PageRank

Ben Burns, Dan Magazu, Lucas Chagas,
Thomas Webster, Trung Do

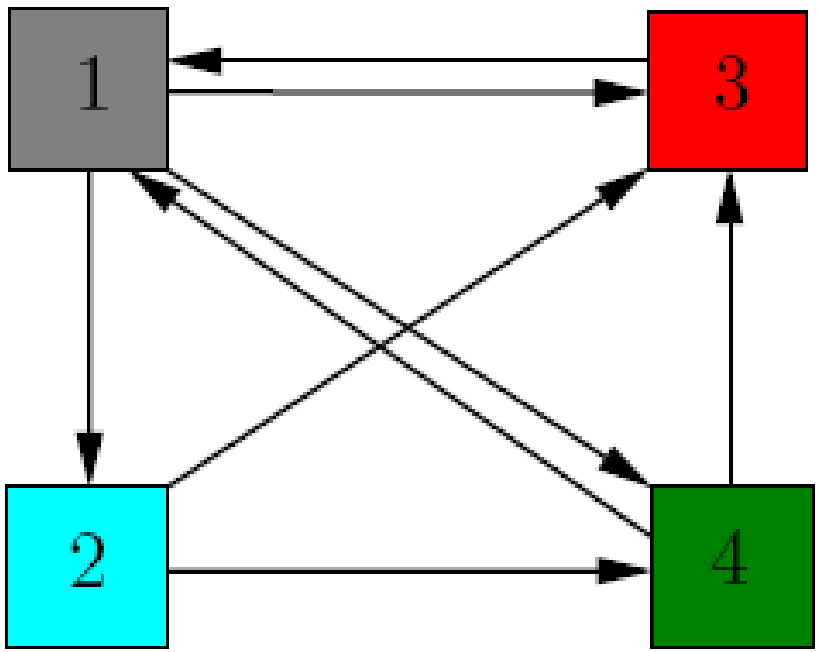
Fall 2021

Table of Contents

PageRank

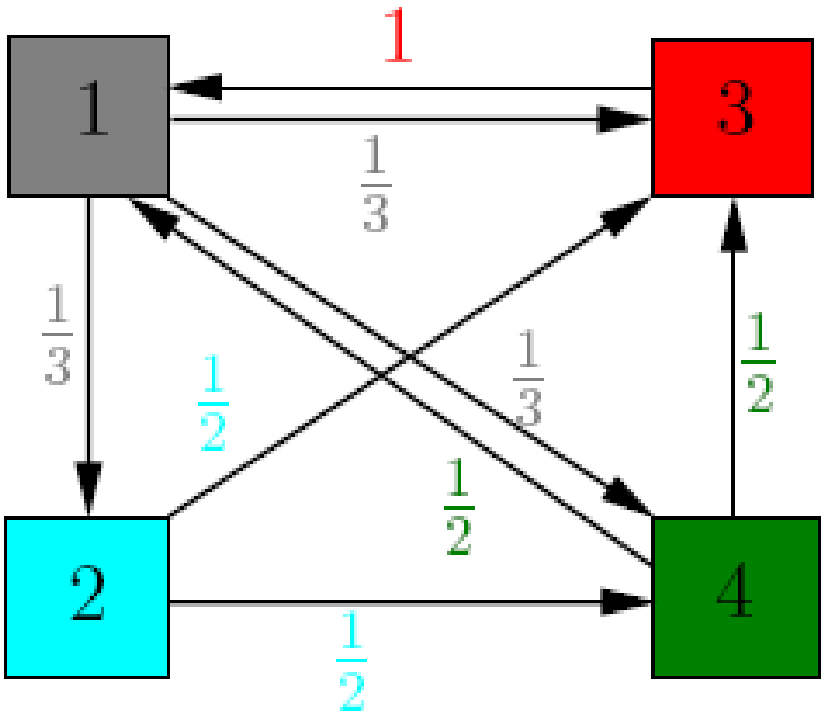
- PageRank is the algorithm to determine the importance of a website relatively to all other websites. The algorithm ranks the importance of website w, i.e., PR(w), based on the number of links points to website w and the *quality* of each pointing link from the other source website.
- **The underlying assumption**: More important website are likely to receive more links from other website. Since the algorithm measures the relative popularity ("ranking") between all websites, websites with higher ranking score are ranked higher. The sum of all ranking score equals 1 (will see why later).

Formalizing the PageRank problem



As a graph: Each website is represented by a node assigned with a PageRank value, denoted by PR(w). If a website w has a link to another website v (meaning there are an outbound link from w and an inbound link to v), then there is a directed edge from node w to node v. Multiple links from w to v is treated as a single edge from node w to v, and all self-links from a website to itself are ignored. Thus, this is a node-weighted, simple, no self-loop directed graph.

Edge Weights



This edge has its weight equal to PR(*w*) divided by the total number of outbound links of w, L(w). Then recipient node v "receives" the PageRank value of w, adding to its own value, i.e., PR(v) += PR(w)/L(w).
In another words, for a given node in the graph:
An outbound link will "give" away the PR value of the source node to the recipient node.
An inbound link will add the PR value from the source node to the recipient node.

Perspectives for Solving

- There are two ways to understand the problem:

1) as an Eigenvector problem

2) as a probability problem

- Both perspectives use linear algebra

Eigenvector Problem

Probability Problem

Power Iteration

Spider Traps and Deadends

Random Teleports + Damping factor

Applications

- PageRank is perhaps the most famous search ranking algorithm. Googles high-quality search engine results are directly correlated to PageRank
- There are a remarkable wide variety of applications of the PageRank algorithm that apply to non-search engine contexts.
- Its simplicity and elegance allow PageRank to be a more general and powerful tool.
- Let's look at the applications of PageRank and its connection to Twitter.

Twitter

- In 2010, Twitter was lagging behind and was lacking a user recommendation service.
- This was perceived both externally and internally as a critical gap in Twitter's product offerings, so quickly launching a high-quality product was a top priority.
- Twitter is unique because of the asymmetric nature of the following relationship—a user can receive messages from another without reciprocation.
- This differs substantially from other social networks such as Facebook or LinkedIn, where social ties can only be established with the consent of both participating members.

Introducing PageRank

- This works well with PageRank because we can determine outbound and inbound links.
- A user u is likely to follow those who are followed by users that are similar to u.
- These users are in turn similar to u if they follow the same (or similar) users.
- Therefore using Page Rank, Twitter is able to offer unique recommendations for users.
- By analyzing who the user follows and who those users follow, the PageRank algorithm will allow Twitter to make specific recommendations for each user.
- Essentially, the more outbound links that an account receives correlate to a higher PageRank score.