32 Perron-Frobenius Theorem: context and applications

As mentioned in the last section, non-negative matrices are closely connected to *graphs*. If B is a non-negative $n \times n$ matrix, the graph $\Gamma(B)$ associated with B has vertex set $\{v_1, \dots, v_n\}$, with an arc directed from v_i to v_j if and only if the entry A_{ij} of A is non-zero (hence positive). If the zero-nonzero pattern of B is *symmetrix* (i.e. if B_{ji} is positive whenever B_{ij} is positive) then the graph associated with A may be regarded as undirected, since it has an arc from v_i to v_j if and only if it has one from v_j to v_i. On the other hand, the *adjacency matrix* $A(\Gamma)$ of a directed graph Γ with vertices $v_1, \dots, v_n$ is the matrix whose (i, j) entry is 1 if there is an from v_i to v_j in Γ, and 0 otherwise.

Everything that has been said here could apply equally well to any real matrix B as to a non-negative one. The reason for restricting attention here to non-negative matrices is related to the meaning, in terms of graphs, of the entries of the positive integer powers of a matrix.

Note that a matrix whose entries are all either 0 or 1 is referred to as a $(0, 1)$-matrix, and the *zero-nonzero pattern* of a non-negative matrix refers to the collection of positions where entries are zero or positive. So for example the matrices $\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}$ and $\begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$ have the same zero-nonzero pattern. If B is a non-negative matrix, then $A(\Gamma(B))$ is the unique $(0, 1)$-matrix that has the same zero-nonzero pattern as B.

Definition 32.1. *Let Γ be a directed graph, with vertices u and v. A* walk *from u to v in Γ is a sequence of arcs, where the intial vertex of the fist arc is u, the terminal vertex of the last arc is v, and the terminal vertex if each arc is the initial vertex of the next one (so informally it's a journey from u to v in the graph, along arcs). The length of a walk is the number of arcs in it.*

Lemma 32.2. *Let Γ be a directed graph with adjacency matrix A (with respect to the ordering $v_1, \dots, v_n$ of the vertices), and let k be a positive integer. The entry in the (i, j) position of A^k is the number of walks from v_i to v_j in Γ.*

Proof. The proof proceeds by induction on k. If $k = 1$ the statement is the the (i, j) entry is 1 or 0 according as there is an arc from v_i to v_j or not, which is clearly true. Now assume that the lemma holds for A^{k-1} and consider the (i, j)-entry of A^k.

$$(A^k)_{ij} = \sum_{m=1}^{n} (A_{im}^{k-1} A_{mj}).$$

By the induction hypothesis, A_{im}^{k-1} is the number of walks of length $k-1$ from v_i to v_m. Also A_{mj} is 1 if there is an arc from v_m to v_j and 0 if not. Thus $A_{im}^{k-1} A_{mj}$ is the number of walks of length k in Γ from v_i to v_j that have v_m as their second last vertex. The sum of these numbers over m is the total number of walks of length k from v_i to v_j. □

Now let B be any non-negative matrix, and let A be the $(0, 1)$-matrix that has the same zero-nonzero pattern as B, so $A = A(\Gamma(B))$. We observe that for every positive integer k, B^k and A^k have the same zero-nonzero pattern. Note that this would not necessarily be true if B had negative entries, since positive and negative contributions could cancel to produce a zero entry in B^k where there is a positive one in A^k. This is the reason for focussing on non-negative matrices in the graph interpretation. If B is a non-negative square matrix and k is a positive integer, the entry in the (i, j) entry of B^k is positive if and only if there is a walk of length k from v_i to v_j in $\Gamma(B)$. Recall that the matrix B is *primitive* if and only if B^k is positive for some positive integer k. This condition may be interpreted now as saying that there is a walk of length k in $\Gamma(B)$ from every vertex u to every vertex v (where the possibility that $u = v$ is included). A directed graph is said to be primitive if there is a positive integer k for which this property holds. This is a graph-theoretic concept of primitivity that does not rely on matrices. However a graph is primitive precisely if its adjacency matrix is a primitive matrix.

The Frobenius-Perron Theorem admits an interpretation in terms of graphs, that has been applied by Google in its PageRank algorithm for assigning relative importance to the results of searches. The webpages that turn up in the search may be regarded as vertices of a graph where there is an arc from vertex v_i to vertex v_j if and only if there is a link from Page i to Page j. Let n be the number of webpages involved. Let A be the transpose of the adjacency matrix of this graph, so that $A_{ij} = 1$ if there is a link from Page j to Page i, 0 otherwise. Suppose that the number of links from Page j is d_j. We assume that a "random surfer" on Page j will leave it via a randomly chosen link with probability 0.85, and will move to a randomly chosen page (linked from Page j or not) with probability 0.15 (these probabilities can obviously be adjusted for different implementations of the algorithm). We also assume that users have the opportunity to move (or stay where they are) at discrete time steps. We assume also that all links are equally likely to be chosen if a move is via a link, and that all pages are equally likely to be chosen if not (these assumptions can also be adjusted of course). In this model, the probability that a surfer at Page j will move to Page i at a given step is given by

- $\dfrac{0.85}{d_j} + \dfrac{0.15}{n}$, if there is a link from Page j to Page i, or
- $\dfrac{0.15}{n}$ if not.

Now let B be the matrix whose (i, j)-entry is the probability that a user at Page j will move to Page i at a given time step. Let D be the diagonal matrix whose jth diagonal entry is $1/d_j$ (if $d_j \neq 0$), or 0 if $d_j = 0$. Note that

$$B = (0.85)AD + (0.15)J,$$

where J is the $n \times n$ matrix whose entries are all 1. Now B is a positive matrix and the sum of the entries in each column of B is 1. So the Perron-Frobenius Theorem applies to B in its strong form. We need two theorems.

Theorem 32.3. *Let A be a positive $n \times n$ matrix with the property that the entries in every column sum to the same number k. Then k is the Perron eigenvalue of A.*

Proof. Let v be the vector of length n whose entries are all equal to 1. Then it is easily observed that $v^T A = kv^T$, since the jth entry of $v^T A$ is the sum of the entries in Column j of A. Thus v^T is a left eigenvector of A, and since v is a positive vector if follows from the Perron-Frobenius Theorem that k is the Perron eigenvalue, or spetral radius, of A. □

Note that the same statement and proof (with v instead of v^T) would apply to a matrix whose row sums are all equal to k.

If follows from Theorem 32.3 that the Perron eigenvalue of the PageRank matrix B above is 1.

Theorem 32.4. *Let A be a positive $n \times n$ matrix with Perron eigenvalue (or spectral radius) 1. Let v be any positive vector in $\mathbb{R}^n$. Then the sequence*

$$v, Av, A^2v, \dots$$

converges to a Perron eigenvector of A (with probability 1) or to the zero vector (with probability zero).

Note: the key statement here is that the sequence converges to a vector w for which $Aw = w$. There are some choices of v for which this will fail but a random choice of v is almost guaranteed to succeed. This technical point is explained in the proof below. The proof is subject to the simplifying assumption that A is diagonalizable. This assumption simplifies the proof but is not necessary for the statement.

Proof. Let $1, \lambda_2, \dots, \lambda_n$ be the eigenvalues of A, so $|\lambda_i| < 1$ for $i \geqslant 2$ by the Perron-Frobenius Theorem. Assume that A is diagonalizable and let $\{v_1, \dots, v_n\}$ be a basis of $\mathbb{R}^n$ consisting of eigenvectors of A, where $Av_1 = v_1$ and $Av_i = \lambda_i v_i$ for $i \geqslant 2$. Now

$$v = a_1v_1 + a_2v_2 + \cdots + a_nv_n,$$

and for a positive integer k

$$A^k v = a_1 A^k v_1 + a_2 A^k v_2 + \cdots + a_n A^k v_n = a_1 v_1 + a_2 \lambda_2^k v_2 + \cdots + a_n \lambda_n^k v_n.$$

Since $|\lambda_i| < 1$ for $i \geqslant 2$, $\lambda_i^k \to 0$ as $k \to \infty$. Hence the sequence $(A^k v)$ converges to $a_1 v_1$, which is a Perron eigenvector of A unless $a_1 = 0$, in which case it is the zero vector. $\square$

We now return to the PageRank matrix B. At the outset, let x be the vector whose jth entry x_j is the proportion of "random surfers" who are on Page j at a given time. The ith entry of Bx is

$$(Ax)_i = \sum_{j=1}^{n} B_{ij} x_j = \sum_{j=1}^{n} P(j \to i) x_j,$$

where $P(j \to i)$ is the probability that a random surfer at page j moves to Page i in a given step. Thus $\sum_{j=1}^{n} P(j \to i) x_j$ is the proportion of the overall population that will be at Page i one step after the step whose population distribution is described by the vector x, and so the vector Bx describes the population distubution one step later. By Theorem 32.4, the sequence $x, Bx, B^2x, \ldots$ converges to a vector y for which $By = y$ and $\sum y_i = 1$. Thus y describes a steady state of the system and its jth entry is the proportion of the surfing population that we be at Page j in the long term. Note that individuals continue to move between pages; what this is saying is that the dynamics settle down to a steady state where the overall proportions of the population at the various pages stabilizes. The pages are then ranked in terms of importance according to the entries of this Perron eigenvector.