数学建模

浙江大学数学系　谈之奕

tanzy@zju.edu.cn

Google中的数学模型

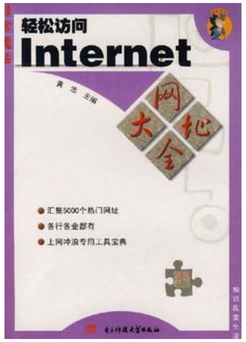
搜索引擎

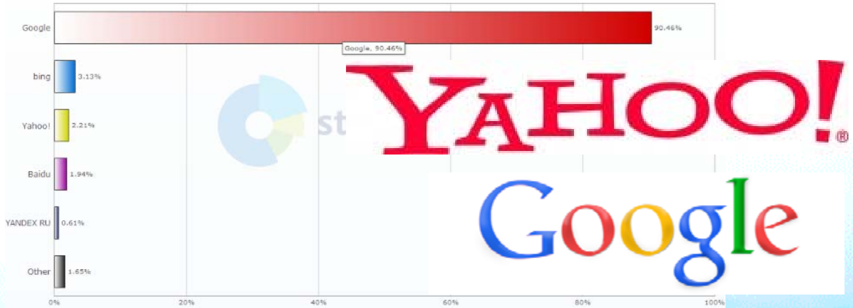
浙江大学 ZheJiang University

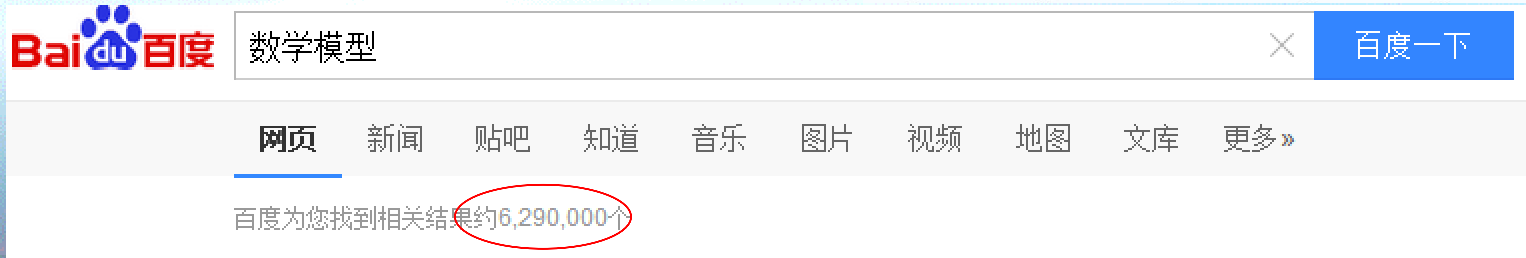
数学建模

- 搜索引擎（search engine）：一种用于帮助用户查找因特网信息的技术工具。其工作原理是先以一定的策略在因特网中搜索，发现信息，然后对信息进行理解、提取。组织和处理，最后为用户提供信息和服务

——摘自《中国大百科全书》







Google与PageRank

浙江大学 ZheJiang University

数学建模

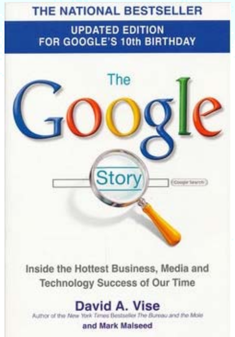
Brin和Page进入斯坦福大学攻读博士学位

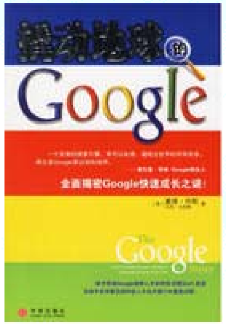
设计Google搜索引擎并创立Google公司

Google公司在Nasdaq上市

Google公司市值达到5400亿美元

1995 1996 1998 2004 2017





Vise DA, *The Google Story*

Page提出了Google搜索引擎的核心——衡量网页重要度的模型和算法PageRank

Googol 10^{100}

Google!

Google创始人：
Larry Page(左)
Sergey Brin(右)

网页重要度

浙江大学
ZheJiang University
数学建模

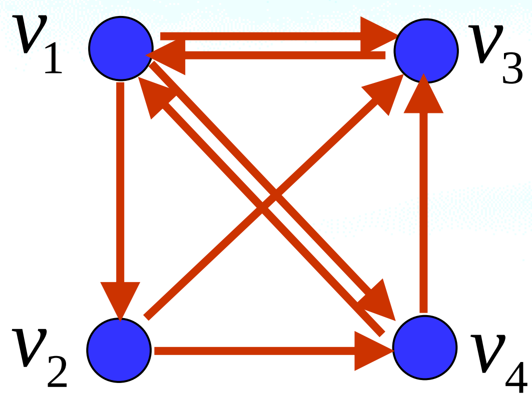
- 源起
  - 互联网上存在海量信息，如何寻找
  - 有相同关键词的网页数量众多，如何展示
  - 互联网变化快，缺乏统一管理，如何组织
- 要求
  - 合理、客观、定量、可操作
- 原则
  - 网页重要度由互联网中网页之间链接关系决定
  - 某网页重要，因为有重要的网页链接到它

规模大
动态
自组织

网络链接图

数学建模

- 由**Internet**上网页链接关系构造网络链接图 $G=(V,A)$
  - 顶点（网页）与顶点集：$V=\{v_1,v_2,\cdots,v_n\}$
  - 弧（链接）(v_i,v_j)：从网页 v_i有链接指向网页 v_j，A为弧的集合
  - 出度 q_i：以 v_i为起点的弧的总数（网页 v_i 上的链接数目）
- G 为一有向图（**digraph**）



网络链接图

网页重要度

数学建模

- 网页重要度由**Internet**中网页之间的链接关系决定

  若有重要的网页链接到某网页**A**，则网页**A**也是重要的

  - （叠加性）链接到网页**A**的网页越多，则网页**A**越重要
  - （传递性）重要度大的网页链接到网页**A**时对**A**重要度的贡献比重要度小的网页链接到**A**时对**A**重要度的贡献更大

    对其它网页的贡献与自身的重要度成正比
  - （等效性）网页链接较多时对它所链接的网页的重要度的贡献比链接较少时对它所链接的网页的重要度的贡献小

    任一网页对其它网页重要度贡献之和与链接的网页数量无关
  - （无关性）网页链接其它网页的多少，与其本身的重要度无关

浙江大学
ZheJiang University
数学建模

重要度计算

- 记 x_i 为网页 v_i 的重要度
  - 网页 v_i 对其它网页重要度贡献之和为 x_i 传递性
  - 网页 v_i 对它链接的 q_i 个网页中的任一个的重要度贡献为 $\dfrac{x_i}{q_i}$ 等效性
  - 若链接到网页 v_i 的网页有 $v_{j_1}, v_{j_2}, \cdots, v_{j_k}$ ，则

$$x_i = \frac{x_{j_1}}{q_{j_1}} + \frac{x_{j_2}}{q_{j_2}} + \cdots + \frac{x_{j_k}}{q_{j_k}}$$

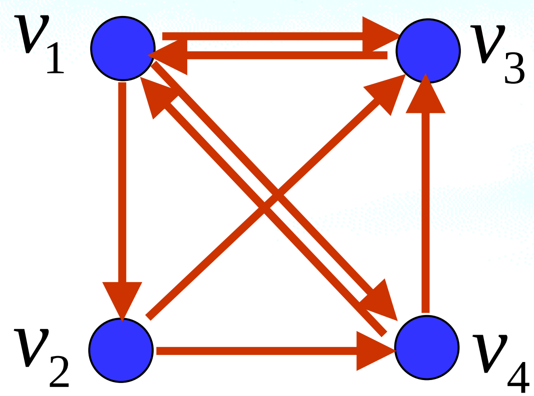
叠加性

无关性

链接矩阵

数学建模

- 令 $p_{ij}=\begin{cases}\dfrac{1}{q_j}, & \text{若有链接自}v_j\text{链向}v_i\\ 0, & \text{其他}\end{cases}\quad i,j=1,\cdots,n$
- $\mathbf{P}=(p_{ij})_{n\times n}$ 为一 n 阶方阵，称为链接矩阵，n 为网页数目
- 记 $\mathbf{X}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 为网页重要度向量
- $\mathbf{X}$ 为线性方程组 $\mathbf{X}=\mathbf{P}\mathbf{X}$ 的解



$$\mathbf{P}=\begin{pmatrix} 0 & \boxed{0} & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix}$$

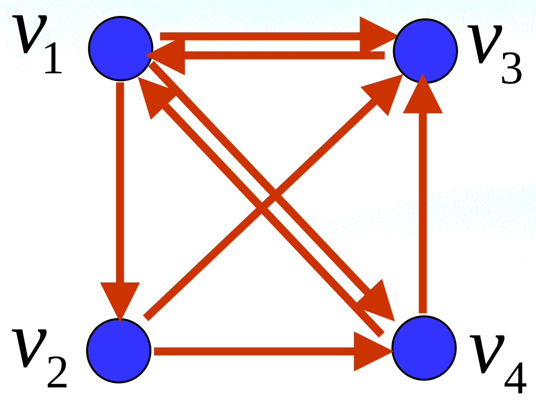
$$x_i=\frac{x_{j_1}}{q_{j_1}}+\frac{x_{j_2}}{q_{j_2}}+\cdots+\frac{x_{j_k}}{q_{j_k}}=p_{ij_1}x_{j_1}+\cdots+p_{ij_k}x_{j_k}$$

$$=p_{i1}x_1+p_{i2}x_2\cdots+p_{in}x_n=\sum_{j=1}^{n}p_{ij}x_j,\ i=1,\cdots,n$$

重要度向量

数学建模

$$\mathbf{X}=\mathbf{P}\mathbf{X} \qquad (\mathbf{I}-\mathbf{P})\mathbf{X}=\mathbf{0}$$



$$\begin{cases} x_1 = x_3+\frac{1}{2}x_4 \\ x_2=\frac{1}{3}x_1 \\ x_3=\frac{1}{3}x_1+\frac{1}{2}x_2+\frac{1}{2}x_4 \\ x_4=\frac{1}{3}x_1+\frac{1}{2}x_2 \end{cases} \Leftrightarrow \begin{pmatrix} 1 & 0 & -1 & -\frac{1}{2} \\ -\frac{1}{3} & 1 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{3} & -\frac{1}{2} & 0 & 1 \end{pmatrix}$$

$$\mathbf{P}=\begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix}$$

$$\left(\frac{12}{31},\frac{4}{31},\frac{9}{31},\frac{6}{31}\right)^{\mathrm{T}} \Leftarrow \begin{pmatrix} 1 & 0 & -1 & -\frac{1}{2} \\ 0 & 1 & -\frac{1}{3} & -\frac{1}{6} \\ 0 & 0 & \frac{1}{2} & -\frac{3}{4} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$v_1 \succ v_3 \succ v_4 \succ v_2$$

非零解存在性

- 对任意矩阵 $\mathbf{P}$，齐次线性方程组 $\mathbf{X}=\mathbf{P}\mathbf{X}$ 有非零解
  - $\mathrm{rank}(\mathbf{I}-\mathbf{P})<n$
  - 矩阵 $\mathbf{P}$ 有特征值 1
  - 矩阵 $\mathbf{P}^{\mathrm{T}}$ 有特征值 1

$$(\mathbf{I}-\mathbf{P})\mathbf{X}=\mathbf{0} \qquad \mathbf{P}\mathbf{X}=1\cdot\mathbf{X}$$

$$|\lambda\mathbf{I}-\mathbf{P}|=|\lambda\mathbf{I}-\mathbf{P}^{\mathrm{T}}|$$

$$\mathbf{P}^{\mathrm{T}}\mathbf{e}=1\,\mathbf{e},\ \mathbf{e}=(1,\cdots,1)^{\mathrm{T}}$$

- 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（**stochastic matrix**）
  - 任一随机矩阵均有特征值 1

$$\mathbf{P}=\begin{pmatrix} 0 & \boxed{0} & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix} \qquad \mathbf{P}^{\mathrm{T}}=\begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}$$

随机矩阵

- 任一随机矩阵的模最大特征值为 1
  - 设 λ 是（行）随机矩阵 $\mathbf{P}=(p_{ij})_{n\times n}$ 的特征值，非零向量 $\mathbf{X}$ $(x_1,\cdots,x_n)^{\mathrm{T}}$ 为属于特征值 的特征向量
  - $|x_i|=\max\limits_{1\le j\le n}|x_j|>0$
  - $\lambda x_i=\sum\limits_{j=1}^{n}p_{ij}x_j$
  - $$|\lambda||x_i|=\left|\lambda x_i\right|=\left|\sum_{j=1}^{n}p_{ij}x_j\right|\le\sum_{j=1}^{n}|p_{ij}||x_j|\le|x_i|\sum_{j=1}^{n}|p_{ij}|=|x_i|$$

实矩阵的特征值可能是复数

$$\mathbf{PX}=\lambda\mathbf{X}$$

悬挂网页

- 某网页不链接任意其它网页
  - 网络链接图中对应顶点的出度为 0
  - 链接矩阵中对应列元素全为 0
- 将该列所有元素修改为 $\frac{1}{n}$，链接矩阵成为随机矩阵
- 链接矩阵为一随机矩阵，该矩阵属于特征值 1 的特征向量即为重要度向量

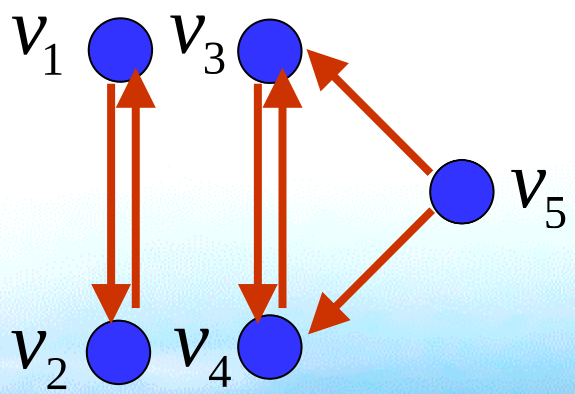
$$\mathbf{P}_0=\begin{pmatrix} 0 & 0 & 1 & 0 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 \end{pmatrix} \quad \mathbf{P}=\mathbf{P}_0+\begin{pmatrix} 0 & 1/4 & 0 & 1/4 \\ 0 & 1/4 & 0 & 1/4 \\ 0 & 1/4 & 0 & 1/4 \\ 0 & 1/4 & 0 & 1/4 \end{pmatrix}=\mathbf{P}_0+\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 0 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

$$=\mathbf{P}_0+\frac{1}{4}\mathbf{e}\mathbf{d}^{\mathrm{T}} \qquad \mathbf{d}=\begin{pmatrix} 0 & 1 & 0 & 1 \end{pmatrix}^{\mathrm{T}}$$

唯一性

- 若 **P** 有两个属于特征值1的线性无关的特征向量（属于特征值 的特征子空间维数大于 ）1，用上述方法可能得到相互矛盾的网页重要度比较结果

$$\mathbf{P}=\begin{pmatrix}0&1&0&0&0\\1&0&0&0&0\\0&0&0&1&1/2\\0&0&1&0&1/2\\0&0&0&0&0\end{pmatrix}\qquad\begin{pmatrix}1\\1\\0\\0\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\\1\\0\end{pmatrix}$$



Google矩阵

- 修改链接矩阵为 $\overline{\mathbf{P}} = \alpha\mathbf{P} + (1-\alpha)\dfrac{1}{n}\mathbf{e}\mathbf{e}^{\mathrm{T}}$，其中参数 $\alpha = 0.85$
- $\overline{\mathbf{P}}$ 所有元素均为正，每列元素之和仍为 1

$$\overline{\mathbf{P}} = 0.85\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.5 \\ 0 & 0 & 1 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} + 0.15\cdot\frac{1}{5}\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0.03 & 0.88 & 0.03 & 0.03 & 0.03 \\ 0.88 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.88 & 0.455 \\ 0.03 & 0.03 & 0.88 & 0.03 & 0.455 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \end{pmatrix}$$

唯一性的初等证明

浙江大学 ZheJiang University

数学建模

- 所有元素为正的（列）随机矩阵仅有 1个属于特征值 1 的线性无关的特征向量
  - 引理：设 $\mathbf{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$为 $\mathbf{P}$的属于特征值 1的非零特征向量，则 $\sum_{i=1}^{n} x_i \neq 0$

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

    - 反证法 若 $\sum_{i=1}^{n} x_i = 0$，则 $\mathbf{x}$ 的分量有正有负
    - $|x_i| = \left|\sum_{j=1}^{n} p_{ij} x_j\right| < \sum_{j=1}^{n} p_{ij} |x_j| \qquad i=1,\cdots,n$
    - $\sum_{i=1}^{n} |x_i| < \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} |x_j| = \sum_{j=1}^{n} |x_j| \left(\sum_{i=1}^{n} p_{ij}\right) = \sum_{j=1}^{n} |x_j|$ 矛盾

唯一性的初等证明

浙江大学 ZheJiang University

数学建模

- 设 $\mathbf{v}=(v_1,\cdots,v_n)^{\mathrm{T}}$ 和 $\mathbf{w}=(w_1,\cdots,w_n)^{\mathrm{T}}$ 是 $\mathbf{P}$ 的两个属于特征值 1的线性无关的特征向量
- 记 $W=\sum_{k=1}^{n} w_k$，$V=\sum_{k=1}^{n} v_k$， 令 $x_i=-\frac{W}{V}v_i+w_i, i=1,\cdots,n$
  - $\mathbf{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 为 $\mathbf{P}$ 的属于特征值 1的非零特征向量
  - $\sum_{j=1}^{n} p_{ij}x_j=\sum_{j=1}^{n} p_{ij}\left(-\frac{W}{V}v_j+w_j\right)=-\frac{W}{V}\sum_{j=1}^{n} p_{ij}v_j+\sum_{j=1}^{n} p_{ij}w_j=-\frac{W}{V}v_i+w_i=x_i$
  - $\sum_{i=1}^{n} x_i=\sum_{i=1}^{n}\left(-\frac{W}{V}v_i+w_i\right)=-\frac{W}{V}\sum_{i=1}^{n} v_i+\sum_{i=1}^{n} w_i=0$ 矛盾

$$\boxed{\sum_{j=1}^{n} p_{ij}v_j=v_i,\sum_{j=1}^{n} p_{ij}w_j=w_i, i=1,\cdots,n}$$

Perron定理

数学建模

- 若矩阵 **A**的所有元素均为正，则
  - **A**的模最大特征值唯一，且为正实数
  - 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
- **Google**矩阵为元素全为正的随机矩阵，**1** 为模最大特征值，重要度向量唯一且分量全为正

Oskar Perron
（**1880－1975**）
德国数学家

Perron—Frobenius定理

浙江大学 ZheJiang University

数学建模

- 若矩阵 **A** 为非负不可约（**irreducible**）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

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \lambda = 0 \qquad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \lambda_1 = \lambda_2 = 1, \; \xi = (0,1)^{\mathrm{T}} \qquad \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

Georg Frobenius
（1849－1917）
德国数学家

不可约矩阵

浙江大学 ZheJiang University
数学建模

- 若干个初等对换矩阵的乘积称为置换矩阵（**permutation matrix**）。置换矩阵每行和每列都恰有一个元素为 **1**，其余元素都为 **0**
- 若存在置换矩阵 $\mathbf{Q}$，使得 $\mathbf{Q}^{\mathrm{T}}\mathbf{A}\mathbf{Q}=\begin{pmatrix}\mathbf{X} & \mathbf{0}\\ \mathbf{Y} & \mathbf{Z}\end{pmatrix}$,

这里 **X**和 **Z**均为方阵，则称 **A**为可约矩阵（**reducible matrix**）；否则 **A**为不可约矩阵

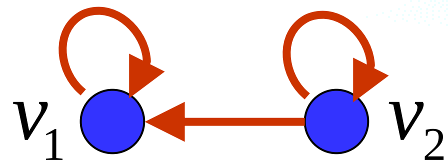
不可约矩阵

浙江大学 ZheJiang University

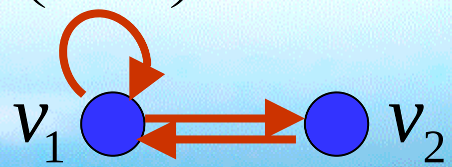
数学建模

- 若对有向图中任意有序顶点对 v_i, v_j，存在一条从 v_i 到 v_j 的有向路，则称有向图是强联通（**strongly connected**）的
- 给定非负矩阵 $\mathbf{A}=(a_{ij})_{n\times n}$，构造有向图 $G(\mathbf{A})=(V,A)$，其中 $V=\{v_1,v_2,\cdots,v_n\}$，弧 $(v_i,v_j)\in A$ 当且仅当 $a_{ij}>0$
- $\mathbf{A}$ 是不可约矩阵当且仅当 $G(\mathbf{A})$ 是强联通的

$\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}$ 可约



$\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$ 不可约



幂法

浙江大学 ZheJiang University

数学建模

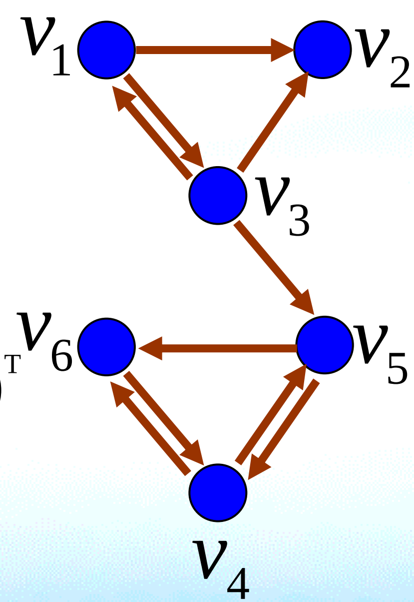
- **The World's Largest Matrix Computation**
  - **Google's PageRank is an eigenvector of a matrix of order 2.7 billion (Moler, 2002.10)**
- 幂法（**power method**）是计算矩阵模最大特征值和对应的特征向量的一种迭代算法 $\mathbf{e}^{\mathrm{T}}\mathbf{x}^{(k)}=\mathbf{e}^{\mathrm{T}}\overline{\mathbf{P}}\,\mathbf{x}^{(k-1)}=\mathbf{e}^{\mathrm{T}}\,\mathbf{x}^{(k-1)}=1$
  - 任取初始向量 $\mathbf{x}^{(0)}, \mathbf{x}^{(0)}>0$ 且 $\mathbf{e}^{\mathrm{T}}\mathbf{x}^{(0)}=1$
  - 计算 $\mathbf{x}^{(k)}=\overline{\mathbf{P}}\,\mathbf{x}^{(k-1)}$

Cleve Barry Moler
（1939－ ）
美国数学家
Matlab创始人

von Mises R, Pollaczek-Geiringer H Praktische verfahren der gleichungsauflosung. ***Zeitschrift für Angewandte Mathematik und Mechanik (Journal of Applied Mathematics and Mechanics*)**, ***9*, 58–77, 152–164, 1929.**

幂法

$$
\overline{\mathbf{P}}=\begin{pmatrix}
0.025 & 0.1667 & 0.3083 & 0.025 & 0.025 & 0.025 \\
0.45 & 0.1667 & 0.3083 & 0.025 & 0.025 & 0.025 \\
0.45 & 0.1667 & 0.025 & 0.025 & 0.025 & 0.025 \\
0.025 & 0.1667 & 0.025 & 0.025 & 0.45 & 0.875 \\
0.025 & 0.1667 & 0.3083 & 0.45 & 0.025 & 0.025 \\
0.025 & 0.1667 & 0.025 & 0.45 & 0.45 & 0.025
\end{pmatrix}
$$



$$\mathbf{x}^{(0)}=\left(0.166667,0.166667,0.166667,0.166667,0.166667,0.166667\right)^{\mathrm{T}}$$

$$\mathbf{x}^{(5)}=\left(0.057165,0.083312,0.063942,0.338898,0.196007,0.260676\right)^{\mathrm{T}}$$

$$\mathbf{x}^{(10)}=\left(0.052057,0.074290,0.057821,0.347973,0.199759,0.268101\right)^{\mathrm{T}}$$

$$\mathbf{x}^{(20)}=\left(0.051706,0.073681,0.057414,0.348701,0.199903,0.268594\right)^{\mathrm{T}}$$

$$\mathbf{x}^{(25)}=\left(0.051705,0.073679,0.057412,0.348704,0.199904,0.268596\right)^{\mathrm{T}}$$

收敛性

- 记 $\mathbf{V}$ 为所有满足 $\sum_{i=1}^{n} v_i = 0$ 的 n 维列向量 $\mathbf{v} = (v_1, \cdots, v_n)^{\mathrm{T}}$ 全体组成的空间。记 $\|\mathbf{v}\|_1 = \sum_{i=1}^{n} |v_i|$

$$\boxed{w_i = \sum_{j=1}^{n} p_{ij} v_j} \qquad \boxed{\sum_{j=1}^{n} p_{ij} = 1}$$

- 引理：对任意的 $\mathbf{v} \in \mathbf{V}$，$\mathbf{w} = \overline{\mathbf{P}}\mathbf{v} \in \mathbf{V}$，且 $\|\mathbf{w}\|_1 \le c\|\mathbf{v}\|_1$，其中 $c < 1$
  - $\sum_{i=1}^{n} w_i = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} p_{ij} v_j\right) = \sum_{j=1}^{n} v_j \left(\sum_{i=1}^{n} p_{ij}\right) = \sum_{j=1}^{n} v_j = 0$
  - 若 $\mathbf{w} = \mathbf{0}$，$\|\mathbf{w}\|_1 \le c\|\mathbf{v}\|_1$ 显然成立，若 $\mathbf{w} \ne \mathbf{0}$，$\mathbf{w}$ 的分量有正有负
  - $\|\mathbf{w}\|_1 = \sum_{i=1}^{n} |w_i| = \sum_{i=1}^{n} e_i w_i = \sum_{i=1}^{n} e_i \left(\sum_{j=1}^{n} p_{ij} v_j\right) = \sum_{j=1}^{n} v_j \left(\sum_{i=1}^{n} p_{ij} e_i\right) \le \sum_{j=1}^{n} |v_j| \left|\sum_{i=1}^{n} p_{ij} e_i\right|$

$$\le c \sum_{j=1}^{n} |v_j| = c\|\mathbf{v}\|_1$$

e_i 不同号

$$e_i = \operatorname{sgn} w_i \qquad \left|\sum_{i=1}^{n} p_{ij} e_i\right| < 1$$

收敛性

浙江大学 ZheJiang University

数学建模

- $\lim_{k\to\infty} \mathbf{x}^{(k)}$ 存在，极限即为重要度向量 $\mathbf{X}$

$$\boxed{\mathbf{X} = \overline{\mathbf{P}}\mathbf{X}}$$

  - 记 $\mathbf{v}_0 = \mathbf{x}^{(0)} - \mathbf{X} \in \mathbf{V},$

$$\mathbf{x}^{(k)} = \overline{\mathbf{P}}^k \mathbf{x}^{(0)} = \overline{\mathbf{P}}^k (\mathbf{X} + \mathbf{v}_0) = \overline{\mathbf{P}}^k \mathbf{X} + \overline{\mathbf{P}}^k \mathbf{v}_0 = \mathbf{X} + \overline{\mathbf{P}}^k \mathbf{v}_0$$

  - $\left\| \overline{\mathbf{P}}^k \mathbf{v}_0 \right\|_1 \le c^k \left\| \mathbf{v}_0 \right\|_1$

This algorithm is very popular for its simplicity. Moreover, it does not involve transformations of the matrix A which is only used through its multiplication with vectors. This property is of special interest for large and sparse matrices. However, the main drawback of the method is its very slow convergence when the ratio $|\frac{\lambda_2}{\lambda}|$ is almost equal to 1 which is a common situation for large matrices.

稀疏矩阵

浙江大学
ZheJiang University
数学建模

$$\bar{\mathbf{P}}\,\mathbf{x}^{(k)}=\left(\alpha\mathbf{P}+(1-\alpha)\frac{1}{n}\mathbf{e}\mathbf{e}^{\mathrm{T}}\right)\mathbf{x}^{(k)}$$
$$=\left(\alpha\left(\mathbf{P}_0+\frac{1}{n}\mathbf{e}\mathbf{d}^{\mathrm{T}}\right)+(1-\alpha)\frac{1}{n}\mathbf{e}\mathbf{e}^{\mathrm{T}}\right)\mathbf{x}^{(k)}$$
$$=\left(\alpha\mathbf{P}_0+\frac{1}{n}\mathbf{e}\left(\alpha\mathbf{d}^{\mathrm{T}}+(1-\alpha)\mathbf{e}^{\mathrm{T}}\right)\right)\mathbf{x}^{(k)}=\alpha\mathbf{P}_0\mathbf{x}^{(k)}+\frac{1}{n}\mathbf{e}\left(\alpha\mathbf{d}^{\mathrm{T}}\mathbf{x}^{(k)}+(1-\alpha)\right)$$

$$\boxed{\begin{aligned}\mathbf{x}^{(k+1)}&=\bar{\mathbf{P}}\,\mathbf{x}^{(k)}\\ \mathbf{e}^{\mathrm{T}}\mathbf{x}^{(k)}&=1\end{aligned}}$$

$$\boxed{\mathbf{P}=\mathbf{P}_0+\frac{1}{n}\mathbf{e}\mathbf{d}^{\mathrm{T}}}\qquad\boxed{\bar{\mathbf{P}}=\alpha\mathbf{P}+(1-\alpha)\frac{1}{n}\mathbf{e}\mathbf{e}^{\mathrm{T}}}$$

链接矩阵 ⟶ 悬挂网页修改 ⟶ **Google**矩阵

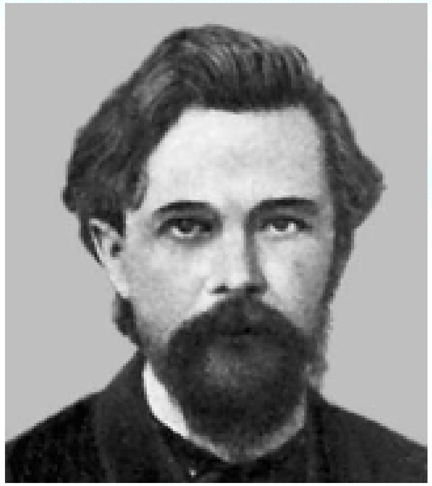
随机浏览

- 网络浏览方式
  - 从当前网页的所有链接中以相同概率随机打开一个新网页　　链接矩阵
  - 在地址栏中输入网址新建一个窗口　　悬挂网页
  - 通过链接打开网页与输入网址新建窗口的比例约为 5:1
- **Random Surfer**　　α
  - 从某网页开始浏览，最终在各网页上停留的概率是网页重要度的一种体现

Markov链

浙江大学
ZheJiang University
数学建模

- 设$\{X_m, m=0,1,2,\cdots\}$为一随机过程，状态空间有限，若 $P\{X_m=i\}$只与 X_{m-1} 有关，而与$X_{m-2}, X_{m-3},\cdots$无关，则称$\{X_m\}$为**Markov链**（**Markov chain**）
- 记 $P\{X_m=i \mid X_{m-1}=j\}=p_{ij}(m)$，若 $p_{ij}(m)$ 与 m 无关，则称**Markov**链为齐次（**homogeneous**）的。$\mathbf{P}=(p_{ij})_{n\times n}$ 为一随机矩阵，称为$\{X_m\}$的转移矩阵（**transition matrix**）



Andrei Markov
（1856－1922）
俄罗斯数学家

Markov链

浙江大学
ZheJiang University
数学建模

- 若 $P\{X_{m-1}=j\}=x_j,\ j=1,\cdots,n$， 则

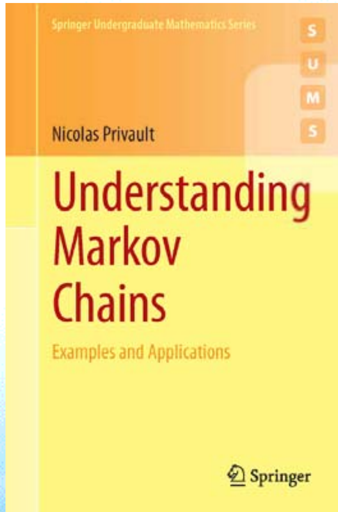
全概率公式

$$P\{X_m=i\}=\sum_{j=1}^{n}P\{X_m=i\mid X_{m-1}=j\}P\{X_{m-1}=j\}=\sum_{j=1}^{n}p_{ij}x_j$$

- 记 $\mathbf{x}^{(m)}=(P\{X_m=1\},P\{X_m=2\},\cdots,P\{X_m=n\})^{\mathrm{T}}$，

则 $\mathbf{x}^{(m)}=\mathbf{P}\mathbf{x}^{(m-1)}$ 幂法

- 不论从何网页开始随机浏览，经过充分长时间，停留在各网页上的概率组成的向量即为重要度向量

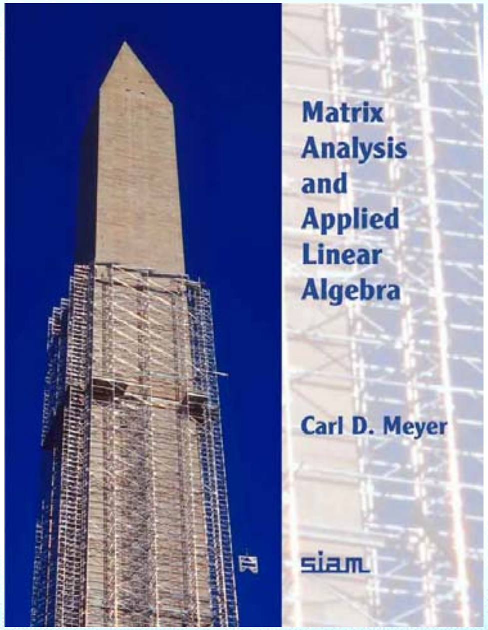


Privault N. *Understanding Markov chains: Examples and applications*. Springer, 2013

数学基石

浙江大学 ZheJiang University
数学建模

1906	Markov	Markov theory
1907	Perron	Perron theorem
1912	Frobenius	Perron–Frobenius theorem
1929	von Mises, Pollaczek-Geiringer	Power method
1998	Brin, Page	PageRank



In addition to saying something useful, the Perron–Frobenius theory is elegant. It is a testament to the fact that beautiful mathematics eventually tends to be useful, and useful mathematics eventually tends to be beautiful.

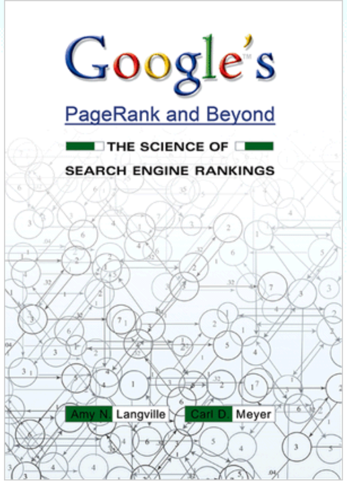
Meyer CD, Matrix Analysis and Applied Linear Algebra, SIAM, 2000.

成功原因

数学建模

- 成功原因
  - 模型简洁、适用范围广
  - 继承与发展了既往评价模型、结合与应用了多项数学成果
    - 递归形式给出的重要性定义
    - 对链接矩阵的改造与存在性、唯一性的理论证明
    - 较快的计算速度与良好的实际效果
  - **Google**公司的成功实践

Jon Michael Kleinberg (1971-)
美国康奈尔大学计算机科学系、信息科学系教授
2006年**Nevanlinna**奖得主



Kleinberg JM. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5): 604-632, 1999

Langville AN, Meyer CD. *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press, 2011.

源流与应用

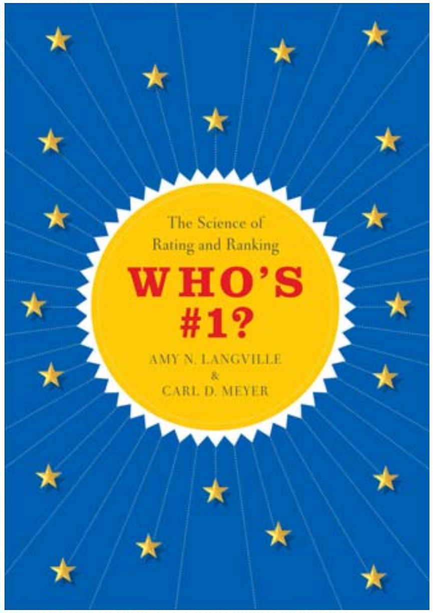
浙江大学
ZheJiang University
数学建模

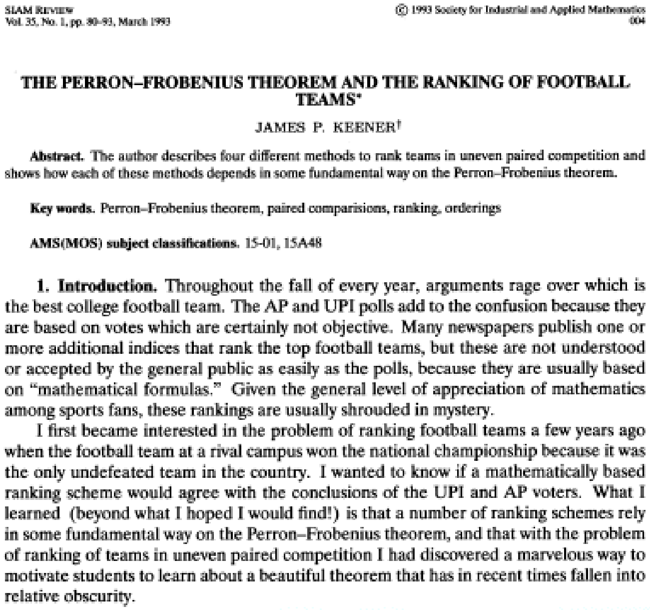
Biology	GeneRank	ProteinRank	IsoRank	
Social networks	BuddyRank	TwitterRank		VisualRank
The web	HostRank	DirRank	TrustRank	BadRank
Databases & knowledge systems	PopRank	FactRank	ObjectRank	FolkRank
Bibliometrics	TimedPageRank	CiteRank	AuthorRank	
Literature	BookRank			
Recommender systems	ItemRank			
Engineered systems	MonitorRank			
Chemistry				
Neuroscience				
Mathematical systems				
Sports				

Gleich DF. PageRank beyond the Web. *SIAM Review*, 2015, 57(3): 321-363.

体育排名

浙江大学
ZheJiang University
数学建模





SIAM REVIEW
Vol. 35, No. 1, pp. 80-93, March 1993

© 1993 Society for Industrial and Applied Mathematics
004

THE PERRON–FROBENIUS THEOREM AND THE RANKING OF FOOTBALL TEAMS*

JAMES P. KEENER†

Abstract. The author describes four different methods to rank teams in uneven paired competition and shows how each of these methods depends in some fundamental way on the Perron–Frobenius theorem.

Key words. Perron–Frobenius theorem, paired comparisions, ranking, orderings

AMS(MOS) subject classifications. 15-01, 15A48

1. Introduction. Throughout the fall of every year, arguments rage over which is the best college football team. The AP and UPI polls add to the confusion because they are based on votes which are certainly not objective. Many newspapers publish one or more additional indices that rank the top football teams, but these are not understood or accepted by the general public as easily as the polls, because they are usually based on "mathematical formulas." Given the general level of appreciation of mathematics among sports fans, these rankings are usually shrouded in mystery.

I first became interested in the problem of ranking football teams a few years ago when the football team at a rival campus won the national championship because it was the only undefeated team in the country. I wanted to know if a mathematically based ranking scheme would agree with the conclusions of the UPI and AP voters. What I learned (beyond what I hoped I would find!) is that a number of ranking schemes rely in some fundamental way on the Perron–Frobenius theorem, and that with the problem of ranking of teams in uneven paired competition I had discovered a marvelous way to motivate students to learn about a beautiful theorem that has in recent times fallen into relative obscurity.

Distinguished Professor of Mathematics
Adjunct Professor of Bioengineering
University of Utah

Langville AN, Meyer CD. *Who's # 1?: the science of rating and ranking*. Princeton University Press, 2012.

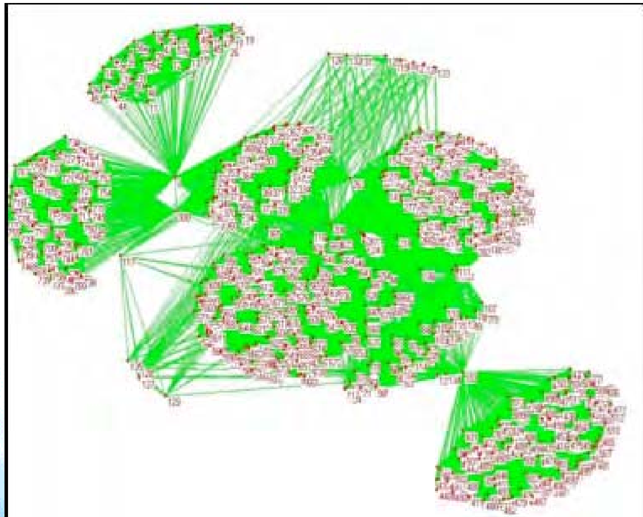
Keener J P. The Perron-Frobenius theorem and the ranking of football teams. *SIAM Review*, 35, 80-93, 1993.

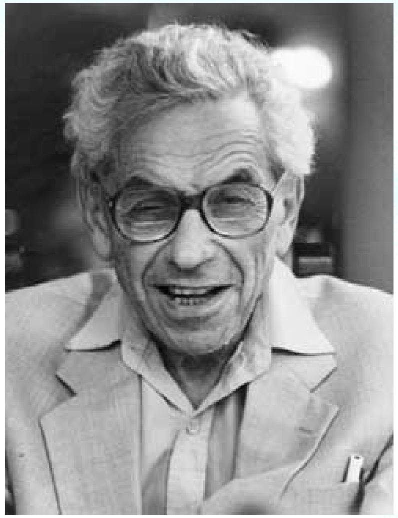
浙江大学
ZheJiang University
数学建模

2014 ICM

- **Using Networks to Measure Influence and Impact**
  - **Erdős number**

Erdös number	0	---	1	person
Erdös number	1	---	504	people
Erdös number	2	---	6593	people
Erdös number	3	---	33605	people
Erdös number	4	---	83642	people
Erdös number	5	---	87760	people
Erdös number	6	---	40014	people
Erdös number	7	---	11591	people
Erdös number	8	---	3146	people
Erdös number	9	---	819	people
Erdös number	10	---	244	people
Erdös number	11	---	68	people
Erdös number	12	---	23	people
Erdös number	13	---	5	people





Paul Erdős
(1913-1996)
匈牙利数学家
1984年Wolf奖得主

2014 ICM

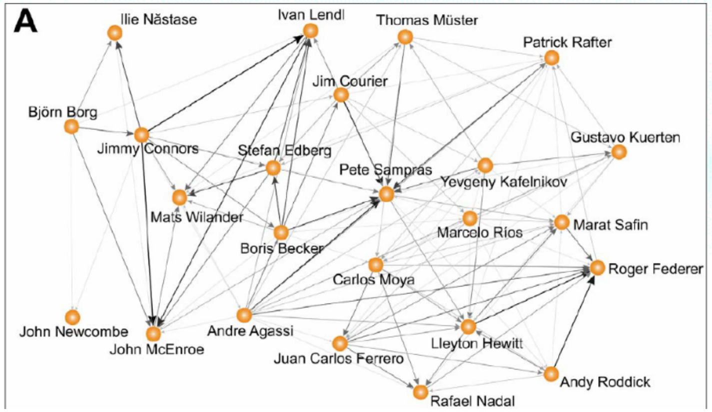
数学建模

- Build the co-author network of the Erdos1 authors (510 authors, 18000 records). Once built, analyze the properties of this network
- Develop influence measure(s) to determine who in this Erdos1 network has significant influence within the network
- Choose some set of foundational papers in the emerging field of network science. Use these papers to analyze and develop a model to determine their relative influence. Is there a similar way to determine the role or influence measure of an individual network researcher? Consider how you would measure the impact of a specific university, department, or a journal in network science?
- Implement your algorithm on a completely different set of network influence data, for instance, movie actors
- Could individuals, organizations, nations, and society use influence methodology to improve relationships, conduct business, and make wise decisions

2014 MCM B

浙江大学
ZheJiang University
数学建模

- **College Coaching Legends**
  - **Build a mathematical model to choose the best college coach or coaches (past or present) from among either male or female coaches in such sports as college hockey or field hockey, football, baseball or softball, basketball, or soccer.**
  - **Does it make a difference which time line horizon that you use in your analysis. Discuss how your model can be applied in general across both genders and all possible sports.**



Radicchi F. Who is the best player ever? A complex network analysis of the history of professional tennis. *PloS one*, 6(2): e17249, 2011.


浙江大学
ZheJiang University
谢 谢