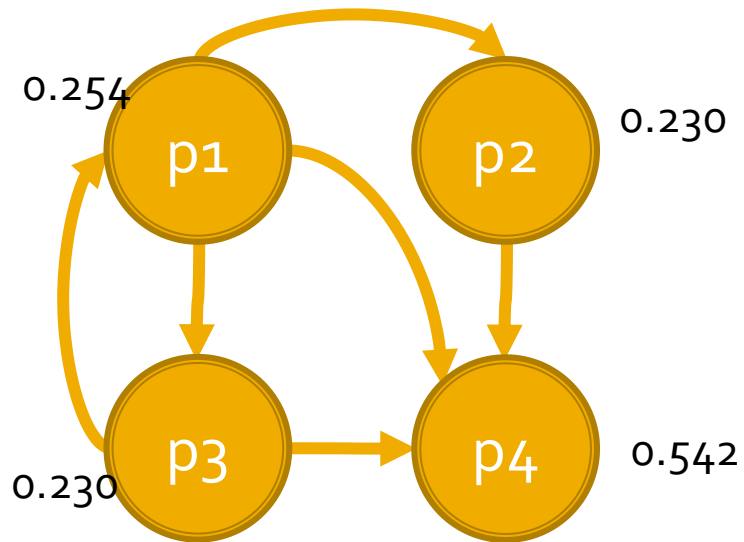
PageRank: Example

- Iteration #3



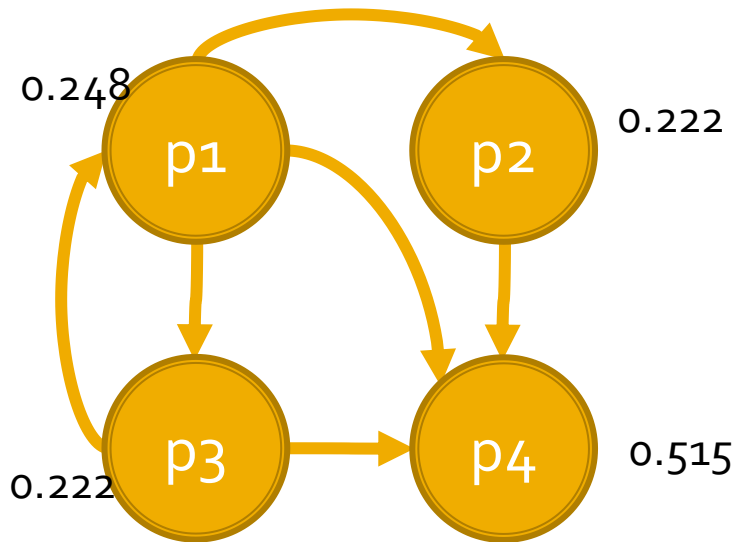
PageRank: Example

- Iteration #4



PageRank: Example

- Iteration #5



PageRank: Example

- Iteration #50

