The Perron Frobenius Theorem and a Few of Its Many Applications

Overview

The Perron-Frobenius Theorem arose from a very theoretical environment over 100 years ago in the study of matrices and eigenvalues.

In the last few decades, it has been rediscovered as a powerful tool in a myriad of applications including Biology, Economics, Dynamical Systems, and even ranking of football teams. We´ll talk about the theorem and few of its interesting applications.

Idea of Theorem

The **PFT** is about real square matrices with entries ≥ 0.

If all the entries are > 0, $\exists!$ largest real eigenvalue, it is positive, and the corresponding eigenvector has strictly positive components.

For *certain classes* of nonnegative matrices, we can say the same thing.

Terminology

A *positive* matrix

has all strictly positive entries (no zeros).

A *nonnegative* matrix

is one with nonnegative entries.

Notation

- Vectors are column vectors unless noted otherwise. e.g., y is a column vector, and y^T is row vector.
- When needed for clarification, **bold** means vector. e.g., 1 is just a number, but **1** is the vector $\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$.

Some Motivation

- In many applications, knowing the largest, or dominant eigenvalue is all that we need to know.
- There´s no reason that for a general matrix, a largest eigenvalue exists. There may be repeated values, or the least negative and greatest positive could be = in absolute value.
- There´s no reason the largest one is positive.
- PFT says: For certain matrices, $\exists!$ positive one.

The authors

The theorem was proved by Oskar Perron and Georg Frobenius.

1907-Perron proved it for positive matrices.

1912-Georg Frobenius extended the proof to nonnegative matrices.

There have been further extensions recently.

Oskar Perron (1880-1975)

- Famous for contributions to PDEs, including the “Perron Method” for solving the Dirichlet problem for elliptical PDEs.

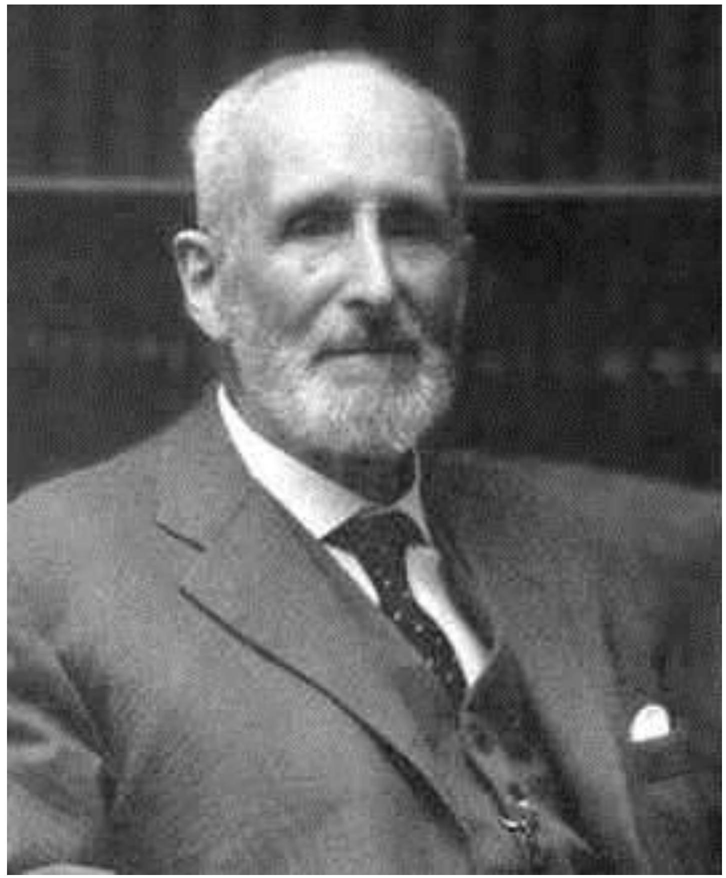


Photo courtesy Wikipedia

- “Perron´s Paradox”
- Thesis at Munich was on Geometry.
- Retired from teaching at 80, but published 18 more papers until ´73.

Ferdinand Georg Frobenius (1849-1917)

Diffy Qs

Elliptical Functions

Group Theory

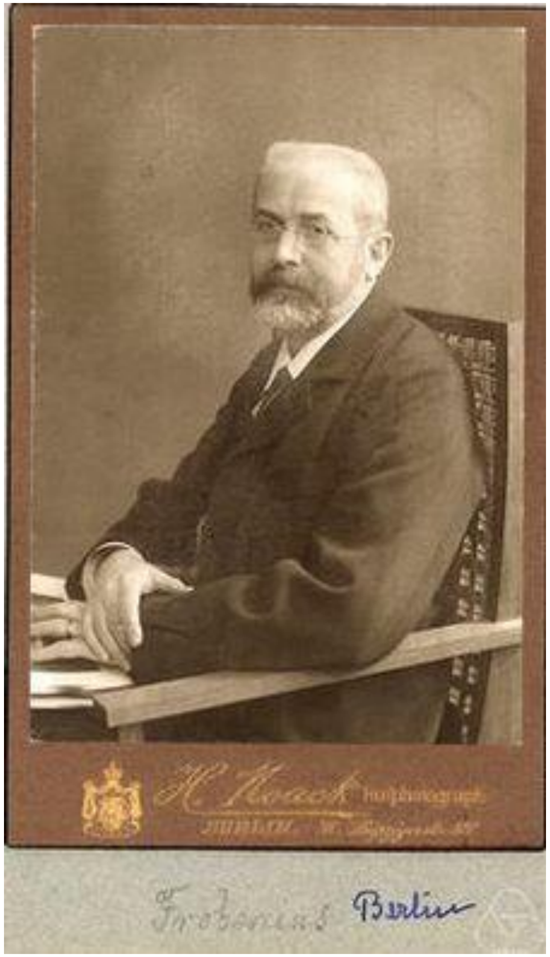


Photo courtesy Wikipedia

Student of Weierstrass

Applications

- Google Page Rank Algorithm
- Inputs to production processes are $\geq 0 \Rightarrow$ nonnegative matrices are important in linear economics.
- Biology

Both Perron and Frobenius were very theoretical. Especially Frobenius, who considered applications something for tech schools. This is one reason applications were mostly overlooked until recently.

Precise Statement of the Theorem

If A is a square positive matrix ,
$\rho(A) = \max \{ |\lambda| ; \lambda$ is an eigenvalue of A$\}$, and
$\mu(A) = \max \{ |\lambda| ; \lambda$ is an eigenvalue of A and $\lambda \neq \rho(A)\}$ then:

- $\exists$ a real eigenvalue $r > 0$, called the Perron Root, and any other eigenvalue is strictly smaller in absolute value. Thus $r = \rho(A) > 0$, $\rho(A)$ is simple and $\mu(A) < \rho(A)$.

- The corresponding eigenvector is positive. No eigenvectors except those associated to r have only positive parts.

- $\lim_{n\to\infty} \left(\frac{A}{\rho(A)}\right)^n = \mathbf{x}\mathbf{y}^T$ where $A\mathbf{x} = \rho(A)\mathbf{x}$ for $\mathbf{x} > 0$, $A^T\mathbf{y} = \rho(A)\mathbf{y}$, $\mathbf{y} > 0$, and $\mathbf{x}^T\mathbf{y} = \mathbf{1}$.

(Note: An entry in **x** is just a value, whereas each entry of $\mathbf{y}^T$ is a column. $\mathbf{y}^T$ is a *row* of *columns*, i.e. a matrix.)

- The Perron–Frobenius eigenvalue satisfies the inequalities:

$$\min_i \sum_j a_{ij} \leq r \leq \max_i \sum_j a_{ij} .$$

Precise Statement of the Theorem

This is an important consequence of the theorem:

If A is a square positive matrix ,
$\rho(A) = \max \{ |\lambda| ; \lambda$ is an eigenvalue of A$\}$, and
$\mu(A) = \max \{ |\lambda| ; \lambda$ is an eigenvalue of A and $\lambda \neq \rho(A)\}$ then:

- For every q, $\mu(A) / \rho(A) < q < 1$, there exists a constant K such that for every n,

$$\left\| \left(\frac{A}{\rho(A)} \right)^{n} - \mathbf{x}\mathbf{y}^{T} \right\|_{\infty} \leq K\, q^{m} \qquad \text{where } \|A\|_{\infty} = \max_{j} \sum_{i=0}^{n} |a_{ij}|$$

The third bullet on the last page said that there is convergence.
This consequence says more: The iterates of the matrix converge *exponentially* (in the max abs row sum norm).

Extensions to Nonnegative Matrices

In 1912, Frobenius extended Perron´s original theorem to several types of nonnegative matrices.

The first extension is that the matrix can have a few zeros, as long as it´s *primitive*.

What is a Primitive Matrix?

A nonnegative matrix is primitive if for some natural k, A^k is positive.

Examples:

1) Of course, if it starts out positive ✓

2) $\begin{bmatrix} 0 & 3 & 2 \\ 0 & 5 & 2 \\ 3 & 0 & 0 \end{bmatrix}$ is.

$$A^2 = \begin{bmatrix} 6 & 15 & 6 \\ 6 & 25 & 10 \\ 0 & 9 & 6 \end{bmatrix}, \quad A^3 = \begin{bmatrix} 126 & 519 & 222 \\ 186 & 805 & 346 \\ 54 & 279 & 126 \end{bmatrix}$$

What´s not a primitive matrix?

$\begin{bmatrix} 0 & 3 \\ 5 & 0 \end{bmatrix}$ is not, but …we got lucky. You may have to try multiplying infinitely many times to know that it´s NOT primitive!

Fortunately, there´s Wielandt´s Theorem:

$$A^{(n-1)^2 + 1} \text{ is positive iff } A \text{ is primitive.}$$

Another test of Primitivity

A matrix is primitive

iff that limit

$$\lim_{n\to\infty}\left(\frac{\text{A}}{\rho(\text{A})}\right)^{n} \text{ exists.}$$

Simple Example

Let A = $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

ρ(A) = max { |λ| ; λ is an eigenvalue of A} = 1

(there´s only one eigenvalue, which we can find)

$\lim_{n\to\infty} \left(\frac{\text{A}}{\rho(\text{A})}\right)^n = \lim_{n\to\infty} (\text{A})^n$, which DNE because A^n alternates between A and I.

Thus A is not primitive.

1st Extension

If the matrix is primitive then…

$$\begin{bmatrix} \textit{same as above} \end{bmatrix}$$

What is an Irreducible Matrix?

Matrix A is *reducible* if ∃ a permutation matrix P, such that $P^{-1}AP = \begin{bmatrix} B & D \\ 0 & C \end{bmatrix}$ where B and C are square matrices. Otherwise, A is *irreducible*.

For $A \geq 0$, TFAE definitions of irreducible:

1. The digraph associated to A is strongly connected.
2. For each i and j, ∃ some k such that $(A^k)_{ij} > 0$.
3. For any partition $J \sqcup K$ of the index set $\{1,2,\ldots,n\}$, $\exists\ j \in J$ and $k \in K$ such that $a_{jk} \neq 0$.
4. A cannot be conjugated into a block upper triangular matrix by a permutation matrix P.

Some irreducible matrices

Notice 2. looks similar to definition of primitive. In fact,...

Primitive $\Rightarrow$ irreducible.

In particular, a positive matrix is irreducible.

Once again, an awkard definition…

Irreducible matrices turn out to be the important classification for the PFT. To say it is "NOT REDUCIBLE", however, may be difficult.

Two more Helpful Theorems

For a nonnegative matrix A,

1. A is irreducible $\Rightarrow (\mathtt{I}+A)^{n-1} > 0$.

2. A is irreducible with at least one non-zero diagonal element $\Rightarrow$ A is primitive.

Revisiting the "Simple Example"

Let A = $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
A is nonnegative, and irreducible (it´s digraph is strongly connected 0↔1)

By the PFT we can say ∃ a real eigenvalue r > 0 and $\min_i \sum_j a_{ij} \leq r \leq \max_i \sum_j a_{ij}$, so it is 1. We **can´t** go on to claim uniqueness (indeed a quick calculation shows, it´s not the strictly largest in abs value.)

Irreducible ⇒ we lose a guarantee of uniqueness, but have the other properties of the PFT.

Markov Chains Example

A transition matrix of a Markov Chain is an example of a nonnegative matrix.

Suppose the transition matrix is such that

a_{ij} = the probability of going from state i to state j

Reducible if the absorbing states are at the bottom of the matrix. Again when can we be sure it is *irr*educible?

Graphs Example

If we considered directed graphs then each has associated with it a nonnegative matrix with all entries 0 or 1 with

$a_{ij} = 1$ if there is an arc from vertex i to vertex j.

Irreducible means that you can go from any vertex to another (may be several steps)

This is called a “Strongly Connected” graph.

Dynamical Systems Example

In studying population, we might have $\mathbf{x}(k+1) = A\,\mathbf{x}(k)$ $\quad k=0,1,\ldots$ where A is reducible.

Then with partitioned matrices, we can rewrite this system as

$$\mathbf{Y}(k+1) = A_{11}\,\mathbf{Y}(k) + A_{12}\,\mathbf{Z}(k)$$
$$\mathbf{Z}(k+1) = A_{22}\,\mathbf{Z}(k)$$

where the first r components of **X** are contained in **Y**, and the last n-r are in **Z**.

Now we can solve for **Z** with no reference to the system involving **Y**, then solve for **Y** with **Z** assumed to be known. We "reduced“ the original system to two simpler systems.

Thus “the matrix A is irreducible” means that “the system cannot be reduced”; When studying the behavior of the system, we must treat it *as a whole*—we can´t break it up.

Applications of the PFT

Nonnegative matrices arise in many fields, e.g.,

- Economics
- Population models
- Graph theory
- Markov chains
- Power control in communications
- Lyapunov analysis of large scale systems
- Ranking systems

Applications of the PFT

The Perron-Frobenius theorem has several uses with matrix problems. It helps determine what types of vectors are special for many types of matrices encountered in the real world, such as stochastic matrices. Most often it is used to state that there is a solution to a problem where it might not be that easy to determine if one exists (such as problems that deal with large matrices).

Here are some detailed examples…

Graph Theory-Path Count

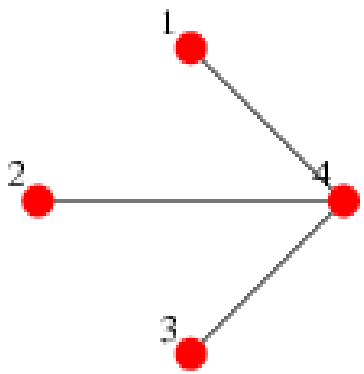
Given a digraph of n nodes with adjacency matrix A.

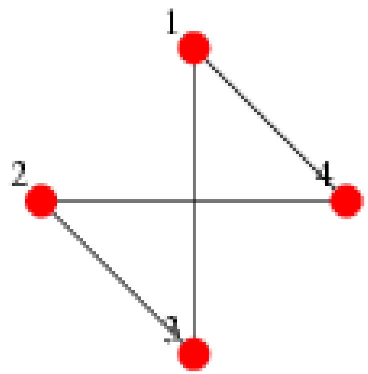
A_{ij} = 1 when there's an edge from node *i* to *j*,

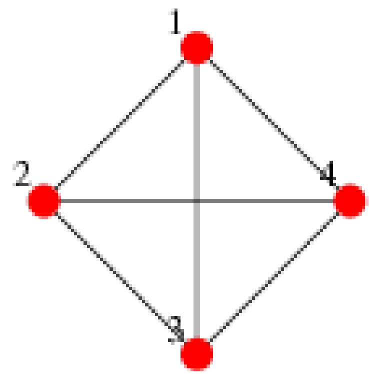
0 otherwise

$(A^k)_{ij}$ = number of paths from *i* to *j* of length *k.*

Adjacency Matrix



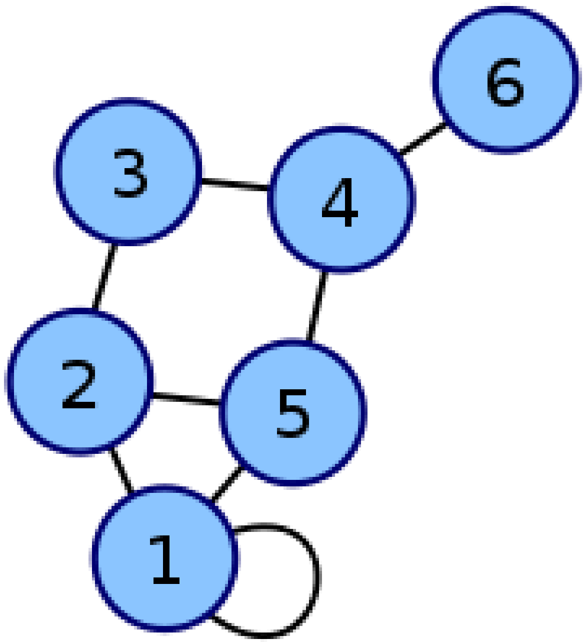




$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

From Wolfram MathWorld

Adjacency Matrix



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

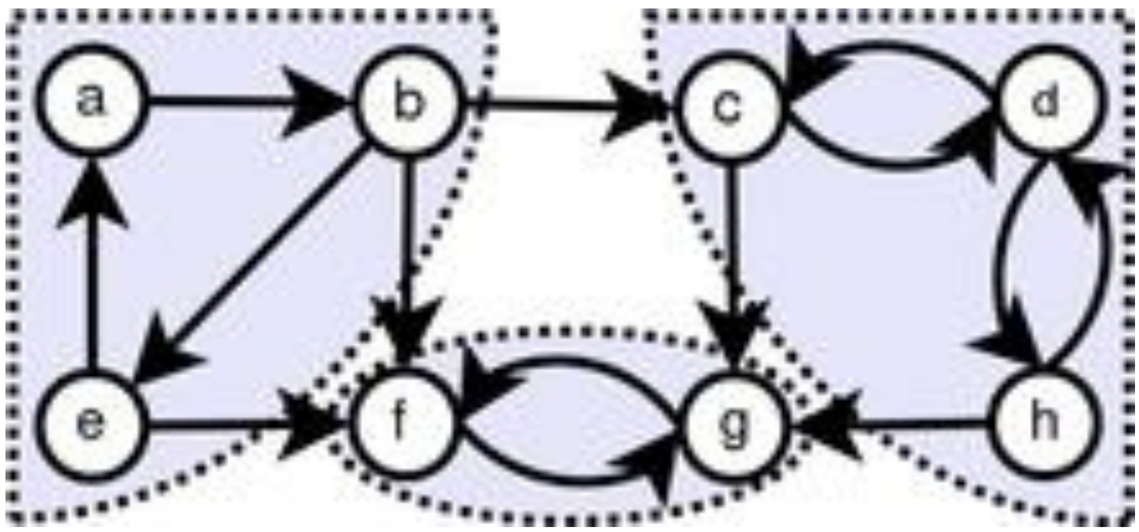
From Wikipedia

Graph Theory-Path Count

A graph is said to be **strongly connected** if every vertex is reachable from every other vertex.

The adjacency matrix of a strongly connected graph is irreducible.

Graph Theory-Path Count



Strongly Connected
Components of a Graph

From Wikipedia

Graph Theory-Path Count

A is primitive $\Rightarrow$ for large k , $A^k \sim r^k xy^T = r^k(\mathbf{1}^Ty)x(y/\mathbf{1}^Ty)^T$
where x is a right eigenvector, so normalized: $\mathbf{1}^Tx = 1$
and y is the left e.v., so $y^Tx = 1$.

The dominant term of $r^k xy^T$ will be a good estimate for large enough k.

- r is the factor of increase in # of paths when length increases by one
- x_i = fraction of length k paths that end at i. Measures importance/connectedness of node i as a sink.
- $y_i/\mathbf{1}^Ty$ = fraction of length k paths that start at j. Measures importance/connectedness of node j as a source.
- multiply these to get fraction of length k paths starting at i and ending at j.

Graph Theory

So much more… just one example:

A *cycle* is a path starting and ending at the same vertex.

If the graph associated to M is strongly connected and has two cycles of relatively prime lengths, then M is primitive.

This was just to point out that problems may be interpreted in the world of graphs or matrices, and go back and forth…

Markov Chains

Suppose we have a stochastic process $X_0, X_1, \ldots$ with values in $\{1, \ldots, n\}$ and $\mathrm{Prob}(X_{t+1} = i | X_t = j) = P_{ij}$.

P is called the transition matrix. Note $P_{ij} \geq 0$

Let $p_t \in R^n$ be the distribution of X_t, i.e., $(p_t)_i = \mathrm{Prob}(X_t = i)$

then we have $p_{t+1} = P\, p_t$

P is a (right) stochastic matrix, i.e., $P \geq 0$ and the sum of the probabilities on any row is 1.

P is non negative, but may not be irreducible, so we can´t apply PFT.

However,…

Markov Chains

...notice that the column vector with each entry 1 is an eigenvector corresponding to the eigenvalue 1.

In fact it is the Perron-Frobenius eigenvalue of P.

(Markov Chains) Quick Calculation

Suppose we have $A = \begin{bmatrix} .6 & .3 & .1 \\ .5 & .2 & .3 \\ 0 & .5 & .5 \end{bmatrix}$ and $B = \begin{bmatrix} .6 & .3 & .1 \\ .4 & .2 & .4 \\ 0 & .5 & .5 \end{bmatrix}$

1) Then for both matrices, the column vector **1** is a right eigenvector corresponding to eigenvalue 1:

$$A\,\mathbf{1} = \begin{bmatrix} .6 & .3 & .1 \\ .5 & .2 & .3 \\ 0 & .5 & .5 \end{bmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \mathbf{1} = 1\,\mathbf{1} \qquad \text{and} \qquad B\,\mathbf{1} = \begin{bmatrix} .6 & .3 & .1 \\ .4 & .2 & .4 \\ 0 & .5 & .5 \end{bmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \mathbf{1} = 1\,\mathbf{1}$$

2) For B, the row vector $\mathbf{1}^T$ is a left eigenvector corresponding to 1:

$$\mathbf{1}^T B = (1 \quad 1 \quad 1)\begin{bmatrix} .6 & .3 & .1 \\ .4 & .2 & .4 \\ 0 & .5 & .5 \end{bmatrix} = (1 \quad 1 \quad 1) = \mathbf{1}^T = 1\,\mathbf{1}^T$$

This second part works because each column of B adds to one, making it a doubly stochastic matrix. The second part need not be true for a (right) stochastic matrix in general: As an example, it is not true for matrix A.

Markov Chains

It might not be the only eigenvalue on the unit circle: and the associated eigenspace can be multi-dimensional.

Markov Chains

A stationary probability vector is defined as a vector that does not change under application of the transition matrix; that is, it is a left eigenvector of the probability matrix, associated with eigenvalue 1: $\boldsymbol{\pi}\mathrm{P} = \boldsymbol{\pi}$

(Recall left e.v. is $\mathbf{u}A = \lambda\mathbf{u}$, or $A^T\mathbf{u}^T = \lambda\mathbf{u}^T$)

Every stochastic matrix has such a vector (at least "1/n" vector on the right, so some left vector is associated to 1 on the left) called the "invariant probability distribution".

PFT ensures that the largest absolute value of an eigenvalue is always 1.

In general, there may be several such vectors. However, for a matrix with strictly positive entries, this vector is unique and can be computed using

$\lim_{n\to\infty}\left(\frac{A}{\rho(A)}\right)^n = \mathbf{x}\mathbf{y}^T$ where $A\mathbf{x} = \rho(A)\mathbf{x}$ for $\mathbf{x} > 0$, $A^T\mathbf{y} = \rho(A)\mathbf{y}$, $\mathbf{y} > 0$, and $\mathbf{x}^T\mathbf{y} = \mathbf{1}$

where we set A = P, our probability matrix.

Markov Chains

ρ(A) = 1, so it simplifies to:

$\lim_{n\to\infty}(\mathrm{P})^n = \mathbf{x}\mathbf{y}^T$ where P**x**=**x** for **x**>0, $\mathrm{P}^T\mathbf{y} = \mathbf{y}$, **y**>0, and $\mathbf{x}^T\mathbf{y}=\mathbf{1}$.

Observing that for any row i, $\lim_{n\to\infty}(\mathrm{P}^n)_{ij} = \boldsymbol{\pi}_j$

(so $\boldsymbol{\pi}_j$ is the j $^{\mathrm{th}}$ element of the row vector $\boldsymbol{\pi}$)

we come to the conclusion that the long-term probability of being in a state *j* is independent of the initial state *i*.

Markov Chains-Genetics

Inbreeding

2 homologous chromosomes take on A or a.

Individual will have AA, Aa=aA, or aa.

Offspring of two such individuals has AA, Aa, aa with probabilities given by the transition matrix:

$$\begin{array}{cc} & \begin{array}{ccc} AA & Aa & aa \end{array} \\ \begin{array}{c} AA \\ Aa \\ aa \end{array} & \begin{bmatrix} 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

Markov Chains-Genetics

	A	A
A	AA	AA
A	AA	AA

$$
\begin{array}{c c}
 & \begin{array}{ccc} \text{AA} & \text{Aa} & \text{aa} \end{array} \\
\begin{array}{c} \text{AA} \\ \text{Aa} \\ \text{Aa} \end{array} & \begin{bmatrix} 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

	A	a
A	AA	Aa
a	Aa	aa

Markov Chains-Genetics

Continue inbreeding, and eventually you get a pure line AA or aa

$$P^n = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2}-\frac{1}{2}^{n+1} & \frac{1}{2}^{n} & \frac{1}{2}-\frac{1}{2}^{n+1} \\ 0 & 0 & 1 \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{bmatrix} \text{as } n \rightarrow \infty$$

Markov Chains-Genetics

Other Ways to Breed

- Reject unfavourable genes during the breeding process
- (More aggressive:) Back-cross to encourage desirable genes
- etc.

Markov Chains-Filling Gas Stations

Suppose that you own 5 gas stations and a refueling truck. Because of availability and cost of hiring a driver to get the fuel from the depot and expenses of running the truck, you schedule one trip for the truck per day, visiting the station that needs the gas the most (most likely to run out). If the truck does not have enough in it to fill the station, it may have to go to the depot first. The truck can hold enough for a few stops (usually 2 to 4).

Your five stations are in Miami: MIA (near the airport), CG1 and CG2 in Coral Gables, OVT (Overtown), and NMB (North Miami Beach). The depot is close to the MIA station

If the truck starts in MIA, these are the probabilities of where it is left at the end of day:

.1 MIA, .4 Coral Gables, .3 OVT, .2 NMB,

From OVT:

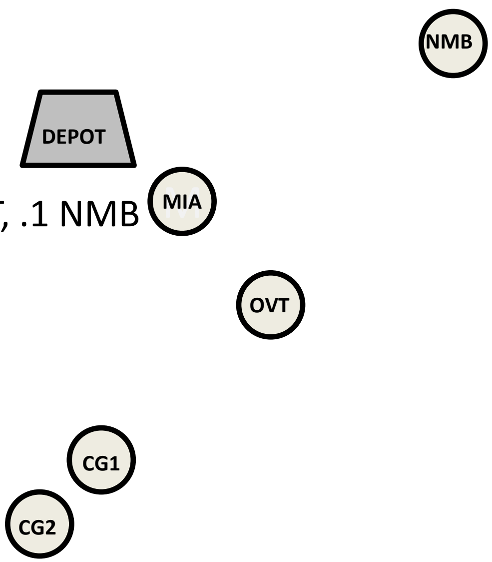
.4 MIA, .4 Coral Gables, .1 OVT, .1 NMB

From Coral Gables:

.3 MIA, .3 the other CG store, .1 the same CG store, .2 OVT, .1 NMB

From NMB:

.3 MIA, .4 Coral Gables, .2 OVT, .1 NMB



Markov Chains-Filling Gas Stations

Ideally, the truck will fill up at the depot before or after a stop at MIA. According to the plan, a truck that is heading to MIA without enough to fill MIA will go to the depot beforehand, and a truck that starts in MIA, after receiving a call to fill some station, will go to the depot first if it doesn´t have enough for that station. One use of this study is to consider if a partially filled truck should stop at the depot when it has enough fuel to fill the station that calls, but the truck happens to be starting out from MIA.

We would like to know how often the truck is in MIA.

We can calculate specific cases of where the truck will be:

e.g, If a car starts today in OVT, what´s the probability that it´s in MIA tomorrow morning?

Markov Chains-Filling Gas Stations

We set up a transition matrix:

$$P = \begin{array}{c} \\ \text{MIA} \\ \text{CG1} \\ \text{CG2} \\ \text{OVT} \\ \text{NMB} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{MIA} & \text{CG1} & \text{CG2} & \text{OVT} & \text{NMB} \end{array} \\ \begin{bmatrix} .1 & .4 & 0 & .3 & .2 \\ .3 & .3 & .1 & .2 & .1 \\ .3 & .1 & .3 & .2 & .1 \\ .4 & .4 & 0 & .1 & .1 \\ .3 & .4 & 0 & .2 & .1 \end{bmatrix} \end{array}$$

(Note: Some probablities specific to each store in CG were lacking, but that didn´t stop us. We grouped the stores together; We just need to be careful to interpret our results accordingly.)

Markov Chains-Filling Gas Stations

Then

$$
\begin{array}{c}
\\
(0 \quad 0 \quad 0 \quad 1 \quad 0)
\end{array}
\begin{array}{c}
\begin{array}{ccccc} \text{MIA} & \text{CG1} & \text{CG2} & \text{OVT} & \text{NMB} \end{array} \\
\begin{bmatrix} .1 & .4 & 0 & .3 & .2 \\ .3 & .3 & .1 & .2 & .1 \\ .3 & .1 & .3 & .2 & .1 \\ .4 & .4 & 0 & .1 & .1 \\ .3 & .4 & 0 & .2 & .1 \end{bmatrix}
\end{array}
= (.4 \quad .4 \quad 0 \quad .1 \quad .1)
$$

tells us where the filling truck will end up.

With .4 probability, it will be in MIA.

Markov Chains-Filling Gas Stations

$$
(.4 \quad .4 \quad 0 \quad .1 \quad .1)
\begin{array}{ccccc}
\text{MIA} & \text{CG1} & \text{CG2} & \text{OVT} & \text{NMB}
\end{array}
$$

$$
(.4 \quad .4 \quad 0 \quad .1 \quad .1)\begin{bmatrix} .1 & .4 & 0 & .3 & .2 \\ .3 & .3 & .1 & .2 & .1 \\ .3 & .1 & .3 & .2 & .1 \\ .4 & .4 & 0 & .1 & .1 \\ .3 & .4 & 0 & .2 & .1 \end{bmatrix} = \begin{array}{ccccc} \text{MIA} & \text{CG1} & \text{CG2} & \text{OVT} & \text{NMB} \\ (.23 & .36 & .04 & .23 & .14) \end{array}
$$

tells us where the truck will end up in two days.

With .23 probability, it will be in MIA.

Markov Chains-Filling Gas Stations

To get the probability distribution π_k for where the car is on day k we calculate $\boldsymbol{\pi_k} = \mathbf{x_o}^{\mathbf{T}}\ \mathbf{P}^{\mathbf{K}}$.

Seems simple enough, but…
An important question in the theory of Markov chains is:

$$\boxed{\text{Does the limit exist?}}$$

(i.e, Does $\pi = \lim_{k\to\infty} (\pi_k)^n$ exist?)

- Can be viewed as the distribution of the time the Markov chain spends in every state.
- Often want to know if π is the same for any intial state x_o

$$\boxed{\text{Does the limit depend on } x_o\text{?}}$$

Markov Chains-Filling Gas Stations

To get answers to our questions, we look at the transition matrix. Three possible cases:

1) Irreducible, primitive → PFT gives us the stationary distribution

2) Irreducible, but not primitive → PFT gives us some information

3) Reducible → can´t use PFT, but we can approximate

Markov Chains-FGS Case 1

P Primitive ⇒ P has a right Perron vector **1**/n, where n is # of elements.

$\boldsymbol{\pi}^{\mathbf{T}}$ is a right Perron vector for $P^{\mathbf{T}}$.

Thus, $\lim_{k\to\infty} (\mathrm{P})^k = \lim_{k\to\infty} \left(\frac{\mathrm{P}}{\rho(\mathrm{P})}\right)^k = \frac{(\mathbf{1}/n)\boldsymbol{\pi}^{\mathbf{T}}}{\boldsymbol{\pi}^{\mathbf{T}}(\mathbf{1}/n)} = \mathbf{1}\boldsymbol{\pi}^{\mathbf{T}} > 0$ by PFT.

So $\lim_{k\to\infty} (\mathbf{x}_{\mathbf{o}}^{\mathbf{T}}\ \mathrm{P}^k) = \mathbf{x}_{\mathbf{o}}^{\mathbf{T}}\mathbf{1}\,\boldsymbol{\pi}^{\mathbf{T}} = \boldsymbol{\pi}^{\mathbf{T}}$ ⇒ The Markov Chain converges to the Perron vector of $P^{\mathbf{T}}$, regardless of the initial state $\mathbf{x}_{\mathbf{o}}$.

Markov Chains-FGS Case 1

So in our example, since the matrix is primitive (irreducible because $\exists$ a path between every state, and at least one diagonal element is not zero $\Rightarrow$ primitive)

$$P = \begin{array}{c} \begin{array}{ccccc} \text{MIA} & \text{CG1} & \text{CG2} & \text{OVT} & \text{NMB} \end{array} \\ \begin{bmatrix} .1 & .4 & 0 & .3 & .2 \\ .3 & .3 & .1 & .2 & .1 \\ .3 & .1 & .3 & .2 & .1 \\ .4 & .4 & 0 & .1 & .1 \\ .3 & .4 & 0 & .2 & .1 \end{bmatrix} \end{array}$$

we know the limit exists and it´s the Perron vector which we calculate: $\boldsymbol{\pi}^T \approx (.2672 \quad .35 \quad .05 \quad .2061 \quad .1267)^T$.

This is the stationary distribution of the Markov Chain.

Markov Chains-FGS Case 2

Even if the matrix were not primitive, but at least irreducible, we can do almost the same thing:

-- at least two simple eigenvectors on ρ(P) = 1

-- **1**/n is still a right Perron vector for P

-- $\lim_{k\to\infty} (\mathrm{P})^k$ doesn´t converge, but the Cesáro sum,

$\lim_{k\to\infty} \frac{1}{k} \sum_{j=0}^{k} \mathrm{P}^j$ does, also to $\mathbf{1}\boldsymbol{\pi}^{\mathbf{T}} > 0 \Rightarrow$

$\lim_{k\to\infty} \left(\mathbf{x_o}^{\mathbf{T}} \frac{1}{k} \sum_{j=0}^{k} \mathrm{P}^j \right) = \mathbf{x_o}^{\mathbf{T}}\mathbf{1}\, \boldsymbol{\pi}^{\mathbf{T}} = \boldsymbol{\pi}^{\mathbf{T}}$ $\Rightarrow$ the stationary distribution π exists independent of the initial state but the Markov chain doesn’t converge to it—it oscillates around it.

Markov Chains-FGS Case 3

So if the stochastic matrix is reducible, we can´t invoke PFT, but we can permute a reducible matrix into upper triangular form:

$$PAP^{T} = \begin{bmatrix} X & Y \\ 0 & Z \end{bmatrix}$$

If X or Z are reducible, we can permute again, …until PAP^{T} is upper triangular and each block is either irreducible or a single zero.

Markov Chains-FGS Case 3

…then we put the matrix in canonical form by permuting blocks with nonzeros only on the diagonal to the bottom

Transient States
(never come back)

$$\begin{matrix} \begin{bmatrix} B & \cdots & B \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B \end{bmatrix} & \begin{bmatrix} B & \cdots & B \\ \vdots & \ddots & \vdots \\ B & \cdots & B \end{bmatrix} \\ \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} & \begin{bmatrix} B & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B \end{bmatrix} \end{matrix}$$

Ergodic/Absorbing States
(each block is irreducible, stochastic)

Markov Chains-FGS Case 3

Now if this is in our matrix is in canonical form, call it C then $\lim_{k\to\infty} \mathrm{C}^k$ can be found if all the blocks of Z from $\mathrm{C} = \begin{bmatrix} \mathrm{X} & \mathrm{Y} \\ 0 & \mathrm{Z} \end{bmatrix}$ are primitive. If one or more are not primitive, we can still find the limit of the Cesáro sum.

For the primitive case the limit is $(\mathrm{I}\text{-}\mathrm{X})^{-1}\mathrm{Y} \lim_{k\to\infty} (\mathrm{P})^k = \begin{bmatrix} 0 & (\mathrm{I}-\mathrm{X})^{-1}\mathrm{Y} \\ 0 & \mathrm{Z} \end{bmatrix}$

Markov Chains-FGS Case 3

$$\lim_{k\to\infty} (\mathrm{P})^k = \begin{bmatrix} 0 & (\mathrm{I}-\mathrm{X})^{-1}\mathrm{Y} \\ 0 & \mathrm{E} \end{bmatrix}$$

where $$\mathrm{E} = \begin{bmatrix} \mathbf{1}\pi_{j+1}^{T} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{1}\pi_{k}^{T} \end{bmatrix}$$

The elements of $(\mathrm{I}-\mathrm{X})^{-1}\mathrm{Y}$ are the probabilities that if we start in state *p* we leave the transient states and eventually hit state *q* in one of the absorbing blocks. "The hitting probability to hit q starting in p".

$(\mathrm{I}-\mathrm{X})^{-1}\mathbf{1}$ tells us the average # of steps to leave the transient state. "The hitting time for reaching the absorbing states".

Markov Chains-FGS Case 3

We found that with about 27% probability the truck from OVT ends in MIA.

New question: How long (days) do we have to wait for the truck to go to MIA?

Answer: If you´re already at MIA that´s “good” so make that an absorbing state, which gives us a reducible matrix.

Markov Chains-FGS Case 3

$$P = \begin{array}{c} \\ \text{MIA} \\ \text{CG1} \\ \text{CG2} \\ \text{OVT} \\ \text{NMB} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{MIA} & \text{CG1} & \text{CG2} & \text{OVT} & \text{NMB} \end{array} \\ \begin{bmatrix} .1 & .4 & 0 & .3 & .2 \\ .3 & .3 & .1 & .2 & .1 \\ .3 & .1 & .3 & .2 & .1 \\ 1 & 0 & 0 & 0 & 0 \\ .3 & .4 & 0 & .2 & .1 \end{bmatrix} \end{array}$$

Now put this in canonical form and look at

$(I-X)^{-1}$ **1** tells us the average # of days to leave the transient state. "The hitting time for reaching the absorbing states". 4 days to get end up in MIA.

Markov Chains-FGS Case 3

$$P = \begin{array}{c c} & \begin{array}{cccc} \text{FLL} & \text{WPB} & \text{MIA} & \text{MCG} \end{array} \\ \begin{array}{c} \text{FLL} \\ \text{WPB} \\ \text{MIA} \\ \text{MCG} \end{array} & \begin{bmatrix} .3 & .3 & .4 & 0 \\ 0 & 1 & 0 & 0 \\ .2 & .2 & .5 & .1 \\ .2 & .2 & .1 & .5 \end{bmatrix} \end{array}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} .3 & .3 & .4 & 0 \\ 0 & 1 & 0 & 0 \\ .2 & .2 & .5 & .1 \\ .2 & .2 & .1 & .5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} =$$

$\}$

$$\begin{array}{cc} \begin{bmatrix} .3 & .3 & .4 & 0 \\ 0 & 1 & 0 & 0 \\ .2 & .2 & .5 & .1 \\ .2 & .2 & .1 & .5 \end{bmatrix} & \begin{array}{l} \text{FLL} \\ \text{MCG} \\ \text{MIA} \\ \text{WPB} \end{array} \end{array}$$

Markov Chains-FGS Case 3

$$P = \begin{array}{c} \\ \text{FLL} \\ \text{WPB} \\ \text{MIA} \\ \text{MCG} \end{array} \begin{array}{c} \begin{array}{cccc} \text{FLL} & \text{WPB} & \text{MIA} & \text{MCG} \end{array} \\ \begin{bmatrix} .3 & .3 & .4 & 0 \\ 0 & 1 & 0 & 0 \\ .2 & .2 & .5 & .1 \\ .2 & .2 & .1 & .5 \end{bmatrix} \end{array}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} .3 & .3 & .4 & 0 \\ 0 & 1 & 0 & 0 \\ .2 & .2 & .5 & .1 \\ .2 & .2 & .1 & .5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} =$$

$$\begin{bmatrix} .3 & .3 & .4 & 0 \\ 0 & 1 & 0 & 0 \\ .2 & .2 & .5 & .1 \\ .2 & .2 & .1 & .5 \end{bmatrix} \begin{array}{l} \text{FLL} \\ \text{MCG} \\ \text{MIA} \\ \text{WPB} \end{array}$$

Markov Chains-FGS Case 3

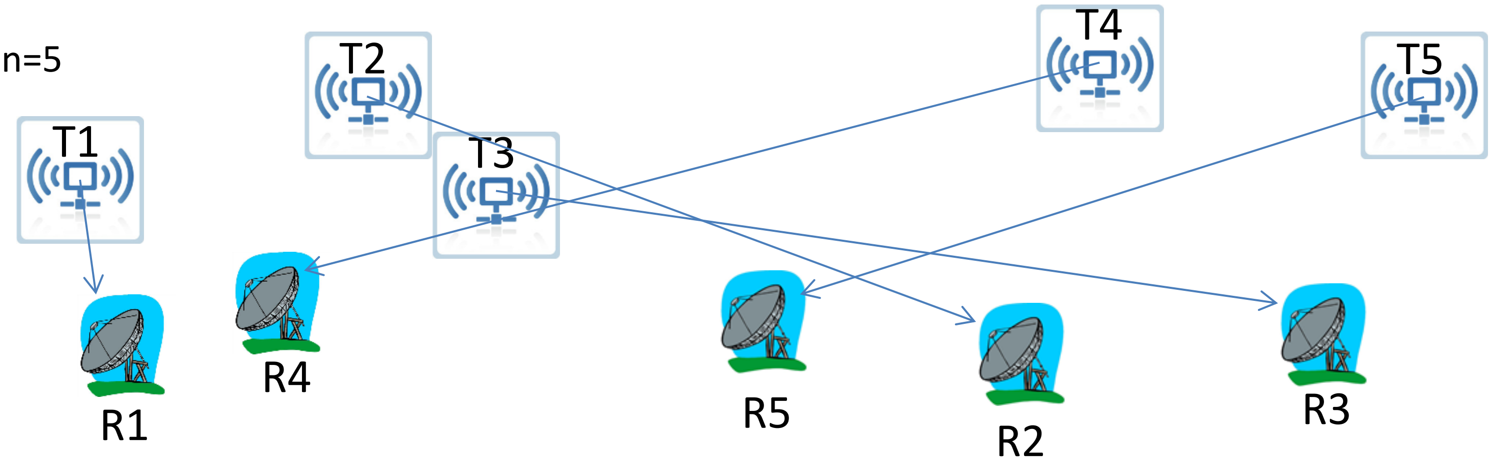
$$X = \begin{bmatrix} .3 & 0 & .4 \\ .2 & .5 & .1 \\ .2 & .1 & .5 \end{bmatrix}$$

$$(I\text{-}X)^{-1}\,\mathbf{1} = \begin{bmatrix} 1.82 & .23 & 1.14 \\ .91 & 2.2 & .98 \\ .91 & .53 & 2.65 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{array}{c} \begin{bmatrix} 4 \\ 4.5 \\ 4.5 \end{bmatrix} \end{array} \begin{array}{l} \text{WPB} \\ \text{MCG} \\ \text{MIA} \end{array}$$

Communications

Let say we have n transmitters with powers $P_1, \dots, Pn > 0$, transmitting to n receivers.

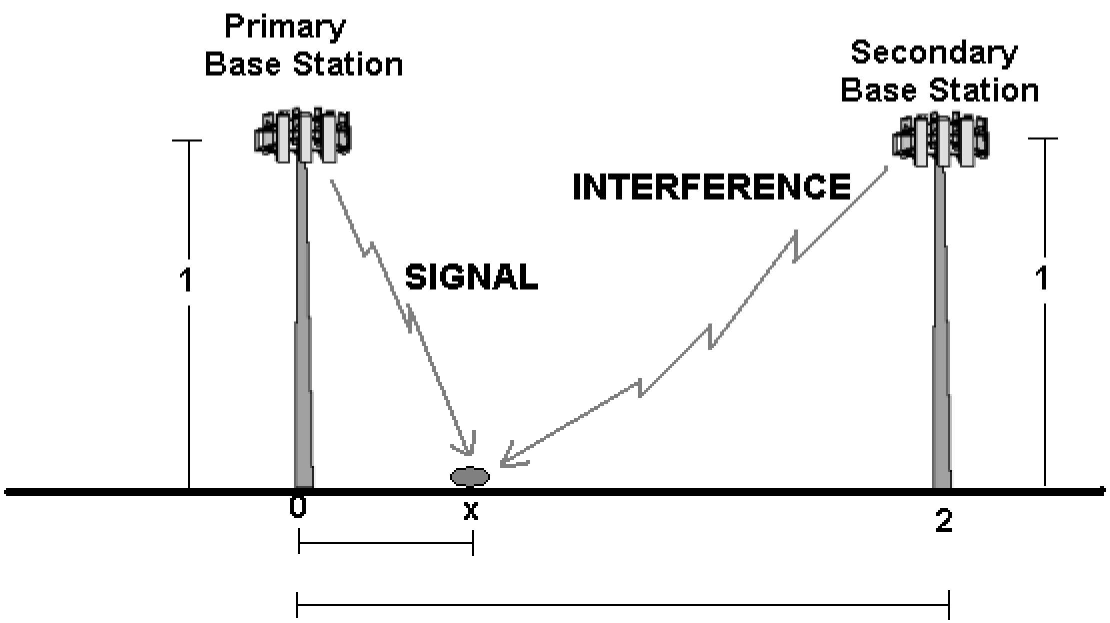
e.g., n=5



Note: Lines do not need to cross for signals from transmitters to interfere with each other. For example, signals from T1 and T4 interfere with each other to some degree, even though T1 has a “clear shot” to the receiver.

Setting the right power for each transmitter *i* is not so simple. Just putting out as much signal as possible , besides being costly, can drowned out other signals.

Signal to Interference Ratio (SIR)



A portion of a simplified cellular network.

Power Control Problem

Path gain from transmitter *j* to receiver *i* is $G_{ij} > 0$
("% sound that arrives to *j* from *i*")
Signal power at receiver *i* is $S_i = G_{ii}P_i$
Interference power at receiver *i* is $\mathrm{I}_i = \sum_{k \neq i} G_{ik}P_k$

The Signal to Interference Ratio (SIR) is

$$S_i / \mathrm{I}_i = G_{ii}P_i / \sum_{k \neq i} G_{ik}P_k$$

This is the question: How do we set transmitter powers to maximize the minimum SIR?

Power Control Problem

The same problem is to minimize the maximum interference to signal ratio, i.e., solve the problem :

Minimize $\max_i (\mathrm{AP})_i$ under the constraint P > 0 , where

$A_{ij} = G_{ij} / G_{ii}$ for $i \neq j$ and 0 for $i = j$

(So A is the matrix of ratios of interference to signal path gains)

A is positive except zeros on the diagonal ⇒ $A^2 > 0$ ⇒ A is primitive ⇒ the solution is given by the PF eigenvector of A

The PF eigenvalue λ of A is the optimal interference to signal ratio

i.e., the maximum possible minimum SIR is 1/λ .

With optimal power allocation, all SIRs are equal.

Economic Applications

Two (of many) applications are:

1. The filling gas stations example, which more generally appears as an optimization problem: Trying to allocate business resources appropriately.
2. An input-output model that is very important in both theory and practice called the Leontif Model.

We looked at 1., now let´s consider 2.

Leontief Input-Output Model

Professor Leontief originally created a system of 500 equations with 500 unknowns; however, at that time, such analysis was much too large for any computer to solve. As a result, he had to narrow it down to a 42 by 42 system. By creating this 42X42 matrix model,

Leontief attempted to answer the following question: What quantity should each sector in an economy produce so that it will be just adequate to meet the total demand for that product?

The main objective of the Leontief Input-Output Model is to equalize the total amount of goods produced and the total demand for that production; in other words, economists try to find X satisfying the following equation:

Production level= intermediate demand + final demand

$$\mathbf{x} = \mathbf{i} + \mathbf{d}$$

Econ App: Leontief I/O Model

In the Leontief Input-Output model, the economic system is assumed to have *n* industries with two types of demands on each industry: *final* demand (the customer is the final place for the product, also called *external* demand being a demand from outside the system) and *intermediate* demand (or *internal* demand – a demand placed on one industry by another within the system).

Economic App: Input-Output Model

So n = number of sectors of the economy

x = production vector = output of each sector

d = external demand = value of goods demanded from outside the system

i = internal demand = inputs required for production

For each sector, there is a unit consumption vector **c** that lists the inputs needed per unit of output. These make up a matrix C, and we define

$$\mathbf{i} = C\mathbf{x} \quad .$$

In the consumption matrix C, entry c_{ij} is the amount of input of a certain good *i* needed to produce one unit output of good *j*.

Economic App: Input-Output Model

$$
\begin{array}{l}
 & \text{Auto} & \text{Construction} & \text{Steel} \\
\begin{array}{l}\text{Automobile}\\ \text{Construction}\\ \text{Steel}\end{array} & \multicolumn{3}{l}{\begin{bmatrix} .15 & .25 & .22 \\ .20 & .14 & .11 \\ .46 & .18 & .25 \end{bmatrix}}
\end{array}
$$

.46 is the number of units of steel that the automobile industry needs to produce one car.

Econ App: Leontief I/O Model

Now we solve $\mathbf{x} = C\mathbf{x} + \mathbf{d}$:

$$\mathbf{x} = (\mathbf{I} - C)^{-1}\mathbf{d}$$

Even though C is nonnegative, **x** could have some negative values. That makes no sense in this model, or similar models where **x** is a supply vector—how much a factory/company/industry should make to meet outside demand.

Econ App: Leontief I/O Model

If we can assume that ρ(C) < 1, then $\sum_{i=0}^{\infty} \mathrm{C}^{\mathrm{i}}$ converges in operator norm.

(Gelfand's formula shows that if ρ(A)<1, then for some n, $\|A_n\|<1$. One can then rewrite the series as $(1+A+\cdots+A_{n-1})\sum_{i=0}^{\infty} \mathrm{A}^{\mathrm{ni}}$, which converges in norm.)

Econ App: Leontief I/O Model

The series $\sum_{i=0}^{\infty} \mathrm{C}^{\mathrm{i}}$ would be a Neumann series converging to $(\mathbf{I} - C)^{-1}$. Since C is nonnegative and **d** is nonnegative, **x** would be nonnegative.

To assume ρ(C) < 1, we need to make some mild assumptions, and then use the PFT.

Econ App: Leontief I/O Model

<u>Assumptions we make about the Matrix</u>

1. Any column or row does not add up to more than one. That makes sense, because a sector should require less than one unit's worth of input to produce one unit of output.

2. The matrix is irreducible. Assuming C to be irreducible is the same as saying that every industry directly or indirectly uses something produced by all other industries, which is reasonable between large (broad) industries.

Econ App: Leontief I/O Model

Now we have a nonnegative, irreducible matrix, with no row or column adding up to more than 1, so by the PFT, ρ(C) < 1, so $\sum_{i=0}^{\infty} \mathrm{C}^{\mathrm{i}}$ converges, and we can use the model to solve for minimal supply needed to meet demand.

Econ App: Leontief I/O Model

Comment

There are different versions of this model for different scales and purposes. The model can also be used in reverse in this sense:

Given a consumption matrix, if the columns (or rows) are less than 1, then that is a sufficient (not necessary--it can still be productive if we are able to find a vector $\mathbf{x}$ such that $\mathbf{x} > C\,\mathbf{x}$) condition for the matrix to be "productive".

Ranking

When all competitors don´t play each other, how to rank them?

College Football rankings and other sports in this situation use the PFT.

Let´s look at a soccer match and then at Google´s web page ranking as an example using the PFT to rank.

Ranking – World Cup

6 teams. 1 pt for win, ½ pt for tie, 0 for loss.

	Brazil	Canada	Colombia	Denmark	Ecuador	France	Total
Brazil	1/2	0	0	1	1	0	2.5
Canada	1	1/2	0	1	1	1	4.5
Colombia	1	1	1/2	0	1	1	4.5
Denmark	0	0	1	1/2	0	0	1.5
Ecuador	0	0	0	1	1/2	1	2.5
France	1	0	0	1	0	1/2	2.5

Ranking – World Cup

2 strange things:

1. Colombia and Canada are tied, even though Colombia won in a head to head contest.

2. Denmark beat #1 Colombia, but no extra credit for this.

Ranking – World Cup

$$\begin{array}{l} \text{Brazil} \\ \text{Canada} \\ \text{Colombia} \\ \text{Denmark} \\ \text{Ecuador} \\ \text{France} \end{array} \begin{bmatrix} .5 & 0 & 0 & 1 & 1 & 0 \\ 1 & .5 & 0 & 1 & 1 & 1 \\ 1 & 1 & .5 & 0 & 1 & 1 \\ 0 & 0 & 1 & .5 & 0 & 0 \\ 0 & 0 & 0 & 1 & .5 & 1 \\ 1 & 0 & 0 & 1 & 0 & .5 \end{bmatrix}$$

The first ranking obtained by comparing the row sums of the matrix. The .5 for playing oneself doesn´t affect anything yet.

COL=CAN > BRA=ECU=FRA > DEN

Ranking – World Cup

The first ranking is $R_1\,(i) = \sum_j A_{ij}$ (sum of the *j* columns of row *i*)

COL=CAN > BRA=ECU=FRA > DEN

The second ranking is $R_2\,(i) = \sum_j A_{ij}\;\; R_1\,(j)$

which equals $\sum_j (A_{ij}\;\; \sum_k A_{jk}\,)$

(This is just multiplication of two matrices.)

… $R_n\,(i) = \sum_j A_{ij}^{\;n}$ or $A^n\,\mathbf{1} = R_n$

Ranking – World Cup

<u>Results</u>:

R_1

4.5 2.5 1.5

COL=CAN > BRA=ECU=FRA > DEN

R_2

14.25 11.25 5.25

COL > CAN > BRA=ECU=FRA=DEN

R_3

34 27 17 13

COL > CAN > DEN>BRA=ECU=FRA

R_4

824 696 426 366

COL > CAN > DEN>BRA=ECU=FRA

Will the rankings keep changing, or eventually stabilize?

Ranking – World Cup

$$
\begin{array}{l}
\text{Brazil} \\
\text{Canada} \\
\text{Colombia} \\
\text{Denmark} \\
\text{Ecuador} \\
\text{France}
\end{array}
\begin{bmatrix}
.5 & 0 & 0 & 1 & 1 & 0 \\
1 & .5 & 0 & 1 & 1 & 1 \\
1 & 1 & .5 & 0 & 1 & 1 \\
0 & 0 & 1 & .5 & 0 & 0 \\
0 & 0 & 0 & 1 & .5 & 1 \\
1 & 0 & 0 & 1 & 0 & .5
\end{bmatrix}
$$

The matrix is nonnegative, with at least one nonnegative on the diagonal. By its digraph we see it´s irreducible, thus it is primitive.

Ranking – World Cup

By the PFT,

$$\lim_{n\to\infty}\left(\frac{\mathrm{A}}{\rho(\mathrm{A})}\right)^{\mathrm{n}} = \mathrm{xy}^{\mathrm{T}}$$

i.e., The final ranking is determined by the eigenvector associated to the largest eigenvalue.

In our example, this turns out to be $\begin{pmatrix} .27 \\ .53 \\ .64 \\ .32 \\ .27 \\ .27 \end{pmatrix}$ so the ranking

converges to **COL > CAN > DEN > BRA=ECU=FRA**.

Ranking – World Cup

There is a glitch if one team is undefeated or never victorious. Then the digraph is not strongly connected. e.g., if Denmark had not upset number 1 Colombia.

There´s also still this annoying fact…Denmark did win against Colombia. It violates the ranking. This is a “pairwise violation”. There are many complicated methods to ranking and often a goal is to minimize the pairwise violations. Finding *the* ranking that minimize the p.w. violations is NP-hard (= *very* hard).

Ranking – (American) Football

This gets more complicated because every team does not play every other team. If we can get an irreducible matrix, we can get primitivity as in the previous example by putting .5 on the diagonal, and then apply the PFT.

Getting to that point requires creating links where they are missing. One example of a way to make links is the sharing of the victory point (based on score). All-or-nothing is unidirectional, but sharing the points makes the graph more connected.

A *weighted* graph is more likely to be strongly connected.

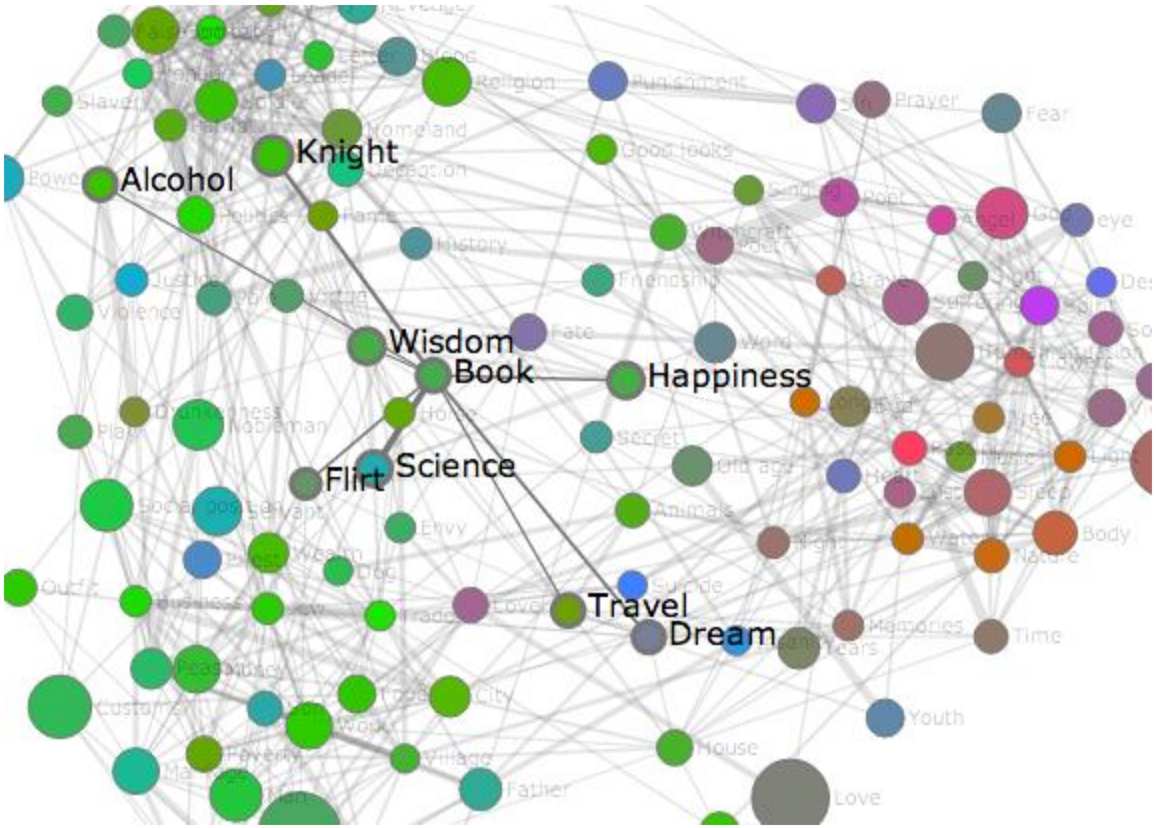
Ranking – Google

Reference: *Google's PageRank and Beyond: The Science of Search Engine Rankings* by Amy N. Langville and Carl D. Meyer.

This is based on their discussion in Chapters 3 and 4.

Ranking – WWW

Think of the internet as a directed graph.



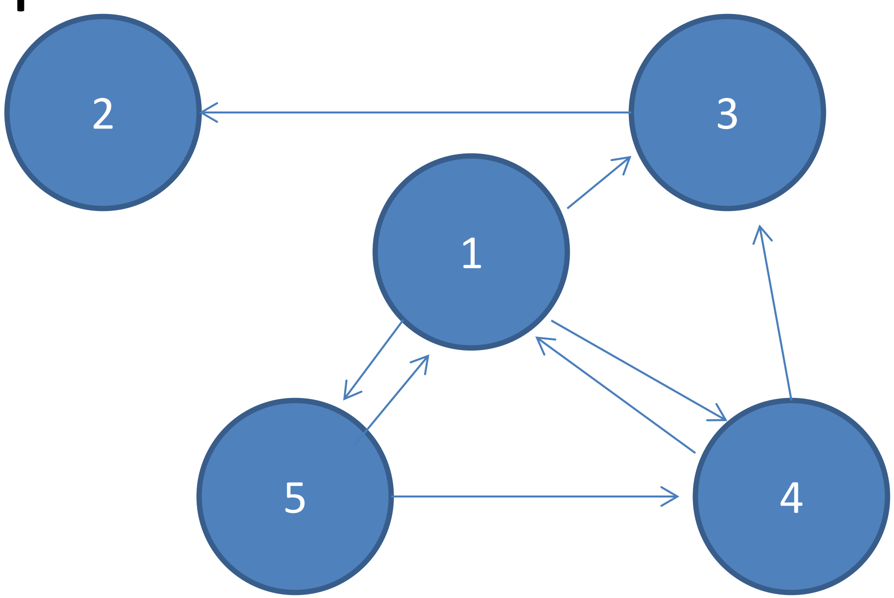
Ranking – WWW

Webpages are nodes

Links between pages are directed edges

# of links from a node = “out-degree” of that node

Here´s a part we´ll look at:



Ranking – WWW

We construct matrix H representing the digraph.

5 rows, 5 columns, each value is 1/(outdegree)

$$H = \begin{bmatrix} 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix}$$

Ranking – Google

Define the rank of a webpage as

$$r(P)= \sum_{Q \in B_P} \frac{r(Q)}{\deg(Q)}$$

deg(Q) is the outdegree out Q

B_P = the set of pages pointing to P

The rank of any website is proportional to the importance of the sites that link to it.

Ranking – Google

But the definition is recursive, so let´s define

$$r_{k+1}(P) = \sum_{Q \in B_P} \frac{r_k(Q)}{\deg(Q)}$$

We need a start point for the iteration, so take the uniform distribution $r_o(P) = 1/n$. We´ll show that it doesn´t matter what the start point is.

Ranking – Google

Define a rank vector **z**.

Then, e.g., $z_0 = (r_0(P_1), r_0(P_2),\ldots r_0(P_n))$

We can write $z_{k+1} = z_k H$ for $k = 0, 1, 2,\ldots$

(Note the nonzero elements on the *i*th row of H are just $\frac{1}{\deg(P_i)}$.)

So $z_k = z_0 H^k$ for $k = 1,2,3,\ldots$

$\lim_{k\to\infty}(z_0 H^k) = \boldsymbol{\pi}$ exists if H is primitive.

Ranking – Google

So that´s what Google does:

1. Make H a stochastic matrix, so that surfing the web becomes a Markov process.

2. Make some adjustments, if necessary so H>0

3. Apply the PFT.

Ranking – Google

Webpages that don´t link to others are called *dangling nodes*.

This puts a row of zeroes in the matrix, making it unstochastic.

We can just put the uniform distribution in place of the zeros. This is artificially creating links from that page to every page on the WWW.

Is this o.k.?

Yes. We have a Markov Process. $\Longrightarrow$

We don´t remember past states (surfing with no back button).

The next site you visit will be any site with equal probability.

Ranking – WWW

H becomes Stochastic

$$S\ \cancel{H} = \begin{array}{c} \begin{matrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 1/3 & 1/3 & 1/3 \\ \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} \\ 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix} \end{array}$$

Ranking – Google

Note:

There are better ways to replace the zeros. That´s one of many things that makes one search algorithm better than another.

Ranking – Google

In this example S is already primitive (strongly connected $\Longrightarrow$ irreducible, and with a non zero diagonal element), but that´s not always the case.

We add the *personalization matrix* (all 1/n entries), call it E.

E is stochastic, but the sum of E + S need not be.

To preserve the stochastic nature we add a linear combination α S + (1 - α) E = The Google Matrix G

Ranking – Google

Ways to tweak the ranking:

1. Different α
2. Different E

E must be stochastic (rows add to 1), but not necessarily uniform. Relative website popularity is a possible choice.

Low α means the hyperlink structure is important, high α means you are more likely to go anywhere randomly. (Google has used $\alpha = .15$)

The linear combination reflects how people surf.

Ranking – Google

G is stochastic and positive $\Longrightarrow$ the largest eigenvalue is 1, is simple, and no other eigenvalue lies on the spectral circle.

$\Longrightarrow$ there $\exists$ a stationary distribution,

$$\lim_{k \to \infty} (z_0 G^k) = \boldsymbol{\pi}$$

Ranking – Google

In our example, with uniform personalization matrix and $\alpha = .15$,

$\boldsymbol{\pi}$ = (.210 .25 .213 .189 .133),

which gives us the ranking

$P_2 > P_3 > P_1 > P_4 > P_5$

References

Google's PageRank and Beyond: The Science of Search Engine Rankings by Amy N. Langville and Carl D. Meyer.

Non-negative Matrices and Markov Chains by E. Seneta. Springer Statistical Series.

The Perron-Frobenius Theorem and the Ranking of Football Teams by James P. Keener SIAM Review March 1993.

and, of course, Wikipedia:

Perron Frobenius Theorem, Stochastic Matrices, Directed Graphs, Markov Chains, etc.