Search Engine

网页重要度

- 由互联网中网页之间的链接关系决定
- 某网页重要，因为有重要的网页链接到它
- 用graph来表示
- 叠加性 传递性 等效性(贡献 $\frac{x_i}{q_i}$) 无关性

重要度，记（q_i为链接网页的数量）

$p_{ij}=\begin{cases}\frac{1}{q_j} & v_j\text{指向}v_i\\ 0 & \text{无链接}\end{cases}$

则有

$x_i=\frac{x_{j_1}}{q_{j_1}}+\ldots+\frac{x_{j_k}}{q_{j_k}}=p_{ij_1}x_{j_1}+\ldots+p_{ij_k}x_{j_k}$

即

$x_i=\sum_{j=1}^{n}p_{ij}x_j, i=1,\ldots,n$

记$X=(x_1,x_2,\ldots,x_n)^T$为重要度向量，则X为线性方程组X=PX的解（*此时问题转化为求链接矩阵关于特征值1的特征向量*）

随机矩阵

- 行（列）元素之和为1的非负方阵称为行（列）随机矩阵
- 任一随机矩阵均有特征值1（*故链接矩阵有特征值1*）
- 任一随机矩阵的模最大特征值为1（考虑了实数复数）
- 所有元素为正的（列）随机矩阵仅有1个属于特征值1的特征向量（证唯一性）
  - 引理：设$X=(x_1,x_2,\ldots,x_n)^T$为Google矩阵属于1的特征向量，则其元素要么恒非负，要么恒非正
- 由于链接矩阵为一随机矩阵，故有性质$\sum_{i=1}^{n}p_{ij}=1$

悬挂网页

- 不链接其他任何网页，网络链接图中对应顶点的出度全为0，链接矩阵中对应列元素全为0(破坏了链接矩阵成为随机矩阵的条件)
- 将该列所有元素修改为$\frac{1}{n}$，链接矩阵则成为随机矩阵
- 再增加一个向量d标记哪几列元素全为0
- 修改为$P=P_0+\frac{1}{n}de^T$(P_0为原始的稀疏矩阵)

唯一性（特征值1对应的特征子空间为一维）

- 修改链接矩阵为Google矩阵：$\bar{P}=\alpha P+(1-\alpha)\frac{1}{n}ee^T$，其中$\alpha=0.85$， P为悬挂网页修改矩阵
- α是通过链接打开网页（链接矩阵）与输入网址新建窗口（悬挂网页）的比例，约为5:1
- 所有元素非负，且列元素和为1
- 由*Perron-Frobenius定理*可以得到，属于特征值1的特征向量唯一，即重要度排序结果唯一

Perron-Frobenius定理

- 若矩阵A的所有元素为正，则
  - A的模最大特征值唯一，且为正实数
  - 该特征值的代数重数为1
  - 存在该特征值的一个特征向量，其分量全为正
- *Google矩阵为元素全正的随机矩阵，1为其最大特征值，重要度向量唯一且分量全为正*
- 若矩阵A为非负不可约矩阵，则
  - A的模最大特征值为正实数
  - 该特征值代数重数为1
  - 存在该特征值的一个特征向量，其分量全为正
- 不可约矩阵
  - 若对应有向图为强连通，则A是不可约矩阵
  - 若不存在初等置换矩阵Q，使得$Q^TAQ=\begin{pmatrix} X & 0 \\ Y & Z \end{pmatrix}$

幂法

- 任取初始向量$x^{(0)}$, $x^{(0)}>0$且$x^{(0)^T}e=1$
- 计算$x^{(k)}=\bar{P}x^{(k-1)}$
- $\lim\limits_{k\to\infty} x^{(k)}$存在，极限即为重要度向量X
- 幂法对稀疏矩阵运算量较小，且上式可简化为$\bar{P}x^{(k)}=\alpha P_0 x^{(k)}+(\frac{\alpha}{n}d+\frac{1-\alpha}{n}e)$

随机浏览Random Surfer

- Markov链：设$\{X_m, m=1,2,3,\ldots\}$为一个随机过程，状态空间有限，若$P\{X_m=i\}$只与X_{m-1}有关，而与X_{m-2}, X_{m-3}等无关，则称$\{X_m\}$为Markov链
- 记$P\{X_m=i|X_{m-1}=j\}=p_{ij}(m)$，若$p_{ij}(m)$与$m$无关，则称Markov链为**齐次**的。$P=p_{ij}(m)$为一随机矩阵，称为$\{X_m\}$的**转移矩阵**。
- 当Markov链为齐次，$P=(p_{ij})_{n\times n}$为一个随机矩阵，称为$\{X_m\}$的转移矩阵
- 若 $P\{X_{m-1}=j\}=x_j \Rightarrow P\{X_m=i\}=\sum\limits_{j=1}^{n} P\{X_m=i|X_{m-1}=j\}P\{X_{m-1}=j\}=\sum\limits_{j=1}^{n} p_{ij}x_j$，记$x^{(m)}=(P\{X_m=1\}, P\{X_m=2\},\ldots,P\{X_m=n\})^T$可得$x^{(m)}=Px^{(m-1)}$，与幂法形式相同（意义：无论从何网页开始随机浏览，经过充分时间，停留在各网页上的概率组成的向量即为重要度向量）

修正后的问题：求解
$X=\bar{P}X$