

## 3 Google PageRank

### 3.1 基本信息

- 1998 年, Page, Brin
- 优点
  - 完全由“网络结构”决定
  - 无广告
  - 速度快
  - 检索结果比较全面

### 3.2 PageRank

**Definition 1** (PageRank). 基于“从许多优秀的网页链接过来的网页必是优秀网页”

回归的定义

**Remark:**

- 网页链接结构定义
- $A \rightarrow B$ , 投票
- 权重

### 3.3 图论

**Definition 2.**  $W$  是所有网页的集合,  $N = |W| =$  网页的个数

**Definition 3** (关联矩阵). 网页  $j$  链向网页  $i$ , 则  $G_{ij} = 1$ , 否则  $G_{ij} = 0$ .

**Remark:**

- ①  $N$  大,  $G$  高度稀疏
- ② 方阵, NonSymmetric
- ③ 第  $i$  行: 很多非零元素,  $\rightarrow i$ , 向内链接  
第  $j$  列: 很多非零元素,  $j \rightarrow$ , 向外链接

**Remark:**

$\Rightarrow$  PR 依赖于进向链接数

$\Rightarrow$  PR 依赖于进向链接中的优质网页

$\Rightarrow$  PR 依赖于进向链接源向外链接数

**Definition 4** (PR 计算). 每条向外链接的值是该链接源的 PR 值除以链接源向外链接数, 网页 PR 是所有进向链接值总和

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \div 5 & \div 1 & \div 2 & \div 3 & \div 4 & \div 2 & \div 1 \end{bmatrix}$$

$$x_1 = \frac{1}{1}x_2 + \frac{1}{2}x_3 + \frac{1}{4}x_5 + \frac{1}{2}x_6$$

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix}$$

$$A\vec{x} = \vec{x}$$