Markov chains, Eigenvector centrality & PageRank

What makes an important webpage?

In days of yore, the web was a small thing

Alta Vista was the search engine of choice

Google started in 1998, based on an algorithm (PageRank) described in a paper of Page, Brin, Motwani and Winograd

Overview

Give each page a rating (of its importance), a recursively defined measure whereby a page becomes important if important pages link to it

Recursive definition: the importance of a page refers back to the importance of other pages that link to it

Random surfer model: a random surfer on the web follows links from page to page. Page rank $\simeq \mathbb{P}$ random surfer lands on a particular page. Popular page $\implies$ higher probability to go there

Example of a Markov chain

Markov chain

A Markov chain is a *stochastic process* in which the evolution through time depends only on the current state of the system (we say the process is *memoryless*)

Markov chains are an interesting combination of matrix theory and graph theory

They form the theoretical foundation for Hidden Markov processes or Markov Chain Monte Carlo (MCMC) methods, are used in ML

Conduct an experiment with a set of r possible outcomes

$$S = \{S_1, \ldots, S_r\}$$

Experiment repeated n times (with n large, potentially infinite)

System has *no memory*: the next state depends only on the present state

Probability of S_j occurring on the next step, given that S_i occurred on the last step, is

$$p_{ij} = p(S_j|S_i)$$

Suppose that S_i is the current state, then one of $S_1, \ldots, S_r$ must be the next state; so

$$p_{i1} + p_{i2} + \cdots + p_{ir} = 1, \quad 1 \leq i \leq r$$

(Some of the p_{ij} can be zero, all that is needed is that $\sum_{j=1}^{r} p_{ij} = 1$ for all i)

Definition 1

An experiment with finite number of possible outcomes $S_1, \ldots, S_r$ is repeated. The sequence of outcomes is a **Markov chain** if there is a set of r^2 numbers $\{p_{ij}\}$ such that the conditional probability of outcome S_j on any experiment given outcome S_i on the previous experiment is p_{ij}, i.e., for $1 \leq i, j \leq r$, $n = 1, \ldots,$

$$p_{ij} = \Pr(S_j \text{ on experiment } n+1 \mid S_i \text{ on experiment } n)$$

Outcomes $S_1, \ldots, S_r$ are **states** and p_{ij} are **transition probabilities**. $P = [p_{ij}]$ the **transition matrix**

The matrix

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ & & & \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix}$$

has

- nonnegative entries, $p_{ij} \geq 0$
- entries less than 1, $p_{ij} \leq 1$
- row sum 1, which we write

$$\sum_{j=1}^{r} p_{ij} = 1, \quad i = 1, \ldots, r$$

or, using the notation $\mathbb{1}^T = (1, \ldots, 1)$,

$$P\mathbb{1} = \mathbb{1}$$

(super simple) Mendelian genetics

A *phenotypic trait* (eye colour, hair colour, etc.) is determined by a specific pair of alleles, each of which may be two types, say G and g

Each individual can have

- GG combination (*dominant*)
- Gg or gG, considered equivalent genetically (*hybrid*)
- gg combination (*recessive*)

Individuals bearing GG or gg alleles are *homozygotes*, hybrids with Gg alleles are called *heterozygotes*

GG and gg combinations lead to different phenotypes, Gg combination leads to expressing the same phenotype as individuals bearing a GG combination, hence the name dominant given to GG

In sexual reproduction, offspring inherit one allele of the pair from each parent

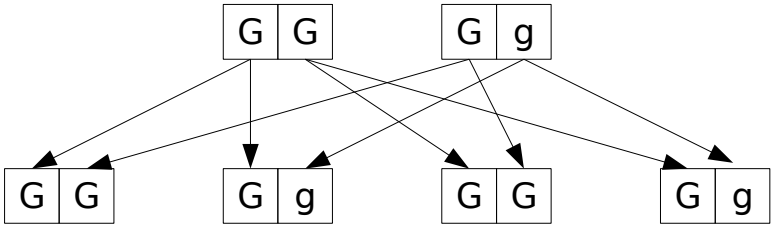
Alleles inherited from each parent are selected at random, independently of each other

This determines probability of occurrence of each type of offspring. The offspring

- of two GG parents must be GG
- of two gg parents must be gg
- of one GG and one gg parent must be Gg
- other cases must be examined in more detail

GG and Gg parents

Suppose one parent GG and the other Gg



		Parent 1	
		G	G
Parent 2	G	GG	GG
	g	Gg	Gg

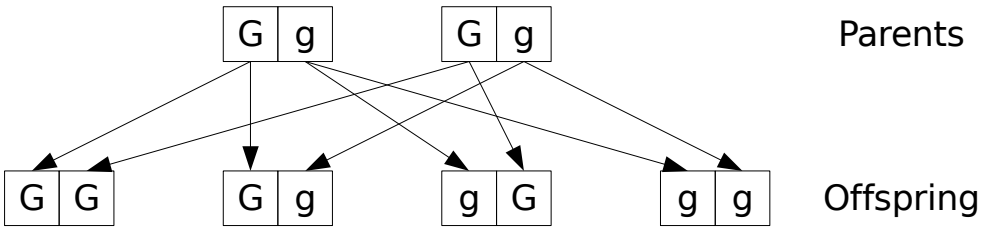
To determine $\mathbb{P}$ that offspring is of a certain type, count number of outcomes of each type (GG and Gg) and divide by 4

$\implies$ offspring have probability

- 1/2 of being GG
- 1/2 of being Gg

Gg and Gg parents

Both parents are hybrid

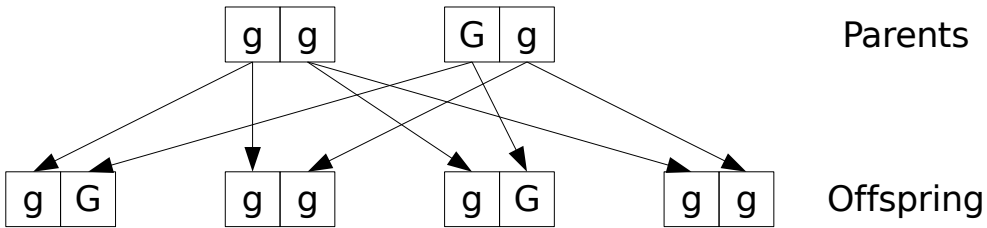


$\Longrightarrow$ offspring have probability

- 1/4 of being GG
- 1/2 of being Gg
- 1/4 of being gg

gg and Gg parents

Recessive and hybrid parents



$\Longrightarrow$ offspring have probability

- 1/2 of being Gg
- 1/2 of being gg

General case

$p_i(n)$: probability that state S_i occurs on the n^{th} repetition of the experiment, $1 \leq i \leq r$

Since one the states S_i must occur on the n^{th} repetition

$$p_1(n) + p_2(n) + \cdots + p_r(n) = 1$$

$p_i(n+1)$: probability that state S_i, $1 \leq i \leq r$, occurs on $(n+1)^{th}$ repetition of the experiment

r ways to be in state S_i at step $n+1$:

1. Step n is S_1. Probability of getting S_1 on n^{th} step is $p_1(n)$, and probability of having S_i after S_1 is p_{1i}. Therefore $P(S_i|S_1) = p_{1i}p_1(n)$
2. We get S_2 on step n and S_i on step $(n+1)$. Then $P(S_i|S_2) = p_{2i}p_2(n)$

..

r. Probability of occurrence of S_i at step $n+1$ if S_r at step n is $P(S_i|S_r) = p_{ri}p_r(n)$

$$\begin{aligned} \implies p_i(n+1) &= P(S_i|S_1) + \cdots + P(S_i|S_r) \\ &= p_{1i}p_1(n) + \cdots + p_{ri}p_r(n) \end{aligned}$$

Therefore,

$$\begin{aligned} p_1(n+1) &= p_{11}p_1(n) + p_{21}p_2(n) + \cdots + p_{r1}p_r(n) \\ &\vdots \\ p_r(n+1) &= p_{1r}p_1(n) + p_{2r}p_2(n) + \cdots + p_{rr}p_r(n) \end{aligned}$$

In matrix form

$$p(n+1) = p(n)P, \quad n = 1, 2, 3, \ldots$$

where $p(n) = (p_1(n), p_2(n), \ldots, p_r(n))$ is a (row) probability vector and $P = (p_{ij})$ is a $r \times r$ *transition matrix*,

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ & & & \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix}$$

So

$$(p_1(n+1), \ldots, p_r(n+1))$$
$$= (p_1(n), \ldots, p_r(n)) \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ & & & \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix}$$

Easy to check that this gives the same expression as before

Stochastic matrices

Definition 2 (Stochastic matrix)

The nonnegative $r \times r$ matrix M is **stochastic** if $\sum_{j=1}^{r} a_{ij} = 1$ for all $i = 1, 2, \ldots, r$

If each row sum and each column sum equals one, the matrix is **doubly stochastic**

Theorem 3

Let M be a stochastic matrix. Then all eigenvalues λ of M are such that $|\lambda| \leq 1$. Furthermore, $\lambda = 1$ is an eigenvalue of M

Long time behaviour

Let $p(0)$ be the initial distribution (row) vector. Then

$$\begin{aligned} p(1) &= p(0)P \\ p(2) &= p(1)P \\ &= (p(0)P)P \\ &= p(0)P^2 \end{aligned}$$

Iterating, we get, for any n,

$$p(n) = p(0)P^n$$

Therefore,

$$\begin{aligned} \lim_{n\to+\infty} p(n) &= \lim_{n\to+\infty} p(0)P^n \\ &= p(0) \lim_{n\to+\infty} P^n \end{aligned}$$

if this limit exists

$$\lim_{n\to+\infty} p(n) = p(0) \lim_{n\to+\infty} P^n$$

Does the limit exist?

Theorem 4

If M, N are nonsingular stochastic matrices, then MN is a stochastic matrix

So the product of any number of stochastic matrices is also stochastic

Corollary 5

If M is a nonsingular stochastic matrix, then for any $k \in \mathbb{N}$, M^k is a stochastic matrix

Regular Markov chains

Definition 6 (Regular Markov chain)

A **regular** Markov chain has P^k (entry-wise) positive for some integer $k > 0$, i.e., P^k has only positive entries

Definition 7 (Primitive matrix)

A nonnegative matrix M is **primitive** if, and only if, there is an integer $k > 0$ such that M^k is positive.

Theorem 8

Markov chain regular $\iff$ *transition matrix P primitive*

Behaviour of a regular MC

Theorem 9

If P is the transition matrix of a regular Markov chain, then

1. *the powers P^n approach a stochastic matrix W*
2. *each row of W is the same (row) vector $w = (w_1, \ldots, w_r)$*
3. *the components of w are positive*

So if the Markov chain is regular

$$\lim_{n \to +\infty} p(n) = p(0) \lim_{n \to +\infty} P^n = p(0)W$$

Computing W

If $p(n)$ converges, then $p(n+1) = p(n)P$, so w is a **fixed point** of the system. Write

$$wP = w$$

and solve for w, i.e., find w as left eigenvector corresponding to the eigenvalue 1 or as (right) eigenvector associated to eigenvalue 1 for the transpose of P

$$P^T w^T = w^T$$

w might have to be normalized (you want a probability vector). Check that the norm $\|w\|$ defined by

$$\|w\| = w_1 + \cdots + w_r$$

is equal to one. If not, use

$$\frac{w}{\|w\|}$$

Back to genetics

Suppose we want to understand what it means to have hybrid individuals in the population

Investigate this using a process of continued matings

- Start with an individual of known or unknown genetic character (dominant, hybrid or recessive) and mate it with a hybrid
- Assume that the mating results in at least one offspring; choose one of the offspring at random and mate it with a hybrid
- Repeat this process through a number of generations

What can we expect in terms of the genetic composition of the population after a while?

$\Longrightarrow$ consider MC with states GG, Gg and gg

3 states: $S_1 = \mathrm{GG}$, $S_2 = \mathrm{Gg}$ and $S_3 = \mathrm{gg}$; we use GG, Gg and gg as well to name the states

↗	GG	Gg	gg
GG	0.5	0.5	0
Gg	0.25	0.5	0.25
gg	0	0.5	0.5

The transition probabilities are thus

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

so

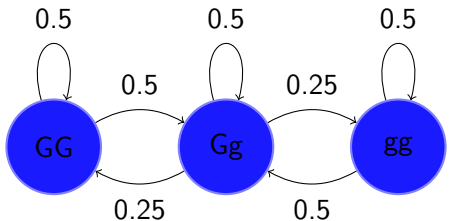
$$P^2 = \begin{pmatrix} \frac{3}{8} & \frac{1}{2} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{8} & \frac{1}{2} & \frac{3}{8} \end{pmatrix}$$

$\Longrightarrow$ P primitive $\Longrightarrow$ Markov chain regular

Theorem 10

M primitive if the associated connection graph is strongly connected and there is at least one positive entry on the diagonal of M

This is checked directly on the transition graph



Compute left eigenvector associated to 1

$$(w_1, w_2, w_3) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} = (w_1, w_2, w_3)$$

$$\begin{aligned} \frac{1}{2}w_1 + \frac{1}{4}w_2 &= w_1 \\ \frac{1}{2}w_1 + \frac{1}{2}w_2 + \frac{1}{2}w_3 &= w_2 \\ \frac{1}{4}w_2 + \frac{1}{2}w_3 &= w_3 \end{aligned}$$

So $w_1 = w_2/2$, $w_3 = w_2/2$ and thus

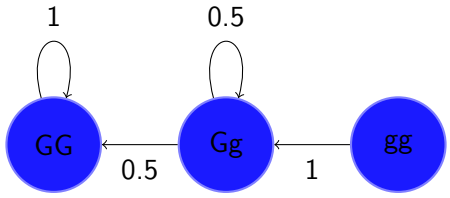
$$\frac{1}{4}w_2 + \frac{1}{2}w_2 + \frac{1}{4}w_2 = w_2,$$

that is, $w_2 = w_2$, i.e., w_2 can take any value

$$\Longrightarrow \quad w = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$$

Changing the setting of the genetic experiment

Suppose now the same type of experiment, but mate each new generation with a GG individual instead of a Gg individual



↗	GG	Gg	gg
GG	1	0	0
Gg	0.5	0.5	0
gg	0	1	0

$$P = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

- leave gg after 1 iteration and can never return
- when we leave Gg, we can never return
- we can never leave GG when we get there

Absorbing Markov chains

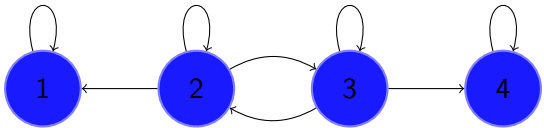
Definition 11 (Absorbing state)

A state S_i in a Markov chain is **absorbing** if whenever it occurs on the n^{th} generation of the experiment, it then occurs on every subsequent step. In other words, S_i is absorbing if $p_{ii} = 1$ and $p_{ij} = 0$ for $i \neq j$

Definition 12 (Absorbing chain)

A Markov chain is **absorbing** if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state. In an absorbing Markov chain, a state that is not absorbing is called **transient**

Suppose we have a chain like the following



1. Does the process eventually reach an absorbing state?
2. What is the average number of steps spent in a transient state, if starting in a transient state?
3. What is the average number of steps before entering an absorbing state?
4. What is the probability of being absorbed by a given absorbing state, when there are more than one, when starting in a given transient state?

The answer to the first question ("Does the process eventually reach an absorbing state?") is given by the following result

Theorem 13

In an absorbing Markov chain, the probability of reaching an absorbing state is 1

To answer the other questions, write the transition matrix in **standard** form

For an absorbing chain with k absorbing states and $r - k$ transient states, write transition matrix as

$$P = \begin{pmatrix} \mathbb{I}_k & \mathbf{0} \\ R & Q \end{pmatrix}$$

with following meaning

	Absorbing states	Transient states
Absorbing states	$\mathbb{I}_k$	$\mathbf{0}$
Transient states	R	Q

with $\mathbb{I}_k$ the $k \times k$ identity matrix, $\mathbf{0}$ an $k \times (r-k)$ matrix of zeros, R an $(r-k) \times k$ matrix and Q an $(r-k) \times (r-k)$ matrix. The matrix $\mathbb{I}_{r-k} - Q$ is invertible. Let

- $N = (\mathbb{I}_{r-k} - Q)^{-1}$ the **fundamental matrix** of the MC
- T_i sum of the entries on row i of N
- $B = NR$

Answers to our remaining questions:

2. N_{ij} average number of times the process is in the jth transient state if it starts in the ith transient state

3. T_i average number of steps before the process enters an absorbing state if it starts in the ith transient state

4. B_{ij} probability of eventually entering the jth absorbing state if the process starts in the ith transient state

Back to the genetic example

The matrix is already in standard form

$$P = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} \mathbb{I}_1 & \mathbf{0} \\ R & Q \end{pmatrix}$$

with $\mathbb{I}_1 = 1$, $\mathbf{0} = (0 \;\; 0)$ and

$$R = \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix} \qquad Q = \begin{pmatrix} \frac{1}{2} & 0 \\ 1 & 0 \end{pmatrix}$$

We have

$$\mathbb{I}_2 - Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ -1 & 1 \end{pmatrix}$$

so

$$N = (\mathbb{I}_2 - Q)^{-1} = 2 \begin{pmatrix} 1 & 0 \\ 1 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix}$$

Then

$$T = N\mathbb{1} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

2. N_{ij} average number of times the process is in the jth transient state if it starts in the ith transient state

$$N = \begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix}$$

3. T_i average number of steps before the process enters an absorbing state if it starts in the ith transient state

$$T = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

4. B_{ij} probability of eventually entering the jth absorbing state if the process starts in the ith transient state

$$B \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Constructing a stochastic matrix from an adjacency matrix

Let A be the adjacency matrix of a simple graph $G = (V, E)$ and D its degree matrix, i.e., the diagonal matrix $D = (d_{ij})$ with diagonal entries

$$d_{ii} = \sum_{j=1}^{n} a_{ji} = \sum_{j=1}^{n} a_{ij}$$

(Recall that A symmetric since G nondirected.) Then the matrix AD^{-1} is column stochastic

Indeed, let $\mathbb{1} = (1, \ldots, 1)^T$, then

$$\mathbb{1}^T AD^{-1} =$$

Eigenvector centrality (undirected graph)

Let $\mathbf{x}$ be an eigenvector corresponding to the largest eigenvalue λ of the non-negative adjacency matrix A of the undirected graph $G = (V, E)$. (We often call λ the **Perron root** of A and $\mathbf{x}$ a **Perron eigenvector**.)

The **eigenvector centrality** (or **prestige score**) of vertex i is the ith component of the eigenvector $\mathbf{x}$ of the (column) stochastic matrix $N := AD^{-1}$ corresponding to the eigenvalue 1:

$$N\mathbf{x} = \mathbf{x}$$

Consider a particular vertex i with its neighbouring vertices $\mathcal{N}(i)$:

$$x_i = \sum_{j \in \mathcal{N}(i)} x_j = \sum_j A_{ij} x_j$$

The eigenvector centrality defined this way depends both on the number of neighbours $|\mathcal{N}(i)|$ and the quality of its connections x_j, $j \in \mathcal{N}(i)$

Let $A = (a_{ij})$ be the adjacency matrix of a graph. The eigenvector centrality x_i of vertex i is given by

$$x_i = \frac{1}{\lambda} \sum_k a_{k,i} \, x_k$$

where $\lambda \neq 0$ is a constant. In matrix form

$$\mathbf{x}^T A = \lambda \mathbf{x}^T$$

Hence the centrality vector $\mathbf{x}$ is the left eigenvector of the adjacency matrix A associated with the eigenvalue λ.

The power method can be used to solve the eigenvector centrality problem. Let $m(v)$ denote the signed component of maximal magnitude of vector v. If there is more than one maximal component, let $m(v)$ be the first one. For instance, $m(-3, 3, 2) = -3$.
Let $x^{(0)}$ be an arbitrary vector. For $k \geq 1$:
repeatedly compute $x^{(k)} = x^{(k-1)}A$; normalize $x^{(k)} = x^{(k)}/m(x^{(k)})$; until the desired precision is achieved. It follows that $x^{(k)}$ converges to the dominant eigenvector of A and $m(x^{(k)})$ converges to the dominant eigenvalue of A. If matrix A is sparse, each vector-matrix product can be performed in linear time in the size of the graph.
The method converges when the dominant (largest) and the sub-dominant (second largest) eigenvalues of A, respectively denoted by λ_1 and λ_2, are separated, that is they are different in absolute value, hence when $|\lambda_1| > |\lambda_2|$. The rate of convergence is the rate at which $(\lambda_2/\lambda_1)^k$ goes to 0. Hence, if the sub-dominant eigenvalue is small compared to the dominant one, then the method quickly converges.

Why use the leading eigenvector?

We want a nonnegative measure, so we want a vector in $\mathbb{R}_+$

We know from the Perron-Frobenius Theorem that the eigenvector corresponding to the dominant eigenvalue of a nonnegative matrix is nonnegative

Furthermore, if the graph is strongly connected, the matrix is irreducible and the eigenvector corresponding to the dominant eigenvalue is *positive*

PageRank

Variant of the Eigenvector centrality measure for directed network

Basic PageRank

- Whenever a vertex i has no outgoing link, we add a self-loop to i such that $k_i^{in} = k_i^{out} = 1$. Therefore $A_{ii} = 1$ for such vertices in the adjacency matrix
- Let D^+ be the diagonal matrix of outdegrees where each element $D_{ii}^+ = k_i^{out}$
- Define a column stochastic matrix $N = A(D^+)^{-1}$
- The PageRank centrality of node i is equal to the eigenvector xi of matrix N (The leading eigenvalue is 1): x = Nx