

# PageRank

---

Antonio Luiz Rosa Teixeira   Gustavo Zambonin

Universidade Federal de Santa Catarina  
Departamento de Informática e Estatística  
INE5413 — Grafos

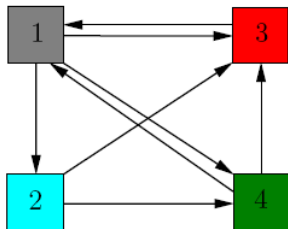
# Introdução

- ▶ Desenvolvido por Page e Brin em 1996
  - ▶ Parte do motor de busca da Google desde 1998
- ▶ Classificação de uma página na web, independente do seu conteúdo
- ▶ Entretanto, depende do número de links que apontam para ela
- ▶ Abrange diversas áreas da matemática
  - ▶ Teoria de grafos
  - ▶ Probabilidade e estatística
  - ▶ Álgebra linear
- ▶ Definições genéricas e aplicáveis a qualquer grafo

- ▶ Motores de busca prévios utilizavam heurísticas facilmente manipuláveis
  - ▶ Tamanho do URL
  - ▶ Conteúdo bruto da página
  - ▶ Título da página
- ▶ Método mais eficiente de relevar páginas de acordo com as palavras-chave da pesquisa
- ▶ Evitar meios artificiais de inflação de popularidade

# Modelagem

- ▶ Possível representar relações entre páginas como um grafo direcionado
- ▶ Seja um grafo  $G(V, A)$ , então:  
 $V = \{p \mid p \text{ é uma página da web}\}$   
 $A = \{(r_1, r_2) \mid r_1, r_2 \in V \text{ e } r_2 \text{ referencia } r_1 \text{ por um } \textit{hyperlink}\}$
- ▶ Assume-se um 'navegador' de páginas, que clicará em links aleatoriamente, até que isto seja impossível



Supondo uma rede com apenas quatro páginas e referências entre estas, tem-se o grafo acima.

# Resolução

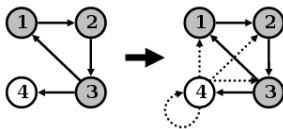
- ▶ Cada página deve transferir sua 'importância'  $\frac{1}{k}$  para seus sucessores
  - ▶  $k$  = grau de emissão do vértice
  - ▶ Processo de valoração de arestas
- ▶ Tem-se uma matriz de transições do grafo
- ▶ Para o exemplo anterior (cada coluna representa uma página):

$$\begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix} \xrightarrow{Av=\lambda v} 1 \cdot \begin{bmatrix} 0.7210 \\ 0.2403 \\ 0.5408 \\ 0.3605 \end{bmatrix}$$

- ▶ O autovetor denota a probabilidade das páginas serem visitadas

# Problemas

- ▶ Se  $k = 0$  para uma página qualquer, então a página não afetaria outras, no modelo acima



A página '4', um sumidouro, deve possibilitar a navegação para o resto do grafo

- ▶ Se o grafo é desconexo, como mover entre componentes?
  - ▶ Considerar que o 'navegador' aleatório pode se cansar de clicar e querer trocar de página de outro modo
- ▶ Adicionar um fator probabilístico que lida com estes cenários

- ▶ Métrica para profundidade e extensão de *web crawling*
- ▶ Funções mais importantes no kernel Linux
- ▶ Recomendação de seguidores no Twitter, produtos na Amazon e filmes no Netflix
- ▶ Impacto científico de pesquisadores e artigos
- ▶ Sistema de ranking entre times ou atletas em esportes
- ▶ Encontrar genes correlacionados de acordo com um certo critério
- ▶ Predição de fluxo de tráfego e pessoas em um sistema urbano
- ▶ Ranking de isomorfismos de grafos

# Referências



Gleich, D. F. (2015).  
PageRank Beyond the Web.  
*SIAM Review*, 57(3):321–363.



Page, L., Brin, S., Motwani, R., and Winograd, T. (1999).  
The PageRank Citation Ranking: Bringing Order to the Web.  
Technical Report 1999-66, Stanford InfoLab.



Tanase, R. and Radu, R. (2009).  
PageRank Algorithm - The Mathematics of Google Search.  
Lecture 3.



Wills, R. S. (2006).  
Google's PageRank: The Math Behind the Search Engine.  
*The Mathematical Intelligencer*, 28(4):6–11.