[S. Unnikrishna Pillai, Torsten Suel, and Seunghun Cha]

© DIGITALVISION

The Perron-Frobenius Theorem

[Some of its applications]

The Perron-Frobenius theorem provides a simple characterization of the eigenvectors and eigenvalues of certain types of matrices with nonnegative entries. The importance of the Perron-Frobenius theorem stems from the fact that eigenvalue problems on these types of matrices frequently arise in many different fields of science and engineering. In this article, we discuss applications of this theorem in such diverse areas as steady state behavior of Markov chains, power control in wireless networks, commodity pricing models in economics, population growth models, and Web search engines. We start out with a review and discussion of the mathematical foundations.

A theorem derived by Oscar Perron (1907) and later generalized by Frobenius (1912) has several interesting applications in engineering and economics: such as power control problem in wireless communications, steady state probability distribution in certain Markov chains, commodity pricing, population dynamics, and ranking techniques in Web search engines.

Perron's theorem deals with positive matrices, i.e., matrices whose entries are strictly positive. More generally, Perron's theorem is true for all primitive matrices (Let A be a nonnegative matrix whose entries a_{ij} are nonnegative numbers. A is said to be primitive if, for some integer m_0, A^{m_0} is a positive matrix. i.e., $a_{ij}^{(m_0)} > 0$ where $a_{ij}^{(m_0)}$ represents the (i, j)th entry of A^{m_0}). For example, the square matrix

1053-5888/05/$20.00©2005IEEE

$$P = \begin{pmatrix} p_0 & p_1 & p_2 & \cdots & p_m \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

is primitive for all $p_i > 0$ since P^2 is a positive matrix. More interestingly, the $m \times m$ matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ p & q & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

is also primitive for $p > 0$, $q > 0$ since P^n is a positive matrix for $n = m^2 - 2m + 2$. To state Perron's theorem, we need to define the spectral radius $\rho(A)$ of a matrix A. The spectral radius of a matrix A represents the maximum of the absolute values of the eigenvalues of A. An eigenvalue/eigenvector pair of the matrix A satisfies the equation $A\underline{x} = \lambda\underline{x}$, where λ and $\underline{x}$ represent the eigenvalue and the corresponding eigenvector, respectively. In this article, both upper and lower case letters with underbar are used to represent vectors, and upper case letters are used to represent matrices. In matrix notation, A' represents the transpose of A. [13]. Thus

$$\rho(A) = \max_i |\lambda_i(A)|, \tag{1}$$

where $\lambda_i(A)$ represents the ith eigenvalue of A. Let r_i and c_j represent the ith row sum and jth column sum of A. Thus

$$r_i = \sum_j a_{ij}, \quad c_j = \sum_i a_{ij}.$$

Then it is easy to show that the spectral radius satisfies the inequalities

$$\min_i r_i \le \rho(A) \le \max_i r_i \tag{2}$$
$$\min_j c_j \le \rho(A) \le \max_j c_j. \tag{3}$$

In general, a matrix need not have an eigenvalue equal to its spectral radius, but as Perron's theorem shows, primitive matrices do possess such an eigenvalue.

THE THEOREMS

PERRON'S THEOREM

For a primitive matrix A with spectral radius $\rho(A)$, we have

i) $\rho(A) > 0$ and $\rho(A)$ is an eigenvalue of A with multiplicity one (i.e., the largest eigenvalue of A is always positive and simple and it equals the spectral radius of A).

ii) The left and right eigenvectors of A corresponding to the eigenvalue $\rho(A)$ are both positive (with positive entries), i.e., there exist positive column vectors $\underline{x}_0$ and $\underline{y}_0$ such that $A\underline{x}_0 = \lambda_0\underline{x}_0$; $\underline{y}'_0 A = \lambda_0 \underline{y}'_0$ where $\lambda_0 = \rho(A)$. We shall refer to $\underline{x}_0$ and $\underline{y}_0$ as the right- and left-Perron vectors of A.

iii) If λ is any other eigenvalue of A, then $|\lambda| < \rho(A)$. In particular, there is no other eigenvalue λ for A such that $|\lambda| = \rho(A)$.

The Perron-Frobenius theorem, on the other hand, refers to nonnegative irreducible matrices. Recall that A is irreducible if there does not exist a permutation matrix S such that [17]

$$SAS' = \begin{pmatrix} B & 0 \\ C & D \end{pmatrix},$$

where B and D are square matrices. Therefore, for reducible matrices, by identical row and column operations it is possible to rewrite A so that the upper right-hand block is full of zeros. In terms of homogeneous Markov chain terminology, let $p_{ij} \ge 0$ represent the one-step transition probability from state e_i to state e_j for the system. Thus, $p_{ij} = P\{X_{n+1} = e_j | X_n = e_i\}$, where X_n represents the state of the system at stage n, and let $P = (p_{ij})_{i,j=1}^m$ represent the probability transition matrix of a finite Markov chain with m states. Since transition probabilities satisfy $\sum_{j=1}^m p_{ij} = 1$, from (2) every probability transition matrix has spectral radius equal to unity. From Perron's theorem, for every probability transition matrix P, there exist two positive vectors $\underline{x}_0$ and $\underline{y}_0$ such that

$$P\underline{x}_0 = \underline{x}_0, \quad \underline{y}'_0 P = \underline{y}'_0 \tag{4}$$

Interestingly $\underline{x}_0 = [1, 1, \ldots, 1]'$ in this case, and the largest eigenvalue λ_0 of P equals unity. Let λ_k, $k = 1, 2, \ldots, m-1$ represent the remaining eigenvalues of P, and $\underline{x}_k, \underline{y}_k$, $k = 1, 2, \ldots, m$ the corresponding eigenvectors of P and P'. Then $|\lambda_n| < 1$, $k = 1, 2, \ldots, m-1$ and [13]

$$P = \underline{x}_0 \underline{y}'_0 + \sum_{k=1}^{m-1} \lambda_k \underline{x}_k \underline{y}'_k \tag{5}$$

which gives

$$P^n = \underline{x}_0 \underline{y}'_0 + \sum_{k=1}^{m-1} \lambda_k^n \underline{x}_k \underline{y}'_k. \tag{6}$$

In this representation, the k-step transition probability $P\{X_{n+k} = e_j | X_n = e_i\} = p_{ij}^{(k)}$ is given by the (i, j)th entry of P^k. If the probability transition matrix P corresponding to a Markov chain is irreducible, then starting from any state e_i, it is possible to get to any other state e_j in a certain number of steps, i.e., P irreducible implies

$$p_{ij}^{(m_{ij})} > 0 \quad \text{for any } e_i, e_j.$$

Here, m may depend on i and j, and in particular, there may not exist an m_0 such that

$$p_{ij}^{(m_0)} > 0 \quad \text{for all } e_i,\ e_j. \tag{7}$$

However, (7) is true for primitive matrices. In the case of primitive matrices, there exists a stage m_0 at which every state is accessible from every other state. All primitive matrices are irreducible, but all irreducible matrices are not necessarily primitive. For example, the matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is irreducible since $p_{12} = p_{21} = 1$, and $P^2 = I = (p_{ij}^2)$ gives $p_{11}^{(2)} = p_{22}^{(2)} = 1$. However P is not primitive since $p_{11}^{(n)} = 0$ if n is odd and $p_{12}^{(n)} = 0$ if n is even. Similarly, the $m \times m$ matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

is irreducible but not primitive!

One interesting problem in this context is to study the long-term behavior (steady state probabilities) of a Markov chain, i.e., what are the limiting probabilities of a Markov chain given that the system started from a specific state e_i? Interestingly, for a primitive chain, these limiting probabilities

$$\lim_{n\to\infty} P\{X_n = e_j \,|X_0 = e_i\} = \lim_{n\to\infty} p_{ij}^{(n)}$$

can be shown to be independent of the starting state e_i and is given by the normalized left-Perron vector of the matrix P. Thus, if

$$\pi_j = \lim_{n\to\infty} P\{X_n = e_j \,|X_0 = e_i\} \tag{8}$$

then $\pi_j,\ j = 1, 2, \ldots, m$ satisfy

$$\underline{\pi} = \underline{\pi}\, P \tag{9}$$

where

$$\underline{\pi} = [\pi_1, \pi_2, \ldots, \pi_m]. \tag{10}$$

This follows quite easily form (6) since

$$P^n \to \underline{x}_0 \underline{y}_0'$$

which gives

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j, \quad \text{the } j\text{th entry of } \underline{y}_0'.$$

Hence, with $\underline{y}_0' = \underline{\pi}$, from (4) we obtain (9).

PERRON-FROBENIUS THEOREM

If A is any nonnegative irreducible matrix, i) and ii) of Perron's theorem are still true for A. However iii) need not be true. Thus if λ is any other eigenvalue of A, then we only have $|\lambda| \le \rho(A)$. In particular, there can be other eigenvalues of A such that

$$|\lambda| = \rho(A). \tag{11}$$

This situation, in fact, corresponds to periodic chains. Recall that an irreducible matrix A is periodic with period T if [20], [21]

$$a_{ii}^{(n)} = 0, \quad n \ne kT. \tag{12}$$

If A is a nonnegative irreducible periodic matrix with period T, then A has exactly T eigenvalues equal to [20]

$$\lambda_i = \rho(A)\, e^{j2\pi ij/T}, \quad i = 1, 2, \ldots, T \tag{13}$$

that are related through the Tth roots of unity, and all other eigenvalues of A are strictly less than $\rho(A)$ in magnitude. Next, we shall examine some interesting applications of Perron's theorem. A theorem by Gersgorin will be useful in this context.

GERSGORIN'S THEOREM

Every eigenvalue of an $m \times m$ matrix A lies in at least one of the discs

$$|\lambda - a_{ii}| \le P_i = \sum_{j\ne i}^{m} |a_{ij}|, \quad i = 1, 2, \ldots, m. \tag{14}$$

In other words, all eigenvalues of A lie somewhere in the union of the closed circles with centers a_{ii} and radii $P_i,\ i = 1, 2, \ldots, m$.

Gersgorin's theorem does not say that every circle in (14) will have one eigenvalue in it. It only says that every eigenvalue of A lie somewhere in the region represented by the union of the circles in (14). However, if the union of k of these circles form a connected region that is disjointed from all the remaining $m-k$ circles, then there are precisely k eigenvalues of A in that region [13] (Figure 1). It follows that when all discs in (14) are nonoverlapping, then every disc there contains an eigenvalue in it.

In particular, if A is diagonally dominant with positive diagonal entries, i.e.,

$$a_{ii} > \sum_{j\neq i}^{m} |a_{ij}| = P_i, \quad i = 1, 2, \ldots, m \tag{15}$$

then A represents a stable matrix (All eigenvalues of A are in the right half plane). This follows since

$$\begin{aligned} |\lambda - a_{ii}| &\leq \sum_{j\neq i}^{m} |a_{ij}| = P_i \\ &\Rightarrow -P_i < \operatorname{Re}\lambda - a_{ii} < P_i \\ &\Rightarrow \operatorname{Re}\lambda > 0; \end{aligned} \tag{16}$$

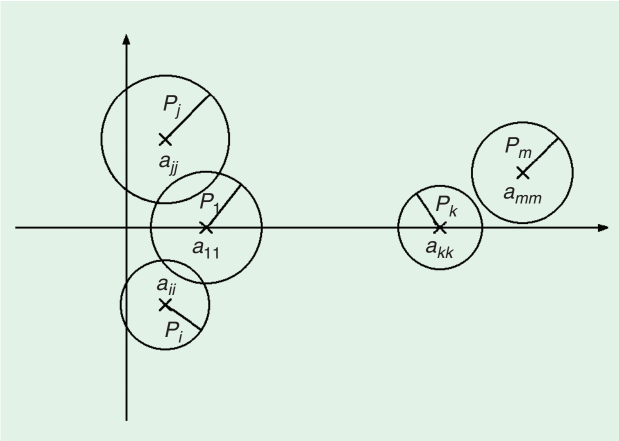
i.e., all eigenvalues of A are in the right-half plane.

Next, we examine some interesting applications of Perron's theorem. In particular, we discuss power control problems in mobile communication, a commodity pricing problem in economics, a population growth model, and finally, an application in the area of Web searching.

POWER CONTROL PROBLEM

Suppose customers at a restaurant are engaged in small talk at each table. There is crosstalk from table to table, and to compensate for that, each group can raise their voice level. But that leads to more crosstalk for some other group who might, in turn, raise their own voice levels. An interesting question in that context is an optimum strategy that allows all customers to converse while maintaining an acceptable level of crosstalk. What is the optimum relative power level for customers at each table?

Let G_{ij} refer to the path gain at the ith table from the jth table and P_j the desired voice power level to be maintained at the jth table. If γ_i represents the acceptable signal power to crosstalk power ratio (SNR) at the ith table, we need



[FIG1] Gersgorin discs.

actual SNR at the ith table

$$= \frac{P_i}{\sum_{i\neq j} G_{ij} P_j} \geq \gamma_i, \quad i = 1, 2, \ldots, m \tag{17}$$

or

$$\gamma_i \sum_{i\neq j} G_{ij} P_j \leq P_i, \quad i = 1, 2, \ldots, m. \tag{18}$$

In matrix form, (18) reads

$$\underbrace{\begin{pmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \gamma_m \end{pmatrix}}_{D} \underbrace{\begin{pmatrix} 0 & G_{1j} & \cdots & G_{1m} \\ G_{21} & 0 & \cdots & G_{2m} \\ G_{i1} & G_{ij} & \cdots & G_{im} \\ & & \ddots & \\ G_{m1} & G_{mj} & \cdots & 0 \end{pmatrix}}_{G} \underbrace{\begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_{m-1} \\ P_m \end{pmatrix}}_{\underline{P}} \leq \underbrace{\begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_{m-1} \\ P_m \end{pmatrix}}_{\underline{P}} \tag{19}$$

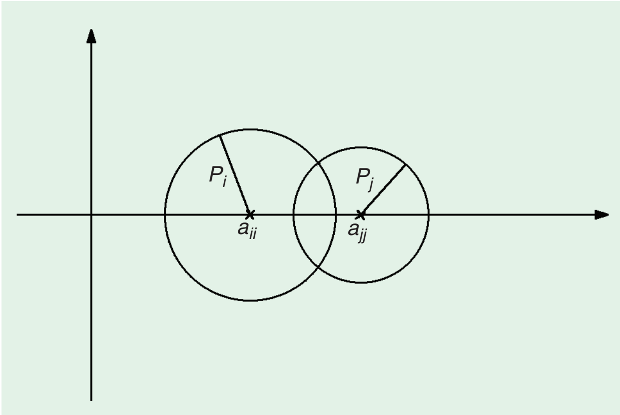
or

$$A\underline{P} \leq \underline{P} \tag{20}$$

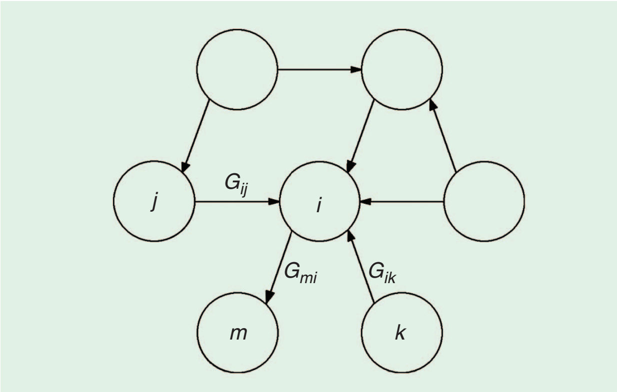
where

$$A = DG. \tag{21}$$

Unless the G_{ij}s have some specific symmetry structure to allow periodicity, A is a primitive matrix with all its diagonal entries equal to zero $(a_{ij}) = 0$. From Perron's theorem, there exist positive vector $\underline{P}$ and $\rho > 0$ such that



[FIG2] Gersgorin discs for a diagonally dominant matrix with positive diagonal entries.



[FIG3] Crosstalk.

$$A\underline{P} = \rho \underline{P}, \quad \rho > 0, \tag{22}$$

and further, all other eigenvalues of A are strictly less than ρ in magnitude. From (20) and (22), it now follows that $\underline{P}$ satisfies (20) if and only if $\rho \leq 1$ [see (31)–(34)]. From Gersgorin's theorem, all eigenvalues λ_i of A lie in union of the discs ($a_{ii} = 0$)

$$|\lambda_i - a_{ii}| = |\lambda_i| \leq \sum_{i \neq j} |a_{ij}| = \gamma_i \sum_{i \neq j} G_{ij}. \tag{23}$$

Hence,

$$\lambda_{\max}(A) = \rho \leq 1 \Rightarrow |\lambda_i| \leq 1 \Rightarrow \gamma_i \leq \frac{1}{\sum_{i \neq j} G_{ij}} \qquad i = 1, 2, \ldots, m, \tag{24}$$

is a sufficient condition, and in that case, the positive eigenvector $\underline{P}$ in (22) corresponding to the largest eigenvalue of A is the desired power vector solution up to a common scaling factor. (Scaling says that if everyone decides to lower or raise their power level by a common factor, that still leads to the same acceptable performance level as before.)

MOBILE SCENE

Interestingly, the power control problem in a mobile communication scene is quite similar to the aforementioned restaurant problem. In addition to the crosstalks, let σ_i^2 represent the ambient noise level at the base station assigned to the ith mobile.

All mobiles communicate using base stations. Suppose there are m mobiles sharing the station B_i in Figure 4. The basic power control problem is to determine the power levels P_i of each mobile so that the signal to interference plus noise ratio for each user is above a certain acceptable level [3], [10], [24], [25]. Let

$|h_{ij}|^2$: Path gain between the base station assigned to the ith mobile and the jth mobile.

γ_i : Signal to interference plus noise ratio (SINR) (desired) required from the ith mobile at its parent base station B_i.

P_i : Transmit power of the ith mobile.

This gives the signal power from the ith mobile to the interference plus noise power ratio received at the parent base station B_i to be (SINR)

$$\gamma_i \leq \frac{|h_{ii}|^2 P_i}{\sum_{j \neq i} |h_{ij}|^2 P_j + \sigma_i^2} = \frac{P_i}{\sum_{j \neq i} G_{ij} P_j + q_i}, \quad i = 1, 2, \ldots, m \tag{25}$$

or

$$\gamma_i \left(\sum_{j \neq i} G_{ij} P_j + q_i \right) \leq P_i, \quad i = 1, 2, \ldots, m \tag{26}$$

where

$$G_{ij} = \frac{|h_{ij}|^2}{|h_{ii}|^2}, \quad q_i = \frac{\sigma_i^2}{|h_{ii}|^2}. \tag{27}$$

In matrix form, (26) reads

$$\underbrace{\begin{pmatrix} \gamma_1 & 0 & \cdots & 0 & 0 \\ 0 & \gamma_2 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \cdots & 0 & \gamma_m \end{pmatrix}}_{D} \left(\underbrace{\begin{pmatrix} 0 & G_{1j} & \cdots & G_{1m} \\ G_{21} & 0 & \cdots & G_{2m} \\ G_{i1} & G_{ij} & \cdots & G_{im} \\ & & \ddots & \\ G_{m1} & G_{mj} & \cdots & 0 \end{pmatrix}}_{G} \right.$$

$$\left. \times \underbrace{\begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_{m-1} \\ P_m \end{pmatrix}}_{\underline{P}} + \underbrace{\begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_{m-1} \\ q_m \end{pmatrix}}_{\underline{q}} \right) \leq \underbrace{\begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_{m-1} \\ P_m \end{pmatrix}}_{\underline{P}} \tag{28}$$

or

$$(I - DG)\, \underline{P} \geq D\underline{q}. \tag{29}$$

Let

$$A = DG \tag{30}$$

so that (28)–(29) reads

$$(I - A)\, \underline{P} \geq D\underline{q} \triangleq \underline{b}. \tag{31}$$

If the path gains do not have any specific structure (that avoids periodicity and reducibility), then A is a nonnegative primitive matrix with all diagonal entries equal to zero. The necessary and sufficient condition for (31) to have a positive solution $\underline{P}$ for every positive vector $\underline{b}$ is that $(I-A)^{-1}$ be nonnegative for $A \geq 0$. However, for any $A \geq 0$,

$$(I-A)^{-1} \geq 0 \quad \text{iff} \quad \rho(A) = |\lambda_{\max}(A)| < 1. \tag{32}$$

Thus, from (32), the necessary and sufficient condition for (31) to have a positive solution $\underline{P}$ is that the spectral radius of A be strictly bounded by unity.

PROOF

Suppose $\rho(A) < 1$. Then $A^k \to 0$ and $(I-A)^{-1} = \sum_{k=0}^{\infty} A^k \geq 0$ converges. Hence, the unique solution given

$$\underline{P} = (I-A)^{-1}\underline{b} = \sum_{k=0}^{\infty} A^k \underline{b} > 0 \tag{33}$$

is positive for every positive vector $\underline{b}$ in (31). Therefore, the condition in (32) in terms of the spectral radius of A is sufficient. To prove its necessity, suppose $A \geq 0$ and $(I-A)^{-1} \geq 0$, as in (32). Let λ and $\underline{x}$ represent any set of eigenvalue-eigenvector pair for A. Then

$$\lambda \underline{x} = A\underline{x} \Rightarrow |\lambda||\underline{x}| = |A\underline{x}| \leq A|\underline{x}|, \quad \text{since } A \geq 0$$

or (here, $|\underline{x}|$ represents entrywise absolute value of $\underline{x}$)

$$(I-A)|\underline{x}| \leq (1-|\lambda|)\,|\underline{x}|.$$

Hence,

$$|\underline{x}| \leq (1-|\lambda|)\,(I-A)^{-1}\,|\mathrm{x}| > 0, \quad \text{since } (I-A)^{-1} \geq 0.$$

But on the left side, $|\underline{x}| \geq 0$. Therefore, we must have $|\lambda| < 1$ on the right side for all λ, which gives

$$\rho(A) < 1 \quad \text{if} \quad (I-A)^{-1} \geq 0$$

thus proving the necessity of the spectral radius condition in (32).

As in (24), if we have

$$\gamma_i < \frac{1}{\sum\limits_{i \neq j} G_{ij}} = \frac{|h_{ii}|^2}{\sum\limits_{i \neq j} |h_{ij}|^2}, \qquad i = 1, 2, \ldots, m, \tag{34}$$

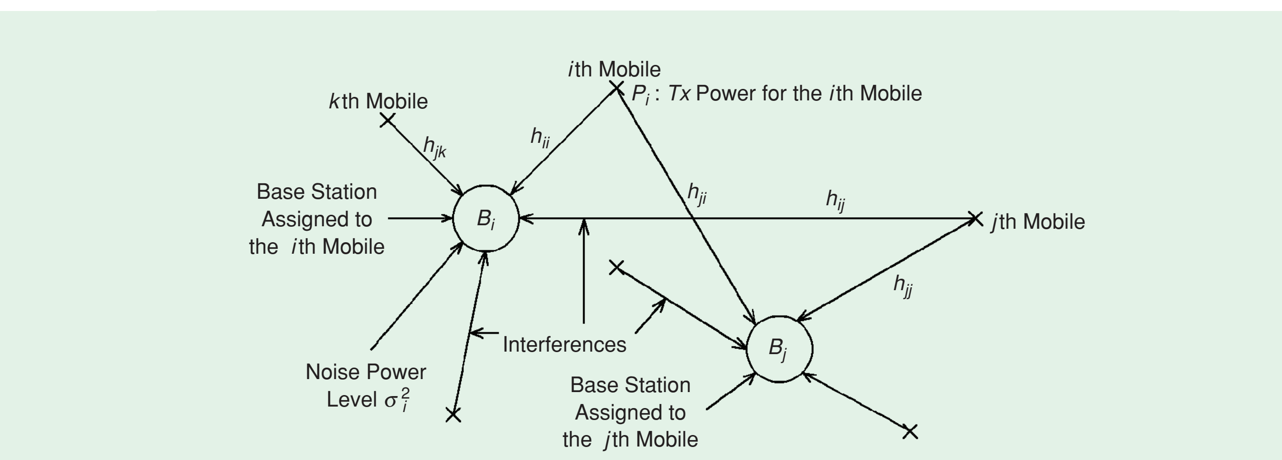
then the spectral radius of A is strictly less than unity and, from (31),

$$\underline{P} = (I-A)^{-1} D\underline{q} = \sum_{k=0}^{\infty} A^k D\underline{q} > 0 \tag{35}$$

represents the unique positive power vector that satisfies (29). However, (34) represents only a sufficient condition on the desired SINRs γ_i, $i = 1, 2, \ldots, m$ for a feasible solution, and it may be too restrictive. Unlike the solution for $\underline{P}$ in (22), the presence of noise makes the solution in (33) unique. If (32) is not satisfied, one possibility is to drop some users so that $\lambda_{\max}(A) < 1$. Another possibility is to swap the assigned base stations for some user so that the matrix G is redefined in (28) and the new path gains satisfy (32).

What if there is more noise at the ith receiver or the path gain $|h_{ii}|^2$ decreases such that G_{ij} are the same? In either case, from (27), q_i increases, affecting the power level structure. To analyze this situation, let Δ_i denote the increment in q_i and $\hat{P}_i$, $i = 1, 2, \ldots, m$ the new optimum power levels. From (35), this gives

$$\underline{\hat{P}} = (I-A)^{-1} D(\underline{q} + \Delta_i \underline{e}_i) \tag{36}$$



[FIG4] Mobile communication scene.

where

$$\underline{e}_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)' \tag{37}$$

with 1 at the ith location. From (37)

$$\begin{aligned}\underline{\hat{P}} &= (I - A)^{-1} D\underline{q} + \Delta_i \gamma_i (I - A)^{-1} \underline{e}_i \\ &= \underline{P} + \Delta_i \gamma_i \underline{M}_i\end{aligned} \tag{38}$$

where $\underline{M}_i$ denotes the ith column of $(I - A)^{-1}$. Thus

$$\hat{P}_k = P_k + \Delta_i \gamma_i (I - A)^{-1}_{ki}, \tag{39}$$

implying that an increase in noise level at any one user results in increase of the power levels for all users. However, the power level of the ith user increases by the highest amount. This follows from the following result: If A in (30) satisfies the additional property that every row sum is strictly less than unity, then [2]

$$(I - A)^{-1}_{ii} > (I - A)^{-1}_{ki}, \quad i \neq k. \tag{40}$$

In summary, any additional disturbance for one user results in power increases for all users, with the highest increase occurring for the user directly affected by the disturbance. If every row sum of A is less than or equal to unity, then the strict inequality in (40) is replaced by an inequality, i.e.,

$$(I - A)^{-1}_{ii} \geq (I - A)^{-1}_{ki}, \tag{41}$$

implying that, with additional disturbance for any one (ith) user, none of the power levels can decrease, and the power level of the ith user increases by the greatest amount, although other power levels can increase by the same amount.

COMMODITY PRICING (LEONTIEF MODEL)

Consider a closed group of n industries, each of which produces one commodity and requires inputs from all other industries, including itself. Let a_{ij} represent the fraction of the jth industry commodity purchased by the ith industry. Then

$$a_{ij} \geq 0, \quad \sum_{i=1}^{n} a_{ij} = 1, \quad j = 1, 2, \ldots, n. \tag{42}$$

That is, A is a nonnegative matrix with each column sum equal to unity, implying that each industry disposes its commodity among all industries in the group. From (3), the spectral radius of A equals unity. In the simplified model, the problem is to determine a "fair price" to be charged for each commodity output so that the total expenses equal the total income for each industry. Let P_i represent the total ith commodity price to be determined. Then, since the ith industry needs a_{ij} fraction of the jth industry, it costs $a_{ij} P_j$ for that portion; and so on. Therefore, the

$$\left.\begin{array}{c}\text{total expenses incurred} \\ \text{by the } i\text{th industry}\end{array}\right\} = \sum_{j=1}^{n} a_{ij} P_j, \tag{43}$$

and this must equal the total income P_i. This gives the set of equations

$$\begin{aligned}\sum_{j=1}^{n} a_{ij} P_j &= P_i, \quad i = 1, 2, \ldots, n \\ \underline{P} &= [P_1, P_2, \ldots, P_n]',\end{aligned} \tag{44}$$

or the price structure must satisfy the equation

$$A\underline{P} = \underline{P}. \tag{45}$$

From (3), we have $\rho(A) = 1$, and from Perron's theorem, the eigenvector $\underline{P}$, mentioned previously, has a unique positive solution given by the right-Perron vector of A corresponding to the eigenvalue unity, and that dictates the price structure.

PROFIT?

If profits are to be introduced in this model, then since

$$\text{Profit} = \text{Total income} - \text{Total expenses}, \tag{46}$$

from (44), we have

$$\gamma_i = P_i - \sum_{j=1}^{n} a_{ij} P_j, \quad i = 1, 2, \ldots, n \tag{47}$$

where γ_i represents the total profit for the ith industry. In matrix form, (47) reads

$$(I - A)\,\underline{P} = \begin{pmatrix}\gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n\end{pmatrix} \stackrel{\Delta}{=} \underline{\gamma}. \tag{48}$$

Notice that this situation is similar to that in (31)–(35) for the mobile power control problem. From (32), the necessary and sufficient condition for (48) to have a positive solution for $\underline{P}$ is that the spectral radius of A be strictly bounded by unity. This situation is unlike the equal income-cost structure in (45), where $\rho(A) = 1$ is necessary and sufficient. In the present case, to sustain a profit structure, we must have $\rho(A) < 1$. Clearly, it follows that the normalization condition in (42) should not be maintained for all columns, and we must have

$$\sum_{i=1}^{n} a_{ij} < 1 \tag{49}$$

for at least one column of A so that $\rho(A) < 1$, implying that one or more industries must overproduce (and sell the excess commodity to external consumers) to maintain a profit structure.

It is interesting that to maintain a profit structure for all industries, it is not necessary that every one of them should sell their commodities to external consumers. Almost all of them can be service industries; however, at least one industry must go outside the support loop and sell their excess product in order for all to make a profit.

In this context, one interesting question is given n dependent industries; what is the best strategy for maximizing the overall profit? Is it better for each industry to be in the support mode as well as the selling mode, or is it more efficient for some industries to be totally in the support mode and others to be in the mixed mode, assuming that there is demand for each or some commodity?

MORE PROFIT?

Suppose the ith industry alone decides to increase its profit from γ_i to $\gamma_i + \Delta$. What happens to the price structure?

This situation is similar to the power control problem in (38)–(39). As before, let

$$\underline{\hat{P}} = \left(\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_n\right)' \tag{50}$$

denote the new prices of the n commodities. Then from (48),

$$\begin{aligned}\underline{\hat{P}} &= (I - A)^{-1} (\underline{\gamma} + \Delta \underline{e}_i) \\ &= \underline{P} + \Delta (I - A)^{-1} \underline{e}_i = \underline{P} + \Delta \underline{M}_i\end{aligned} \tag{51}$$

where $\underline{e}_i = [0, 0, \ldots, 0, 1, 0, \ldots, 0]'$ with 1 at the ith location, and $\underline{M}_i$ denotes the ith column of $(I - A)^{-1}$. Thus

$$\hat{P}_k = P_k + \Delta (I - A)^{-1}_{ki}, \tag{52}$$

implying that the price structure increase for all industries. However, from (40), the price of the ith commodity increases by the highest amount, since in this case [see also (41)],

$$(I - A)^{-1}_{ii} > (I - A)^{-1}_{ki}, \ \ i \neq k.$$

In a socioeconomic context, this result has an interesting interpretation as well. Suppose a family of n members depend on each other for support to various degrees. Then part of the efforts of every family member goes to support other members and part is spent on activities involving self interest. In this context, an interesting question in terms of maximizing profitability for each person is whether it is necessary for everyone to work outside the family loop, or whether some can be in a totally supportive mode. The above analysis states that in order to have profit for everyone, it is not necessary that all must work outside the family loop. From (50)–(52), to maximize profit, perhaps the best skilled person (the one whose profit is largest) must go outside the family loop; all others can be totally in the support mode and still maintain the desired profit. (The family situation is more complicated because of psychological aspects of playing a secondary roll and related ego issues.)

Another way to introduce a profit model is to maintain that a fixed fraction of the income equals profit. From (46)–(47), this gives for some $\epsilon > 0$

$$\text{Profit} = P_i - \sum_{j=1}^{n} a_{ij} P_j = \epsilon P_i, \quad i = 1, 2, \ldots, n \tag{53}$$

or

$$A \underline{P} = (1 - \epsilon)\, \underline{P} = \lambda\, \underline{P}. \tag{54}$$

To maintain a nonzero profit, we must have $\lambda < 1$ so that $\epsilon = 1 - \lambda > 0$, and once again from Perron's theorem, it follows that we must have $\rho(A) < 1$. The right Perron vector of A gives the desired pricing structure. Notice that the profit model in (53)–(54) is fair in a global sense since the percentage profit margin is the same for all industries. Interestingly, that margin cannot be preassigned by each industry, and it is determined by the spectral radius of A. The model in (48), on the other hand, allows the highly desirable situation where the profit is preassigned by each industry, and if $\rho(A) < 1$, then the unique pricing vector in that case is given by

$$\underline{P} = (I - A)^{-1}\, \underline{\gamma} = \sum_{k=0}^{\infty} A^k\, \underline{\gamma} > 0. \tag{55}$$

POPULATION GROWTH MODELS (LESLIE MODEL)

The Leslie model describes the growth of the female population in any closed society (humans or animals) by classifying them into n equal age groups $e_1, e_2, \ldots, e_n$, where e_i represents the ith age group

$$e_i = \{(i - 1)M/n,\ iM/n\}, \quad i = 1, 2, \ldots, n \tag{56}$$

with M representing the life span of the population. The age distribution changes over time because of birth, death, and aging. Let

$$a_i = \begin{cases}\text{average number of daughters born to a single} \\ \text{female in the } i\text{th group, } i = 1, 2, \ldots, n\end{cases} \tag{57}$$

and

$$b_i = \begin{cases} \text{the percentage of females in the } i\text{th group that are} \\ \text{expected to pass into the } (i+1)\text{ th group. } (0 < b_i \le 1) \end{cases} \tag{58}$$

Let

$$\underline{p}_m = \left[p_1^{(m)}, p_2^{(m)}, \ldots, p_n^{(m)}\right]' \tag{59}$$

represent the age distribution of females at time t_m, where $p_i^{(m)}$ is the number of females in the ith group at t_m. From (57)–(59), we get

$$p_1^{(m)} = \sum_{i=1}^{n} a_i\, p_i^{(m-1)} \tag{60}$$

and

$$p_i^{(m)} = b_{i-1}\, p_{i-1}^{(m-1)}, \quad i = 2, 3, \ldots, n. \tag{61}$$

In matrix form, (60)–(61) can be written as

$$\underline{p}_m = L\, \underline{p}_{m-1} = L^m\, \underline{p}_0 \tag{62}$$

where

$$L = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{n-1} & a_n \\ b_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & b_2 & 0 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & b_{n-1} & 0 \end{pmatrix} \tag{63}$$

represents the Leslie matrix. The matrix L is nonnegative and hence the asymptotic behavior of the age-vector $\underline{p}_m$ in (62) is governed by only the left-and right-Perron vectors of L together with the spectral radius $\rho(L)$ of L. Since $\rho(L)$ is the largest eigenvalue of L, we can use the characteristic polynomial of L to determine $\rho(L)$. By direct computation the characteristic polynomial of L is given by

$$|\lambda I - L| = \begin{vmatrix} \lambda - a_1 & -a_2 & -a_3 & \cdots & -a_{n-1} & -a_n \\ -b_1 & \lambda & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \cdots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \cdots & -b_{n-1} & \lambda \end{vmatrix}$$

$$= (\lambda - a_1)\,\lambda^{n-1} + b_1 \begin{vmatrix} -a_2 & -a_3 & \cdots & -a_{n-1} & -a_n \\ -b_2 & \lambda & \cdots & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & -b_{n-1} & \lambda \end{vmatrix}$$

$$= \lambda^{n-1} - a_1\,\lambda^{n-1} + b_1 \times \left\{ -a_2\,\lambda^{n-2} + b_2 \begin{vmatrix} -a_3 & -a_4 & \cdots & -a_{n-1} & -a_n \\ -b_3 & \lambda & \cdots & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & -b_{n-1} & \lambda \end{vmatrix} \right\}$$

$$\begin{aligned} &= \lambda^n - a_1\lambda^{n-1} - a_2 b_1 \lambda^{n-2} - a_3 b_2 b_1 \lambda^{n-3} \\ &\quad + \cdots - a_n b_1 b_2 \cdots b_{n-1}. \end{aligned} \tag{64}$$

Since all a_is cannot be zeros, (64) has at least one positive root. From Perron's theorem, the largest positive root of (64) represents $\rho(L)$. By direct substitution, it is easy to verify that the right-Perron vector $\underline{x}_0$ of L is given by

$$\underline{x}_0 = [1,\ b_1/\rho(L),\ b_1 b_2/\rho^2(L),\ b_1 b_2 \cdots b_{n-1}/\rho^{n-1}(L)]' \tag{65}$$

[verify that $L\,\underline{x}_0 = \rho(L)\,\underline{x}_0$]. Let $\underline{y}_0 = [y_1, y_2, \ldots, y_n]'$ represent the left Perron vector of L. Then $\underline{y}_0'\,\rho(L) = \underline{y}_0'\,L$ gives

$$[y_1, y_2, \ldots, y_n] = \lambda_0^{-1}\,[y_1, y_2, \ldots, y_n]L, \tag{66}$$

where $\lambda_0 = \rho(L)$. Expanding (66), we obtain

$$\begin{aligned} y_n &= \lambda_0^{-1}\, a_n \\ y_{n-1} &= \lambda_0^{-1}(y_1 a_{n-1} + y_n b_{n-1}) \\ &= \lambda_0^{-1}\, y_1 \left(a_{n-1} + a_n b_{n-1} \lambda_0^{-1}\right) \\ y_{n-2} &= \lambda_0^{-1}(y_1 a_{n-2} + y_{n-1} b_{n-2}) \\ &= \lambda_0^{-1}\, y_1 \left(a_{n-2} + a_{n-1} b_{n-2} \lambda_0^{-1} + a_n b_{n-1} b_{n-2} \lambda_0^{-2}\right) \\ &\ \vdots \\ y_i &= \lambda_0^{-1}(y_1 a_i + y_{i+1} b_i) \\ &\ \vdots \\ y_1 &= \lambda_0^{-1}(y_1 a_1 + y_2 b_1) \\ &= \lambda_0^{-1}\, y_1 \left(a_1 + a_2 b_1 \lambda_0^{-1} + a_3 b_1 b_2 \lambda_0^{-2} \right. \\ &\quad \left. + \cdots + a_n b_1 b_2 \cdots b_{n-1} \lambda_0^{-(n-1)}\right) \\ &= y_1 \end{aligned}$$

since from (64), $a_1\lambda_0^{-1} + a_2 b_1 \lambda_0^{-2} + \cdots + a_n b_1 b_2 \cdots b_{n-1}\,\lambda_0^{n-1} = 1$. Therefore, we may select

$$
\begin{aligned}
y_1 &= 1 & (67)\\
y_2 &= \lambda_0^{-1}\Big(a_2 + a_3 b - 2\,\lambda_0^{-1} + a_4 b_2 b_3 \lambda_0^{-2} \\
&\qquad + \cdots + a_n b_2 b_3 \cdots b_{n-1}\,\lambda_0^{-(n-2)}\Big) & (68)\\
y_3 &= \lambda_0^{-1}\Big(a_3 + a_4 b_3\,\lambda_0^{-1} + a_5 b_3 b_4\,\lambda_0^{-2} \\
&\qquad + \cdots + a_n b_3 b_4 \cdots b_{n-1}\,\lambda_0^{(n-3)}\Big) & (69)\\
&\ \ \vdots \\
y_{n-2} &= \lambda_0^{-1}\Big(a_{n-2} + a_{n-1} b_{n-2}\,\lambda_0^{-1} + a_n b_{n-2} b_{n-1}\,\lambda_0^{-2}\Big) & (70)\\
y_{n-1} &= \lambda_0^{-1}\Big(a_{n-1} + a_n b_{n-1}\,\lambda_0^{-1}\Big) & (71)\\
y_n &= \lambda_0^{-1}\,a_n, & (72)
\end{aligned}
$$

which gives the left-Perron vector of L. The normalization condition $\underline{y}_0'\,\underline{x}_0 = 1$ gives $\underline{y}_0/\underline{y}_0'\,\underline{x}_0$ to be the normalized left-Perron vector of L. This gives the spectral representation [13]

$$L = \rho(L)\,\frac{\underline{x}_0\,\underline{y}_0'}{\underline{y}_0'\,\underline{x}_0} + \sum_{k=2}^{n} \lambda_k\,\underline{x}_k\,\underline{y}_k', \quad |\lambda_k| < \rho(L) \qquad (73)$$

and hence [see (6)]

$$L^m = \rho^m(L)\,\frac{\underline{x}_0\,\underline{y}_0'}{\underline{y}_0'\,\underline{x}_0} + \sum_{k=2}^{n} \lambda_k^m\,\underline{x}_k\,\underline{y}_k'. \qquad (74)$$

It follows that if $|\lambda_k| < 1,\ k = 2, 3, \ldots, n$, then

$$L^m \to \rho^m(L)\,\frac{\underline{x}_0\,\underline{y}_0'}{\underline{y}_0'\,\underline{x}_0} \qquad (75)$$

and hence

$$\underline{p}_m = L^m\,\underline{p}_0 \to \rho^m(L)\,\frac{\underline{x}_0\,\underline{y}_0'\,\underline{p}_0}{\underline{y}_0'\,\underline{x}_0} = \alpha\,\rho^m(L)\,\underline{x}_0 \qquad (76)$$

where

$$\alpha = \frac{\underline{y}_0'\,\underline{p}_0}{\underline{y}_0'\,\underline{x}_0} \qquad (77)$$

is a constant that depends on the initial population distribution $\underline{p}_0$. From (76), in the long run, the relative distribution at any time depends only on the right-Perron vector $\underline{x}_0$ of L. From (76), it also follows that

$$\underline{p}_m = \rho(L)\,\underline{p}_{m-1}, \qquad (78)$$

i.e., in the long run, the age distribution is a scalar multiple of the previous age distribution. From (76), if the scalar multiple $\rho(L) > 1$, the population model tends to explode as time goes on; and if $\rho(L) < 1$ the population becomes extinct in the long run. Therefore, in the long run, there are only two stable (absorbing) states for any population. Controlling $\rho(L)$ for a stable population is an interesting problem and a tricky issue. From (74), to maintain (75)–(78), we must have

$$|\lambda_k| < 1, \quad k = 2, 3, \ldots, n. \qquad (79)$$

A sufficient set of conditions to maintain (79) can be once again derived using Gersgorin's theorem in (14). Suppose the first row in (63) satisfies the condition

$$a_1 > \sum_{k=2}^{n} a_k + 1 \triangleq P_1 + 1. \qquad (80)$$

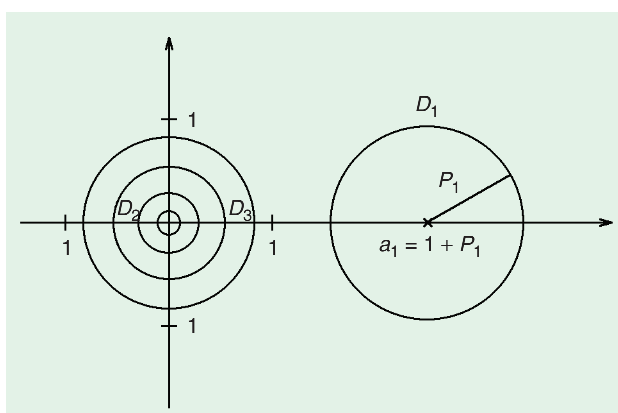
and let

$$0 < b_i < 1.$$

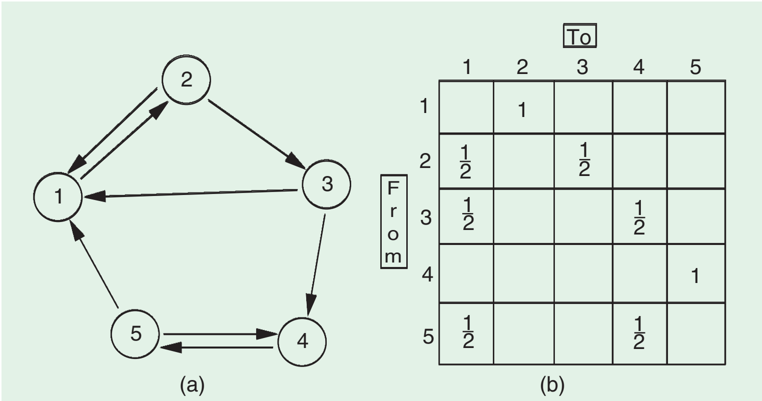
Then the corresponding Gersgorin discs are as shown in Figure 5.

Notice that there are two nonoverlapping sets of discs with one set containing only one disc with center at a_1 and radius equal to P_1. All other discs are concentric circles centered at the origin with radius b_i less than unity. It follows that the disc D_1 centered at a_1 contains exactly one eigenvalue λ_1 of L with $|\lambda_1| > 1$, and the remaining disjointed discs contain the remaining eigenvalues $\lambda_k,\ k = 2, 3, \ldots, n$ of L. Since the disjointed set of discs centered around the origin are bound by the unit circle, we have $|\lambda_k| < 1,\ k = 2, 3, \ldots, n$ as required in (79). From Perron's theorem, since $\lambda_1 > 0$ is the largest eigenvalue of L, we have $\lambda_1 = \rho(L) > 1$ because of (80).

The condition (80) states that if the youngest generation outperforms the rest of the population in terms of the reproduction rate, then the population explodes; a condition that is practically impossible to be met by the humans irrespective of any arbitrary number of age partitioning scheme in (56). Interestingly, in the animal/insect kingdom, it appears that (80) is often met without much difficulty, perhaps due to the lack of severe social structure and responsibilities.



[FIG5] Gersgorin discs when the youngest generation outperforms the rest of the population.



[FIG6] **Example of a Web graph with five pages and the corresponding 5 × 5 matrix.**

PAGE RANKING: HOW DOES GOOGLE DO IT?

Suppose a user types a query into the Google Web search engine to perform a search. Most of the time, there are thousands of results for the given query, and the primary challenge for a search engine is to return these results to the user in an appropriate ordering, so that the best pages are returned first and listed at the top. In the case of the Google engine [4], this is done with the help of a global rank computation called "Pagerank" that can be modeled by an iterative process defined by an irreducible matrix. In the following, we relate this process to Perron's theorem and also give a more detailed discussion of some practical issues in implementing this computation.

SOME BACKGROUND ON SEARCH ENGINES

Most details about Google and other search engines are closely guarded trade secrets, but the following basic description suffices for our purpose. Most major search engines work by periodically or continually downloading (crawling) large numbers of pages from the Web, and thus at any point in time, the engine has a slightly outdated and incomplete snapshot of the entire Web stored on its disks. Given a query, a search engine then assigns to each page it knows a "score," which is a real value that measures the relevance of the page with respect to a given query, and then returns the pages ordered from highest to lowest score. Many different scoring functions for pages and other text documents have been proposed that take into account various factors, e.g., how many times the query words occur in the document, how common or rare the words are in the overall collection, whether the query words occur close to each other in the document, or are in a title or subheading, to name just a few. Such functions and efficient ways to evaluate them without looking at all pages for each query have been studied extensively over the last 30 years in the field of information retrieval [1]. However, modern search engines rely on an important additional ingredient for good results; the hyperlink structure of the Web. One particularly powerful method, Pagerank, was developed by the Google engine and is described in the following.

THE IDEA BEHIND PAGERANK

Recall that pages on the Web usually contain a number of hyperlinks (or links) to other pages, and therefore, the Web can be seen as a giant graph of pages (nodes) connected by edges (links). Fairly early in the development of Web search engines, the idea of using this link structure to improve the quality of the ranking was developed. For example, it might seem like a good idea to boost the scores of pages that many other pages link to, since each of the authors of those other pages made the human judgment that the page is worth linking to. However, one can do better than this naive approach, based on ideas earlier studied in the context of citation analysis [9] and social network analysis [15]. For example, to determine influential publications in the scientific literature, we would not look just at the number of citations a paper receives, but also at the importance of the citing paper. Similarly, we know that influence in a social network depends not just on how many people you know but how influential those people are.

This is the idea underlying the ranking of pages by the Google search engine. The goal is to compute for each Web page an absolute rank value (measure of importance or quality) based on the link structure of the Web, called the "Pagerank value" of the page. The actual score of a page for a given query is then determined by combining (e.g., adding after some normalization) this query-independent Pagerank value with a query-dependent score based on the content of the page as mentioned earlier, and then the top ten pages, according to the total score, are returned to the user. In Google, the goal is to assign a Pagerank score $s(p)$ to each page p that is proportional to the sum of the degree-scaled importances of the pages linking to it, i.e.,

$$s(p) = \sum_{q \to p} \frac{s(q)}{d(q)},$$

where $d(q)$ is the out-degree of page q, i.e., the number of hyperlinks on page q. In matrix notation, we are interested in the Pagerank vector $\underline{s}$ that is the solution to the equation

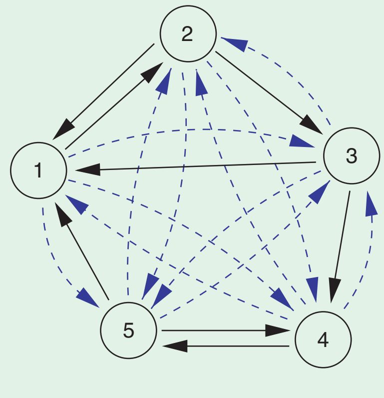
$$\underline{s}L = \underline{s},$$

where

$$(l_{ij}) = \begin{cases} 1/d(p_i), & \text{if there is a hyperlink from } p_i \text{ to } p_j, \text{ and} \\ 0, & \text{otherwise.} \end{cases}$$

Figure 6 shows an example of a small Web graph and the corresponding matrix. Before going more into the structure of the previous matrix, we consider the problem from the perspective of random walks. In particular, note that the solution $\underline{s}$ (if it exists) corresponds to the steady distribution of a random walk on the Web graph where we start out at an arbitrary page and then choose one of the outgoing links at random to go to the next page. Therefore, the Pagerank score corresponds to the likelihood of being at this page under the random walk, and pages with high in-degree or with links from other often visited pages will tend to have a high score. This view of the problem is also referred to as the "random surfer model."

Using this view, we can observe some properties of the problem. First, there are many pages on the Web that have ingoing but no outgoing hyperlinks; such pages are also called "leaks." Moreover, since search engines can only explore (crawl) a subset of the entire Web, some pages with outgoing links to unexplored pages will appears as leaks in the data set. Second, there are groups of pages, called "sinks," where each page only has hyperlinks to other pages in the group, plus some incoming hyperlinks from outside the group. (Some very large commercial Web sites do not have any hyperlinks going outside the site as a matter of policy.) In general, the Web is not strongly connected at all, but consists of a large, strongly connected core of pages plus many unconnected or weakly connected subsets of pages [5]. These properties imply that the matrix L in the formulation is not primitive, and in fact not even irreducible, and that the random walk has no steady distribution.



(a)

To

From	1	2	3	4	5
1	$\frac{b}{5}$	$1-b+\frac{b}{5}$	$\frac{b}{5}$	$\frac{b}{5}$	$\frac{b}{5}$
2	$\frac{1-b}{2}+\frac{b}{5}$	$\frac{b}{5}$	$\frac{1-b}{2}+\frac{b}{5}$	$\frac{b}{5}$	$\frac{b}{5}$
3	$\frac{1-b}{2}+\frac{b}{5}$	$\frac{b}{5}$	$\frac{b}{5}$	$\frac{1-b}{2}+\frac{b}{5}$	$\frac{b}{5}$
4	$\frac{b}{5}$	$\frac{b}{5}$	$\frac{b}{5}$	$\frac{b}{5}$	$1-b+\frac{b}{5}$
5	$\frac{1-b}{2}+\frac{b}{5}$	$\frac{b}{5}$	$\frac{b}{5}$	$\frac{1-b}{2}+\frac{b}{5}$	$\frac{b}{5}$

(b)

[FIG7] Example graph with additional links for random jumps with jump parameter $\beta = b$ and the corresponding matrix.

A MODIFIED APPROACH

To deal with this problem, the graph is usually pruned by repeatedly removing all nodes that do not have any outgoing edges, until we are left with a collection of strongly connected components. In addition, the random walk is modified as follows: we select an additional parameter β, say $\beta = 0.15$, and in each step of the random walk, we follow a random outgoing link with probability $(1-\beta)$, and jump to a random page with probability β. In matrix notation, the process becomes

$$\underline{s}M = \underline{s}, \tag{81}$$

where

$$(m_{ij}) = \begin{cases} (1-\beta)/d(p_i) + \beta/N & \text{if there is a hyperlink from } p_i \text{ to } p_j, \text{ and} \\ \beta/N & \text{otherwise,} \end{cases}$$

where N is the total number of pages. Note that the modified matrix is primitive and that

$$\sum_{j=0}^{N} m_{ij} = 1$$

for all i. Therefore, Perron's theorem implies the existence of an Eigenvalue $\rho = 1$ and a solution $\underline{s}$. This solution is commonly computed by a simple iterative process that repeatedly multiplies an arbitrary start vector by M until convergence. We choose a start vector $\underline{s}_0$ with $||\underline{s}_0|| = 1$, by assigning an initial value of $1/N$ to each element. Note that (81) can be expressed as

$$\underline{s}M = (1-\beta)\underline{s}L + \underline{\beta} = \underline{s}, \tag{82}$$

where $\underline{\beta}$ is the vector with every entry equal to β/N. In other words, one iteration of the process can be computed by a multiplication with the sparse matrix L followed by a vector addition. More details on the efficiency of computation will be given later.

Figure 7 shows our example graph and its corresponding matrix M. Apart from making the transition matrix primitive, the introduction of the random jump with probability β has several other interesting aspects. It has been argued that the modified process is a better model for actual user surfing behavior, where a user can either follow an outgoing link from the current page or jump to a completely unrelated page using his bookmarks or a search engine. The process is also known to make the ranking more robust against noise in the input data arising from the fact that the input graph is only a subset of the entire Web [19] and that pages (and links) may be missed during a crawl for a variety of reasons. Finally, a value for β around 0.15 limits the impact of pages to a small neighborhood of pages that are often related in topic to the page and speeds up convergence of the iterative process. The precise role of the value of β and the random jump in Pagerank are still not completely understood.

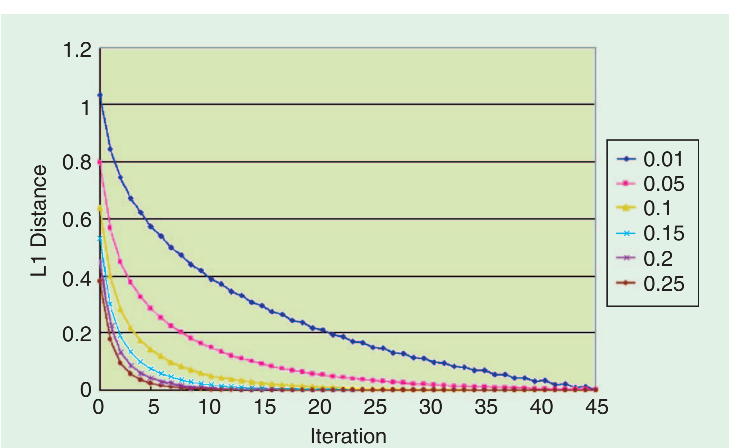
COMPUTING PAGERANK IN PRACTICE

We now present some experimental results that we computed on a actual Web graph of significant size. As mentioned before, the input matrix for the Pagerank computation is obtained by periodically performing a "Web crawl" where a special software tool collects a larger number of pages by starting at a small set of pages and following outgoing hyperlinks in a breadth-first fashion, i.e., by visiting pages in order of their distance from the starting pages. For our experiments, we used the PolyBot research Web crawler [23] to visit a total of about 120 million

Web pages with a total of about 1.8 TB of HTML and text data. There are many system design and administration issues in such large scale runs that are outside the scope of this article; see [23] for details. Clearly, the visited pages, including a number of pages visited several times, are only a small fraction of the more than 3 billion pages available on the Web, and we are only able to visit pages that are reachable from the starting pages. As observed in [18], most pages with high Pagerank value will be visited during the first few million visits of a breadth-first crawl, and by running Pagerank on a subgraph of significant size we obtain a reasonably precise ordering of the pages with high Pagerank.

We extracted the link structure from the crawled data, and then performed repeated pruning of the graph to remove leak nodes. This resulted in a graph with 44.8 million nodes and 665.9 million hyperlinks. Using $\beta = 0.15$ as well as several other values of β, we computed Pagerank values for this subgraph by running the iterative procedure implied by (82) for 50 iterations, at which point the procedure had converged reasonably well. From Figure 8, we see that larger random jump parameters β, such as 0.15 (Google) or 0.25, result in significantly faster convergence of the Pagerank procedure.

Some remarks concerning the efficiency of the computation: a naive implementation based on repeatedly multiplying a matrix with 45 million rows and columns would be extremely inefficient. As mentioned above, we can reduce this to a multiplication by the sparse link matrix L followed by an addition of the value $\_/N$ to each elements of the resulting vector. However, even with a sparse matrix or graph adjacency list representation, a graph with 44.8 million nodes and 665.9 million edges will usually not fit into the main memory of current computer workstations. This requires special out-of-core techniques that perform efficient computation based on repeated scans of disk-resident data; techniques for this problem are described in [11], [6]. After preprocessing to extract the graph from the crawled data, the actual Pagerank iteration takes a few minutes per iteration on a typical workstation.



[FIG8] Convergence of the Pagerank algorithm over 50 iterations on a large Web graph, for different values of the random jump parameter β. The horizontal axis plots the iterations, and the vertical axis shows the L_1-distance between the current Pagerank vector and the vector of the final converged solution.

ADDITIONAL ISSUES

Up to this point, we have only described the most basic version of the Pagerank technique, and there have been a number of attempts to modify and generalize Pagerank. We discuss two issues in this context, search engine manipulation and search personalization.

SEARCH ENGINE MANIPULATION

Due to the importance of search engines as a method for locating relevant Web pages, it is crucial for commercial Web sites to be listed close to the top on those queries that potential customers might use to find the site. E.g., for an online bookstore, it is highly desirable to be ranked among the top ten results on the query "books," since most search engine users will only look at the first page of results that are returned. Because of this, Web sites make considerable efforts to optimize their sites, by adding appropriate keywords and hyperlinks, so that they are ranked high on the most popular search engines. This process is known as "search engine optimization," "search engine manipulation," or "search engine spam" (not to be confused with the e-mail spam more commonly discussed in mass media) when used very aggressively, and there are a large number of companies and consultants that provide such optimizations as service.

Given the popularity of the Google engine, there have been many attempts to manipulate the Pagerank technique by creating sets of pages that link to each other in ways that increase the Pagerank score of a particular page. Search engines on the other hand are interested in identifying such attempts and deleting such nepotistic links [7], or even penalizing sites involved in such behavior. Note that similar issues also arise in the off-line world; e.g., in the scientific literature a small clique of scientists could conspire to increase their impact under common citation metrics by aggressively citing each others work, and this is one reason for omitting self-citations when looking at scientific impact.

In Pagerank, the random jump provides one opportunity for such manipulation, since this means that every page has a Pagerank of at least 0.15. Thus, we could create a large number of pages that are used to collect Pagerank that is then routed via hyperlinks to certain target pages to increase their Pagerank. In fact, many sites employ multiple "doorway pages" that point to a site, and sites are carefully designed to achieve maximum Pagerank for certain important pages. Other entrepreneurs build large sets of pages filled with random junk, say words chosen at random from a set of relevant terms [8], to collect Pagerank that can be passed on the other pages. Moreover, search engine optimization companies create large "link farms" involving many sites that agree to link to each other even though there is not relationship between their content. Search engine companies, in turn, try to identify such link farms via data mining, e.g., by looking for dense link structures between seemingly unrelated pages. Thus, while Pagerank gives a significant improvement in search result quality, constant efforts are needed to protect it against manipulation. There are two basic approaches: "cleaning" the Web graph before the Pagerank computation to remove suspicious edges, or modifying the computation itself, say by introducing appropriate edge weights.

PERSONALIZED SEARCH

Pagerank assigns to each page a global importance score that is independent of the query or the preferences of the user. There has been a lot of recent interest in link-based ranking techniques that consider these aspects [16], [22], [12], [14]. One interesting approach modifies the random jump in the Pagerank process so that the destination is chosen from a smaller set of pages that are known to be relevant to a particular topic [12] or popular with the user posing the query [14]. Since it is not realistic to repeat the Pagerank computation for each user or each query, recent research [14] has focused on how to compute such adaptive versions of Pagerank efficiently from precomputed base data, taking into account the local and global structure of the Web graph (e.g., the fact that a page is unlikely to significantly influence the Pagerank of another page that is far away in the graph.

In general, while Pagerank is probably the simplest and most widely known technique for link-based ranking in search engines, there is significant interest in other more advanced techniques. Almost all of these techniques can be stated in a matrix or Markov chain framework, and many are also based on iterative computations on a graph that can be studied through the theorems of Perron and Frobenius.

ACKNOWLEDGMENTS

We thank Yen-Yu Chen and Qingqing Gan for help with the Pagerank computation experiments.

AUTHORS

S. Unnikrishna Pillai received his B.Tech degree in electronics engineering from the Institute of Technology, India, in 1977, the M.Tech degree in electrical engineering from I.I.T. Kanpur, India in 1982, and the Ph.D. degree in systems engineering from the University of Pennsylvania, Philadelphia, in 1985. From 1978 to 1980 he was with Bharat Electronics Limited, Bangalore, India. In 1985 he joined the Department of Electrical Engineering at Polytechnic University, Brooklyn, New York, as an assistant professor and since 1995 he has been a professor of electrical and computer engineering. He is the author of *Array Signal Processing*, *Spectrum Estimation and System Identification*, and *Probability, Random Variables and Stochastic Processes* (fourth edition). His present research activities include radar signal processing, space based radar, waveform diversity, blind identification and deconvolution, spectrum estimation and system identification.

Torsten Suel is an associate professor in the Department of Computer and Information Science at Polytechnic University in Brooklyn, New York. He received a diploma degree from the Technical University of Braunschweig, Germany, and a Ph.D. from the University of Texas at Austin. After postdoctoral research at the NEC Research Institute, University of California at Berkeley, and Bell Labs, he joined Polytechnic University in the fall of 1998. His current main research interests are in the areas of Web search engines and Web data mining, Web and content distribution protocols, peer-to-peer and distributed systems, and algorithms.

Seunghun Cha received the B.S. degree from Konkuk University, Korea, in 1998 and the M.S. degree from Polytechnic University, Brooklyn, New York, in 2002. He is currently a Ph.D. candidate in the department of Electrical and Computer Engineering at Polytechnic University. His current research interests are in signal processing, optimum transmitter and receiver filter design for Papoulis' wireless applications, and time reversed signal processing.

REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Reading, MA: Addision-Wesley, 1999.

[2] A. Berman and R.J. Piemmons, *Nonnegative Matrices in the Mathematical Sciences*. Philadelphia, PA: SIAM, 1994.

[3] H. Bertoni, *Radio Propagation for Modern Wireless Systems*. New York: Prentice Hall, 2000.

[4] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," in *Proc. 7th World Wide Web Conf.*, 1998, pp. 107–117.

[5] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the Web: Experiments and models," in *Proc. 9th Int. World Wide Web Conf.*, 2000, pp. 309–320.

[6] Y. Chen, Q. Gan, and T. Suel, "I/O-efficient techniques for computing pagerank," in *Proc. 11th Int. Conf. Information and Knowledge Management*, Nov. 2002, pp. 549–557.

[7] B. Davison, "Recognizing nepotistic links on the Web," in *AAAI Workshop Artificial Intelligence Web Search*, July 2000, pp. 23–28.

[8] D. Fetterly, M. Manasse, M. Najork, and J. Wiener, "A large-scale study of the evolution of Web pages," in *12th Int. World Wide Web Conf.*, 2003.

[9] E. Garfield, "Citation analysis as a tool in journal evaluation," *Science*, vol. 178, no. 4060, pp. 471–479, 1972.

[10] S.A. Grandhi, R. Vijayan, and D.J. Goodman, "Distributed power control in cellular radio sytems," *IEEE Trans. Commun.*, vol. 42, pp. 226–228, Feb.–Apr., 1994.

[11] T.H. Haveliwala, "Efficient computation of pagerank," Stanford Univ., Tech. Rep. Oct. 1999. [Online]. Available: http://dbpubs.stanford.edu:8090/pub/1999-31

[12] T.H. Haveliwala, "Topic-sensitive pagerank," in *Proc. 11th Int. World Wide Web Conf.*, May 2002, pp. 517–526.

[13] R.A. Horn and C.R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1992, vol. 1 and vol. 2.

[14] G. Jeh and J. Widom, "Scaling personalized Web search," in *12th Int. World Wide Web Conf.*, 2003, pp. 271–279.

[15] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 8, no. 1, pp. 39–43, 1953.

[16] J. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[17] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities*. New York: Dover, 1992.

[18] M. Najork and J. Wiener, "Breadth-first search crawling yields high-quality pages," in *Proc. 10th Int. World Wide Web Conf.*, 2001, pp. 114–118.

[19] A. Ng, A. Zheng, and M. Jordan, "Link analysis, eigenvectors, and stability," in *Proc. 17th Int. Joint Conf. Artificial Intelligence*, 2001, pp. 903–910.

[20] A. Papoulis and S.U. Pillai, *Probability Random Variables and Stochastic Processes*. New York: McGraw Hill, 2002.

[21] C.R. Rao and M.B. Rao, *Matrix Algebra and Its Applications to Statistics and Econometrics*. Singapore: World Scientific, 1998.

[22] M. Richardson and P. Domingos, "The intelligent surfer: Probabilistic combination of link and content information in pagerank," in *Advances in Neural Information Processing Systems*, 2002, pp. 1441–1448.

[23] V. Shkapenyuk and T. Suel, "Design and implementation of a high-performance distributed Web crawler," in *Proc. Int. Conf. Data Engineering*, Feb. 2002, pp. 357–368.

[24] A.J. Viterbi, A.M. Viterbi, and E. Zehavi, "Other-cell interference in cellular power-controlled CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 1501–1504, Feb.–Apr., 1994.

[25] Q. Wu, "Performance of Optimum Transmitter Power Control in CDMA Cellular Mobile Systems," *IEEE Trans. Veh. Technol.*, vol. 48, no. 2, pp. 571–575, 1999.