

A Measure of Similarity between Graph Vertices: Applications to Synonym Extraction and Web Searching

Vincent D. Blondel, Anahí Gajardo, Maureen Heymans, Pierre Senellart,
and Paul Van Dooren,
SIAM (Society for Industrial and Applied Mathematics)
Review, 46(4) (2004)

Apresentador: Eduardo Hargreaves
01/04/2014

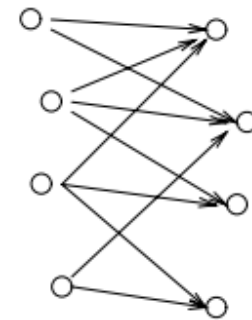
Hub-Authority

- ▶ Conceito desenvolvido por Kleinberg no algoritmo HITS (Hyperlink-induced topic search ou Authoritative Sources)
- ▶ Utiliza os papéis de Hubs e Autoridades para criar um algoritmo de respostas a buscas feitas na Internet
- ▶ Uma página é um bom *hub* se aponta para páginas que são boas autoridades
- ▶ Uma página é uma boa autoridade se é apontada por bons *hubs*
- ▶ É uma medida de centralidade

Hub-Authority

- ▶ Cálculos das Pontuações

$$\begin{cases} h_j \leftarrow \sum_{i:(j,i) \in E} a_i, \\ a_j \leftarrow \sum_{i:(i,j) \in E} h_i. \end{cases}$$



hubs

authorities

- ▶ Relação de Reforço Mútuo - Recursividade
- ▶ Este sistema de equações pode ser reescrito

como:

$$\begin{bmatrix} h \\ a \end{bmatrix}_{k+1} = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \begin{bmatrix} h \\ a \end{bmatrix}_k, \quad k = 0, 1, \dots,$$
$$x_{k+1} = M x_k, \quad k = 0, 1, \dots,$$

Equações de Reforço Mútuo Generalizadas para qualquer grafo

$$x_{ij} \leftarrow \sum_{r:(r,i) \in E_B, s:(s,j) \in E_A} x_{rs} + \sum_{r:(i,r) \in E_B, s:(j,s) \in E_A} x_{rs}.$$

$$X_{k+1} = BX_kA^T + B^T X_kA, \quad k = 0, 1, \dots,$$

THEOREM 3. Let G_A and G_B be two graphs with adjacency matrices A and B , fix some initial positive matrix $Z_0 > 0$, and define

$$Z_{k+1} = \frac{BZ_kA^T + B^T Z_kA}{\|BZ_kA^T + B^T Z_kA\|_F}, \quad k = 0, 1, \dots$$

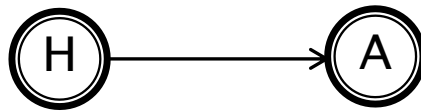
Then the matrix subsequences Z_{2k} and Z_{2k+1} converge to Z_{even} and Z_{odd} . Moreover, among all the matrices in the set

$$\{Z_{\text{even}}(Z_0), Z_{\text{odd}}(Z_0) : Z_0 > 0\},$$

the matrix $Z_{\text{even}}(\mathbf{1})$ is the unique matrix of the largest 1-norm.

Equivalência Regular

- ▶ De acordo com o conceito de equivalência regular, a matriz transposta de $\begin{bmatrix} h \\ a \end{bmatrix}$ pode ser interpretado como a equivalência regular entre um grafo G_B qualquer e o grafo orientado G_A (chamado de grafo estruturante) dado por:



- ▶ Esta equivalência pode ser interpretada como a similaridade entre um vértice de G_B e um papel de hub (ou autoridade)

Algoritmo

1. Set $Z_0 = \mathbf{1}$.
2. Iterate an even number of times

$$Z_{k+1} = \frac{BZ_k A^T + B^T Z_k A}{\|BZ_k A^T + B^T Z_k A\|_F}$$

and stop upon convergence.

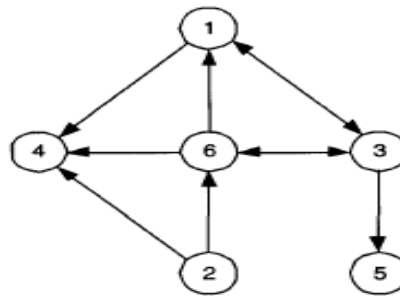
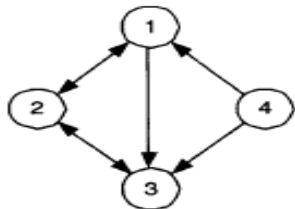
3. Output \mathbf{S} is the last value of Z_k .

A complexidade é dada por:

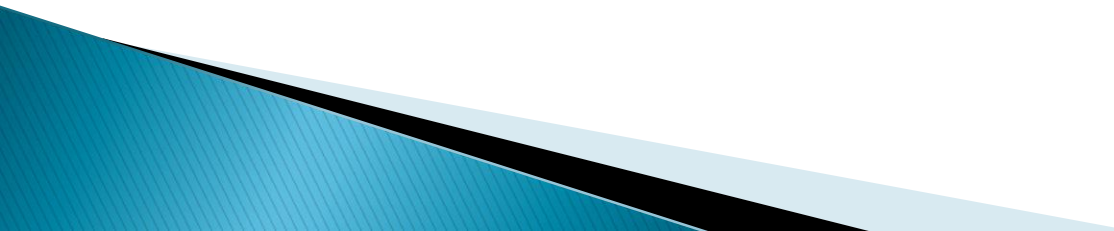
$$8n_A n_B \frac{(\alpha_A + \alpha_B) \log \epsilon}{(\log \mu - \log \rho)}$$

Aplicações

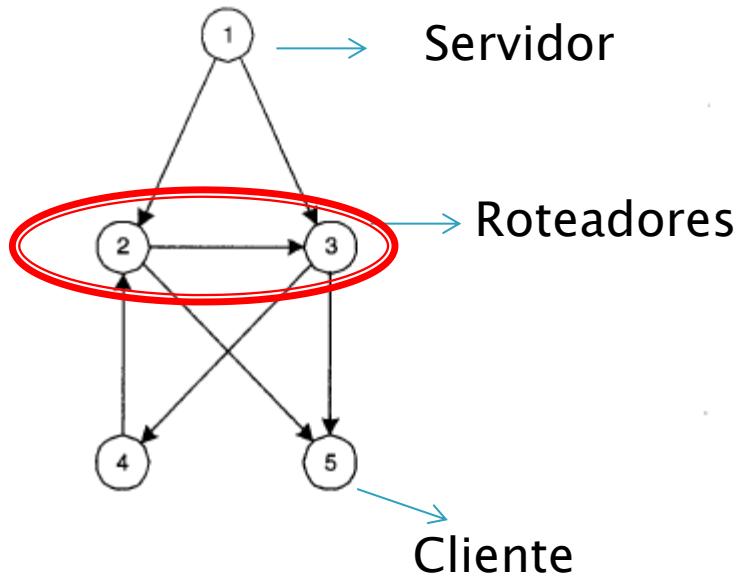
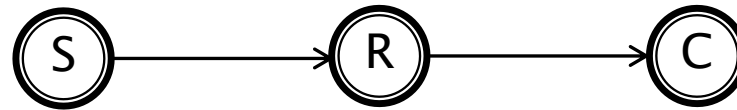
- ▶ Existem duas forma de interpretar estas equações
- ▶ A primeira é encontrar a equivalência direta entre dois grafos quaisquer – isomorfismo?



0.2636	0.2786	0.2723	0.1289
0.1286	0.1286	0.0624	0.1268
0.2904	0.3115	0.2825	0.1667
0.1540	0.1701	0.2462	0
0.0634	0.0759	0.1018	0
0.3038	0.3011	0.2532	0.1999

- ▶ A segunda, que ao meu ver é a mais poderosa, usa o conceito de equivalência estrutura (apesar desta equivalência não ser citada no artigo).
 - ▶ Os papéis e relacionamentos são modelados em um grafo estruturante e a similaridade entre a estrutura investigada e o grafo estruturante é calculada
- 

Exemplo: Rede de computadores



0.4433	0.1043	0
0.2801	0.3956	0.0858
0.0858	0.3956	0.2801
0.2216	0.0489	0.2216
0	0.1043	0.4433

Aplicação – Extração de Sinônimos

- ▶ Sinônimos possuem palavras em comum nas suas definições, e aparecem juntos da definição de muitas palavras

- ▶ Grafo estruturante



- ▶ Primeiro é construído um grafo G com todas as palavras do dicionário. Se uma palavra aparece na definição de outra, é criada uma aresta
- ▶ Para cada palavra w procurada, é construído um grafo com os vizinhos de w .
- ▶ Sinônimos são ordenados pela similaridade com 2