



# Pagerank algorithm

Author: Patryk Wrona

Subject: “Social Networks and  
Recommendation Systems”



# Algorithm background

A cited paper introduced a large-scale web search engine – in order to respond to increasing demands of growing number of users, **pagerank** algorithm was used.

Pagerank goals as one of the system's features:

- Improved Search Quality
- Academic Search Engine Research



The graph of the entire web consisting of links of the web is an important resource – we can make use of such a data structure thanks to calculated **pagerank**.

The described algorithm is based on:

<http://infolab.stanford.edu/pub/papers/google.pdf>

# Pagerank

This metric is widely used to “measure” the relative importance of a websites within their set of hyperlinked pages.

All webpages together constitute a network.

The described algorithm is based on:

<http://infolab.stanford.edu/pub/papers/google.pdf>

# Algorithm description

Page A is associated with  $T_1, \dots, T_n$  pages linking to page A. Indeed, their relationship could be associated with digraph data structure.

$C(A)$  is defined as the number of links going out of A – outgoing vertices.

$d$  – an arbitrary parameter from the interval  $[0,1]$  – usually 0.85

$$PR(A) = (1-d) + d (PR(T_1)/C(T_1) + \dots + PR(T_n)/C(T_n))$$

In order to calculate given page (e.g A) pagerank (PR), one must calculate all pageranks associated with the pages pointing at A – that is why it is a **relative** measure.

The described algorithm is based on:

<http://infolab.stanford.edu/pub/papers/google.pdf>

# Algorithm features

- All pageranks of  $A$ ,  $T_1, \dots, T_n$  form a probability distribution.
- Pagerank of  $A$  is the eigenvector corresponding to the largest eigenvalue of the standardized matrix of the web.
- Pagerank of  $A$  can be calculated iteratively as it consists of calculating the respective websites' pageranks until the convergence of the algorithm.





Thank You  
for your attention

