数学建模

浙江大学数学系　谈之奕

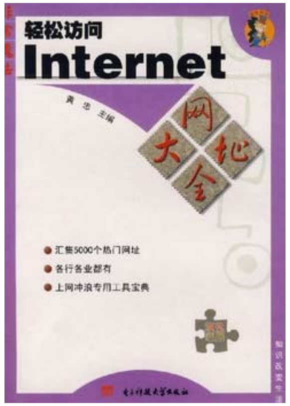
tanzy@zju.edu.cn


浙江大学
ZheJiang University
Google中的数学模型

Google中的数学模型

浙江大学
ZheJiang University
数学建模

- **Internet浏览**
  - 网址大全
  - 分类链接



- 搜索引擎

存储网页 ⟶ 读取短语 ⟶ 建立索引 ⟶

关键字匹配 ⟶ 相关性优先 ⟶ 列表显示结果

搜索引擎

- 符合要求的网页数量过多，无法保证所需内容在列表靠前位置出现，用户体验不佳
- Internet vs 图书馆
  - 大
  - 动态
  - 自组织
  - 超链接

"When Larry Met Sergey"

浙江大学 ZheJiang University

数学建模

- **1995**年 **Brin** 和 **Page** 进入斯坦福大学攻读博士学位，并开始探索搜索引擎的设计
- **1996**年左右 **Page** 给出了衡量网页重要度的模型和算法**PageRank™**，成为后来**Google**搜索引擎的核心

Google创始人

Larry Page(左**)(1973.3-)**

和 **Sergey Brin(**右**)(1973.8-)**

Google的诞生

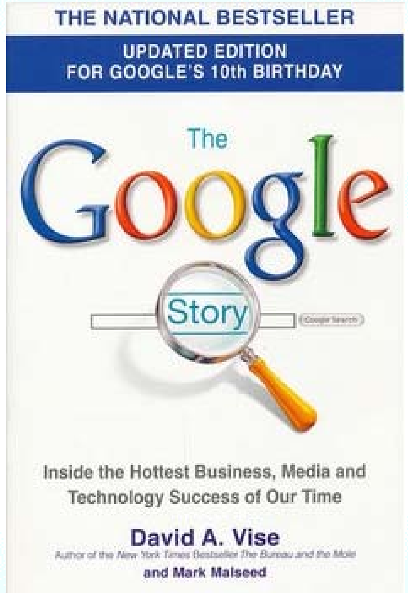
浙江大学
ZheJiang University
数学建模

- **1998年，Brin和Page创立了Google公司，2004年，公司在Nasdaq上市**

Googol 10^{100}

Googolplex $10^{10^{100}}$

- **2012年，Google公司市值已达2500亿美元，利润100多亿美元，Google是公认的全球规模最大的搜索引擎**



The Google Story（中译名：撬动地球的Google）

网页重要度

- 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

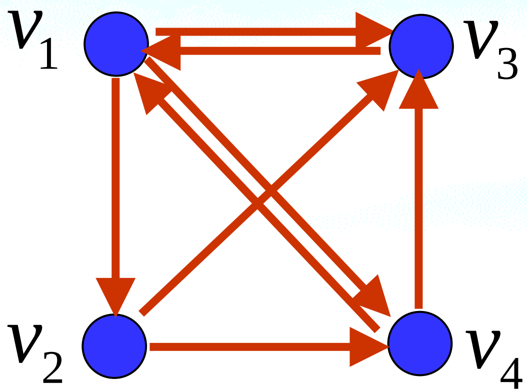
    对其它网页的贡献与自身的重要度成正比
  - （平等性）网页链接较多时对它所链接的网页的重要度的贡献比链接较少时对它所链接的网页的重要度的贡献小

    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

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**）

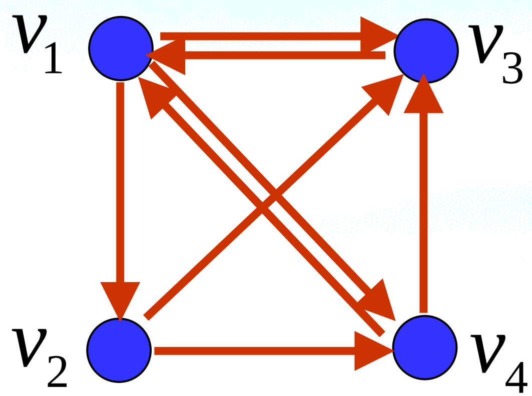


链接矩阵

- **定义**

$$p_{ij}=\begin{cases}\dfrac{1}{q_j}, & \text{若有链接自} v_j \text{链向} v_i \\ 0, & \text{其他}\end{cases}$$

则 $\mathbf{P}=(p_{ij})_{n\times n}$ 为一 n 阶方阵，称为链接矩阵，n 为网页数目



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

重要度向量

浙江大学 ZheJiang University

数学建模

- 记 x_i 为网页 v_i 的重要度。称 $\mathbf{X}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 为网页重要度向量
- 若链接到网页 v_i 的网页有 $v_{j_1},v_{j_2},\cdots,v_{j_k}$ ，则

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

$$=p_{i1}x_1+p_{i2}x_2\cdots+p_{in}x_n$$

$$=\sum_{j=1}^{n}p_{ij}x_j,\qquad i=1,\cdots,n$$

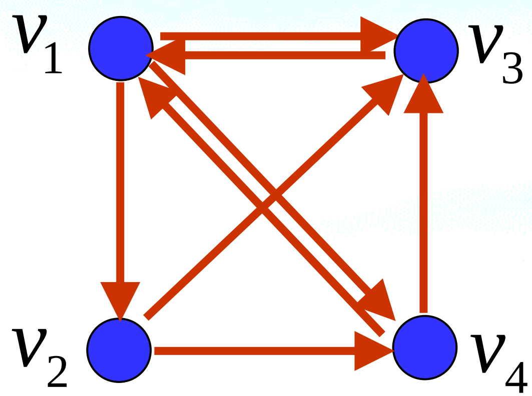
$$p_{ij}=\begin{cases}\dfrac{1}{q_j}, & (v_j,v_i)\in A\\ 0, & 其他\end{cases}$$

- $\mathbf{X}$ 为线性方程组 $\mathbf{X}=\mathbf{PX}$ 的解

v_j

重要度向量

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



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

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

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

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

非零解

- 对任意矩阵 $\mathbf{P}$， $\mathbf{X}=\mathbf{P}\mathbf{X}$ 是否总有非零解
  - $\operatorname{rank}(\mathbf{I}-\mathbf{P})<n$
  - $\mathbf{P}$ 是否总有特征值 1
  - $\mathbf{P}^{\mathrm{T}}$ 是否总有特征值 1
  - $\mathbf{P}^{\mathrm{T}}\mathbf{e}=1\,\mathbf{e},\ \mathbf{e}=(1,\cdots,1)^{\mathrm{T}}$

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

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

随机矩阵

浙江大学 ZheJiang University

数学建模

- 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**
- 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
- **1** 是任一随机矩阵的模最大特征值

随机矩阵

数学建模

- 设 λ 是（行）随机矩阵 $\mathbf{P}=(p_{ij})_{n\times n}$ 的特征值，则 $|\lambda|\le 1$
  - 设 $\mathbf{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 为属于特征值 λ 的特征向量，$|x_i|=\max\limits_{1\le j\le n}|x_j|>0$
  - $$\lambda x_i=\sum_{j=1}^{n}p_{ij}x_j$$

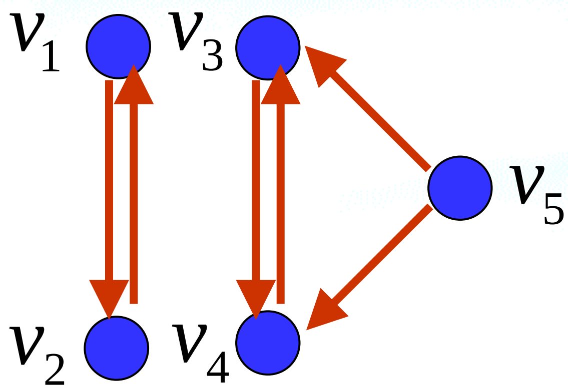
$$|\lambda||x_i|=\left|\lambda x_i\right|=\left|\sum_{j=1}^{n}p_{ij}x_j\right|\le\sum_{j=1}^{n}|p_{ij}||x_j|\le|x_i|\sum_{j=1}^{n}|p_{ij}|=|x_i|$$

悬挂网页

- 若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

唯一性

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



- 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 **1** 的特征子空间维数大于**1**），则用上述方法可能得到相互矛盾的网页重要度比较结果

Google矩阵

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

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

唯一性的初等证明

浙江大学 ZheJiang University

数学建模

- 设 $\mathbf{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 为 $\overline{\mathbf{P}}$ 的属于特征值 **1** 的特征向量
  - $x_1,x_2,\cdots,x_n$ 要么恒非负，要么恒非正
    - 反证法 $x_i=\sum_{j=1}^{n}p_{ij}x_j, i=1,\cdots,n$

$$\left|x_i\right|=\left|\sum_{j=1}^{n}p_{ij}x_j\right|<\sum_{j=1}^{n}p_{ij}\left|x_j\right|, i=1,\cdots,n$$

$$\sum_{i=1}^{n}\left|x_i\right|<\sum_{i=1}^{n}\sum_{j=1}^{n}p_{ij}\left|x_j\right|=\sum_{j=1}^{n}\left|x_j\right|\left(\sum_{i=1}^{n}p_{ij}\right)=\sum_{j=1}^{n}\left|x_j\right| \quad \text{矛盾}$$

唯一性的初等证明

浙江大学 ZheJiang University 数学建模

- 若 $\mathbf{v}=(v_1,\cdots,v_n)^{\mathrm{T}}$ 和 $\mathbf{w}=(w_1,\cdots,w_n)^{\mathrm{T}}$ 是 $\overline{\mathbf{P}}$ 的两个属于特征值 **1** 的特征向量
  - $\sum_{j=1}^{n} p_{ij}v_j = v_i, \sum_{j=1}^{n} p_{ij}w_j = w_i, i=1,\cdots,n$
  - 记 $W=\sum_{k=1}^{n} w_k$, $V=\sum_{k=1}^{n} v_k \neq 0$ （特征向量非零）
  - 令 $x_i = -\frac{W}{V}v_i + w_i, i=1,\cdots,n$

唯一性的初等证明

浙江大学 ZheJiang University

数学建模

$$x_i = -\frac{W}{V}v_i + w_i, i = 1, \cdots, n$$
$$\sum_{j=1}^{n} p_{ij} v_j = v_i, \sum_{j=1}^{n} p_{ij} w_j = w_i$$

- $$\sum_{j=1}^{n} p_{ij} x_j = \sum_{j=1}^{n} p_{ij}\left(-\frac{W}{V}v_j + w_j\right)$$
$$= -\frac{W}{V}\sum_{j=1}^{n} p_{ij} v_j + \sum_{j=1}^{n} p_{ij} w_j = -\frac{W}{V}v_i + w_i = x_i$$
  - $\mathbf{x} = (x_1, \cdots, x_n)^{\mathrm{T}}$是 $\overline{\mathbf{P}}$ 的属于特征值 1 的特征向量
- $$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n}\left(-\frac{W}{V}v_i + w_i\right) = -\frac{W}{V}\sum_{i=1}^{n} v_i + \sum_{i=1}^{n} w_i = 0$$
  - 与 $x_1, x_2, \cdots, x_n$ 要么恒非负，要么恒非正矛盾

Perron定理

- 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

浙江大学 ZheJiang University

数学建模

Perron—Frobenius定理

- 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

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

Georg Frobenius
（**1849－1917**）
德国数学家

浙江大学
ZheJiang University
数学建模

不可约矩阵

- 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 $\mathbf{Q}^{\mathrm{T}}\mathbf{A}\mathbf{Q}=\begin{pmatrix}\mathbf{X} & \mathbf{0}\\ \mathbf{Y} & \mathbf{Z}\end{pmatrix}$，

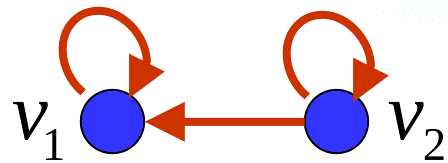
  这里 **X**和 **Z** 均为方阵，则称 **A**为可约矩阵（**reducible matrix**）；否则 **A**为不可约矩阵

A

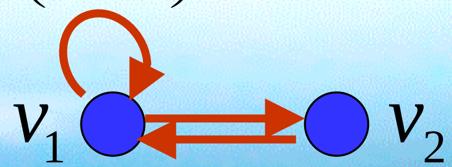
不可约矩阵

- 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

- **The World's Largest Matrix Computation**
  - **Google's PageRank is an eigenvector of a matrix of order 2.7 billion**

——Cleve Moler（Matlab创始人），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

浙江大学 ZheJiang University
数学建模

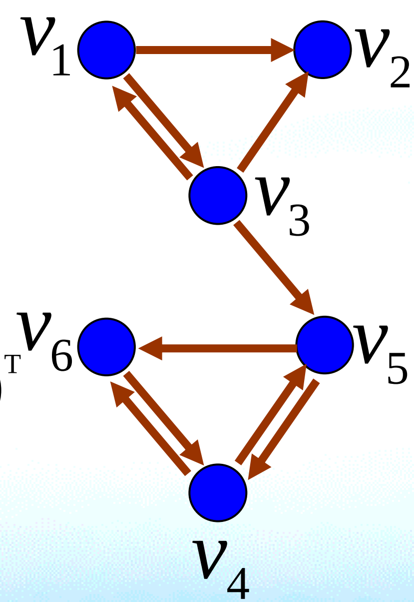
幂法

- 任取初始向量 $\mathbf{x}^{(0)}, \mathbf{x}^{(0)} > 0$ 且 $\mathbf{x}^{(0)^{\mathrm{T}}}\mathbf{e} = 1$
- 计算 $\mathbf{x}^{(k)} = \overline{\mathbf{P}}\,\mathbf{x}^{(k-1)}$
- $\lim\limits_{k\to\infty} \mathbf{x}^{(k)}$ 存在，极限即为重要度向量 $\mathbf{X}$

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 (1929), 152–164.

幂法

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



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

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

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

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

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

收敛性

浙江大学 ZheJiang University

数学建模

- 记 $\mathbf{V}$ 为所有满足 $\mathbf{v}^{\mathrm{T}}\mathbf{e}=0$ 的 n 维列向量 $\mathbf{v}$ 组成的空间。定义 $\|\mathbf{v}\|_1=\sum_{i=1}^{n}|v_i|$
- 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 $\mathbf{w}=\mathbf{0}$ ， 结论显然成立；若 $\mathbf{w}\neq\mathbf{0}$ ，$\mathbf{w}$ 的分量有正有负

收敛性

- 记 $e_i = \operatorname{sgn} w_i, i = 1, \cdots, n$， 存在 $e_{i_1} = 1, e_{i_2} = -1$
- $$\sum_{i=1}^{n} p_{ij} e_i = \sum_{i \neq i_1}^{n} p_{ij} e_i + p_{i_1 j} e_{i_1} \geq -\sum_{i \neq i_1}^{n} p_{ij} + p_{i_1 j} = -\sum_{i=1}^{n} p_{ij} + 2 p_{i_1 j} > -1$$
  $$\sum_{i=1}^{n} p_{ij} e_i = \sum_{i \neq i_2}^{n} p_{ij} e_i + p_{i_2 j} e_{i_2} \leq \sum_{i \neq i_2}^{n} p_{ij} - p_{i_2 j} = \sum_{i=1}^{n} p_{ij} - 2 p_{i_2 j} < 1$$
- $$\|\mathbf{w}\|_1 = \sum_{i=1}^{n} |w_i| = \sum_{i=1}^{n} e_i w_i = \sum_{i=1}^{n} e_i \left(\sum_{j=1}^{n} p_{ij} v_j \right) = \sum_{j=1}^{n} v_j \left(\sum_{i=1}^{n} p_{ij} e_i \right)$$
  $$\leq \sum_{j=1}^{n} |v_j| \left| \sum_{i=1}^{n} p_{ij} e_i \right| \leq c \sum_{j=1}^{n} |v_j| = c \|\mathbf{v}\|_1$$

收敛性

数学建模

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

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

- 由 $\left\| \overline{\mathbf{P}}^k \mathbf{v}_0 \right\|_1 \le c^k \left\| \mathbf{v}_0 \right\|_1$，即得 $\lim\limits_{k\to\infty} \mathbf{x}^{(k)} = \mathbf{X}$
- 幂法收敛速度较慢，但实现简单，存储量小，对稀疏矩阵运算量显著减小，适于**Google**矩阵重要度向量计算

链接矩阵 $\longrightarrow$ 悬挂网页修改 $\longrightarrow$ **Google**矩阵

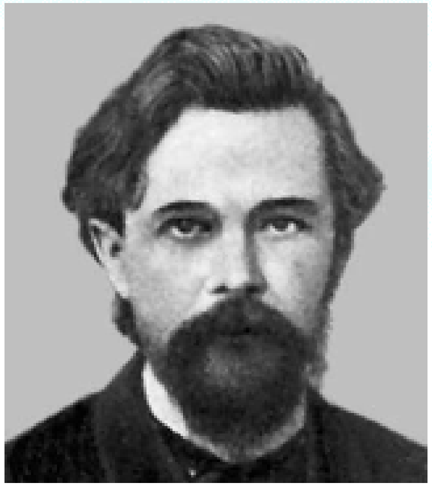
随机浏览

- 链接矩阵：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
- 悬挂网页：在地址栏中输入网址新建一个窗口
- 参数α：通过链接打开网页与输入网址新建窗口的比例约为 5:1
- 从某网页开始按上述模式浏览，在各网页上停留的概率为何

Markov链

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**）



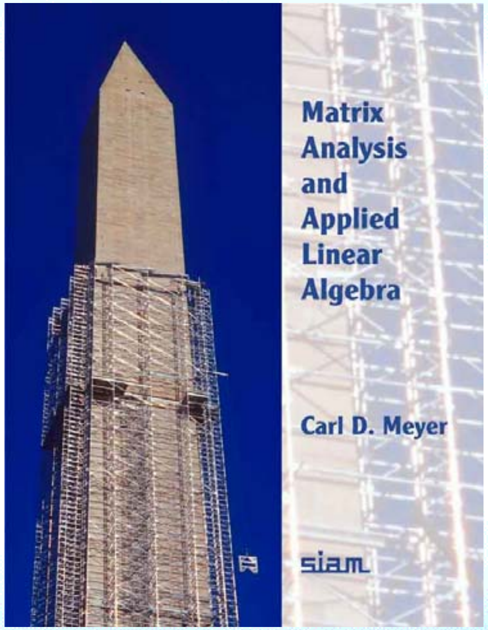
Andrei Markov
（1856－1922）
俄罗斯数学家

Markov链

- 若 $P\{X_{m-1}=j\}=x_j,\ j=1,\cdots,n$ ，则
  $$P\{X_m=i\}=\sum_{j=1}^{n}P\{X_m=i\,|\,X_{m-1}=j\}P\{X_{m-1}=j\}=\sum_{j=1}^{n}p_{ij}x_j$$
- 记 $\mathbf{x}^{(m)}=(P\{X_m=1\},P\{X_m=2\},\cdots,P\{X_m=n\})^{\mathrm{T}}$ ，
  则 $\mathbf{x}^{(m)}=\mathbf{P}\mathbf{x}^{(m-1)}$
- 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

数学基石

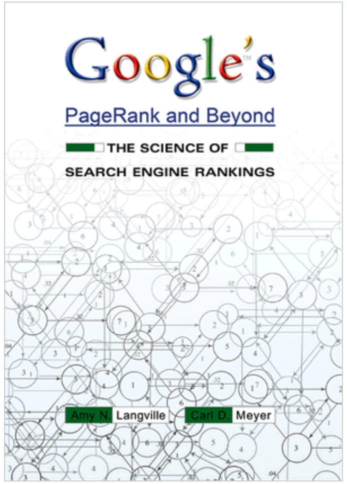
1906	**Markov**	**Markov theory**
1907	**Perron**	**Perron theorem**
1912	**Frobenius**	**Perron–Frobenius theorem**
1929	**von Mises, Pollaczek-Geiringer**	**Power method**
1998	**Brin, Page**	**PageRank**

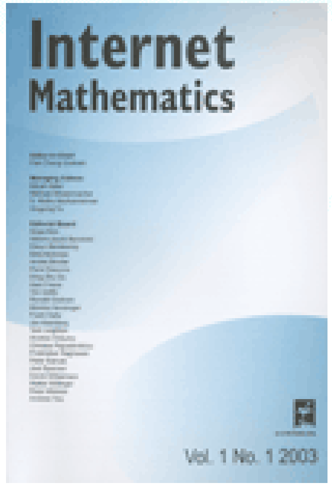


In addition to saying something useful, the Perron–Frobenius theory is elegant. It is a testament to the fact that beautiful mathematics eventually tends to be useful, and useful mathematics eventually tends to be beautiful.

未尽探索

- 计算与更新
- 个性搜索
- 与搜索引擎优化（**SEO**）的博弈
- 当前算法





Internet Mathematics publishes research papers that address fundamental problems, both conceptual and algorithmic, that arise in dealing with large complex information networks such as the Internet.

他山之石

数学建模

- 文献计量学（**Bibliometrics**）
  - **1960**年，**Eugene Garfield**创立了美国科学情报研究所（**Institute for Scientific Information, ISI**），**Science Citation Index**（**SCI**）是其主要数据库之一
  - **1963**年，**Eugene Garfield**和**Irv Sher**提出了影响因子（**Impact Factor**）的概念，用发表的论文被引用次数衡量学术期刊的优劣
  - **1976**年，**Gabriel Pinski**和**Francis Narin**提出了一种改进指标
    - **A journal is influential if it is cited by other influential journals**

文献计量学

数学建模

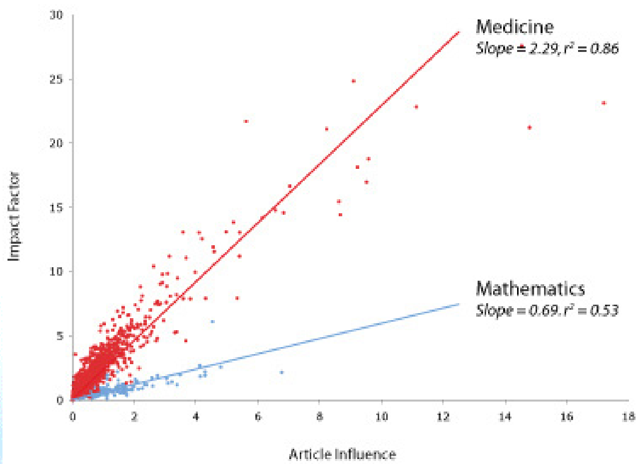
Mark	Rank	Abbreviated Journal Title (linked to journal information)	ISSN	JCR Data						Eigenfactor® Metrics	
				Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Articles	Cited Half-life	Eigenfactor® Score	Article Influence® Score
☐	1	J ACM	0004-5411	4536	2.353	2.709	0.273	22	>10.0	0.00789	2.721

Journal Impact Factor

Cites in 2011 to items published in: 2010 = 49
2009 = 111
Sum: 160

Number of items published in: 2010 = 32
2009 = 36
Sum: 68

Calculation: Cites to recent items / Number of recent items = 160 / 68 = **2.353**



eigenFACTOR.org®
RANKING AND MAPPING SCIENTIFIC KNOWLEDGE

eigenfactor search | mapping | information | well-formed | cost effectiveness | recommendation | gender | about

why eigenFACTOR?

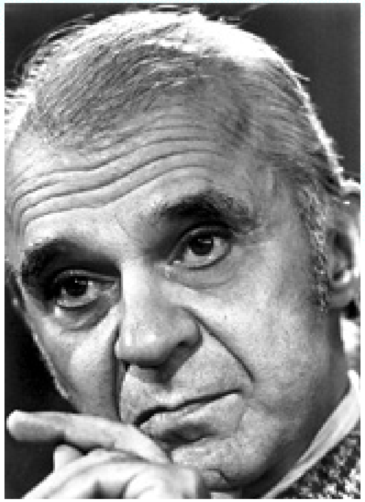
1 *Eigenfactor®* scores and *Article Influence®* scores rank journals much as Google ranks websites.
Scholarly references join journals together in a vast network of citations. Our algorithms use the structure of the entire network (instead of purely local citation information) to evaluate the importance of each journal.

http://www.webofknowledge.com/
http://www.eigenfactor.org/

投入产出模型

- 投入产出模型由**Leontief**于**1936**年创立并逐步发展完善的
  - **Input-output model gives economic science an important tool of analysis for studying the complicated interdependence within the production system in a modern economy.**
  - **Not only constructed the theoretical foundations of the input-output method, also developed the empirical data that are necessary to utilize the method on important economic problems as well as to test empirically various economic theories.**

——selected from Nobel Prize Award Ceremony Speech by Assar Lindbeck



Wassily Leontief
（**1905－1999**）
美籍俄裔经济学家
1973年诺贝尔经济学奖得主

投入产出表

浙江大学 ZheJiang University

数学建模

	农业	工业	服务业	最终需求	总产出
农业	15	20	30	35	100
工业	30	10	45	115	200
服务业	20	60	0	70	150

为实现100亿农业产值需投入农业产值15亿，工业产值30亿，服务业产值20亿

100亿农业产值中，15亿用于农业，20亿用于工业，30亿元用于服务业，35亿用于满足最终需求

投入产出模型

浙江大学 ZheJiang University 数学建模

- x_i：部门 i 的总产出
- d_i：部门 i 的最终需求
- a_{ij}：部门 i 的产出中用于部门 j 的产值

$$\sum_{j=1}^{n} a_{ij} + d_i = x_i \quad \Longrightarrow \quad \sum_{j=1}^{n} t_{ij} x_j + d_i = x_i$$

- $t_{ij} = \dfrac{a_{ij}}{x_j}$：部门 j 生产单位产值的产品需投入部门 i 的 t_{ij} 个单位的产值（直接消耗系数）

投入产出模型

- 令 $\mathbf{T}=(t_{ij})_{n\times n}$ ，$\mathbf{d}=(d_1,\cdots,d_n)^{\mathrm{T}}$, $\mathbf{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$,
投入产出模型可表示为

$$(\mathbf{I}-\mathbf{T})\mathbf{x}=\mathbf{d}$$

$$\boxed{\sum_{j=1}^{n} t_{ij}x_j + d_i = x_i}$$

- 若 $\mathbf{d}\neq \mathbf{0}$，模型称为开放（**open**）的；
若 $\mathbf{d}=\mathbf{0}$，模型称为封闭（**closed**）的
- 封闭投入产出模型和**PageRank**模型形式相同

投入产出模型

浙江大学 ZheJiang University

数学建模

- 直接消耗系数短期内变化不大，其值可通过统计获得
- 若对任何最终需求，方程总有非负解，则经济系统可行（**feasible**）
- 给定直接消耗系数，如何判断经济系统是否可行，如何求出一定最终需求下各部门的产值

国家统计局 国家发展和改革委员会 财政部关于认真做好2012年全国投入产出调查工作的通知

字体：[大 中 小] 日期：2012-06-27 16:53:10 访问次数：142 信息来源：浙江省统计局

国家统计局 国家发展和改革委员会 财政部

关于认真做好2012年全国投入产出调查工作的通知

国统字〔2012〕16号

各省、自治区、直辖市统计局、发展改革委、财政厅（局）及国务院有关部门：

按照《国务院办公厅关于进行全国投入产出调查的通知》（国办发〔1987〕18号）的要求，2012年将开展全国投入产出调查和编制投入产出表。为认真做好2012年全国投入产出调查和编制投入产出表工作，现将有关事项通知如下：

一、调查目的和意义

投入产出调查是编制国家和地区投入产出表的重要基础。投入产出表是国民经济核算体系的重要组成部分，是开展政策模拟和定量分析的有力工具，对宏观经济管理和决策具有重要意义。

二、调查对象和范围

这次投入产出调查的对象是我国的重点法人单位，涉及除农林牧渔业外的所有国民经济行业。具体范围包括：采矿业，制造业，电力、热力、燃气及水的生产和供应业，建筑业，批发和零售业，交通运输、仓储和邮

投入产出表

	农业	工业	服务业	最终需求	总产出
农业	15	20	30	35	100
工业	30	10	45	115	200
服务业	20	60	0	70	150

$$\mathbf{T}=\begin{pmatrix} 0.15 & 0.10 & 0.20 \\ 0.30 & 0.05 & 0.30 \\ 0.20 & 0.30 & 0.00 \end{pmatrix} \qquad \mathbf{A}=\mathbf{I}-\mathbf{T}=\begin{pmatrix} 0.85 & -0.10 & -0.20 \\ -0.30 & 0.95 & -0.30 \\ -0.20 & -0.30 & 1.00 \end{pmatrix}$$

浙江大学 ZheJiang University

数学建模

投入产出表

- $\mathbf{A}=\mathbf{I}-\mathbf{T}$ 的非主对角元素非正（Z-矩阵）

$$\mathbf{A}^{-1}=\begin{pmatrix}1.3459 & 0.2504 & 0.3443\\ 0.5634 & 1.2676 & 0.4930\\ 0.4382 & 0.4304 & 1.2167\end{pmatrix}\geq \mathbf{0}$$

$$\boxed{(\mathbf{I}-\mathbf{T})\mathbf{x}=\mathbf{d}}$$

$$\mathbf{d}=\begin{pmatrix}100\\ 200\\ 300\end{pmatrix},\mathbf{x}=\mathbf{A}^{-1}\mathbf{d}=\begin{pmatrix}287.96\\ 457.76\\ 494.91\end{pmatrix}$$

$$\mathbf{d}=\begin{pmatrix}300\\ 200\\ 300\end{pmatrix},\mathbf{x}=\begin{pmatrix}557.14\\ 570.44\\ 582.55\end{pmatrix}$$

M-矩阵

- 开放投入产出模型可行，当且仅当$\mathbf{A}^{-1} \geq \mathbf{0}$，满足上述条件的矩阵也称为 **M-矩阵**
- 若 **A** 为**Z-**矩阵，则 **A** 为**M-**矩阵当且仅当 **A** 的所有主子式为正
  - **1949**年，**David Hawkins**和**Herbert Alexander Simon**（**1975**年**Turing**奖、**1978**年**Nobel**经济学奖得主）证明了上述开放投入产出模型可行的条件。事实上，条件的等价性早在**1937**年已由**Alexander Ostrowski**证明


浙江大学
ZheJiang University
谢　谢