Lecture 31: Markov chains

(example of a stochastic process), It basically means random variables evolving over time.

$X_0, X_1, X_2, \ldots$ (think of X_n as state of system at (discrete) time n) $\rightarrow$ non-negative integer

Stat 171 (stochastic processes)

(Here, we have finitely many states and each of above X's is one of those states and we ~~just~~ have this process that's bouncing around randomly from state to state.)

Markov property

$$P(X_{n+1}=j \mid X_n=i, X_{n-1}=i_{n-1}, X_{n-2}=i_{n-2}, \ldots, X_0=i_0)$$

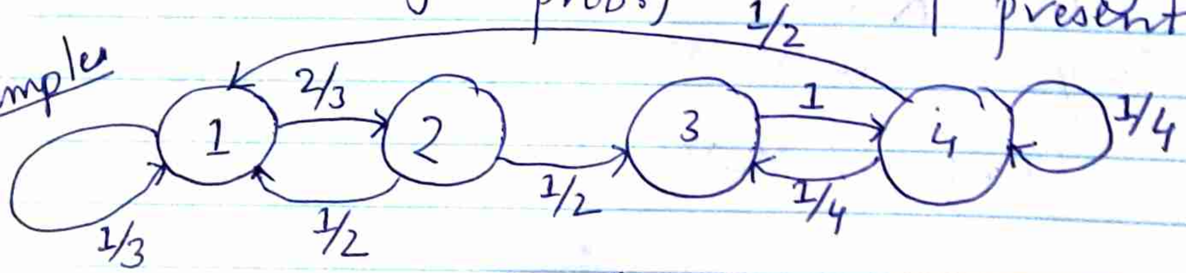
$$= P(X_{n+1}=j \mid X_n=i)$$

"homogeneous"

$= q_{ij}$ (transitional prob.)

past, future are conditionally indep. given present.

Example



transition matrix: $Q = (q_{ij})$

$$= \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \\ 1/2 & 0 & 1/4 & 1/4 \end{pmatrix}$$

Each row sums to 1.

Markov chain Monte Carlo (MCMC)

The idea of MCMC is that we don't have to worry about whether the actual process we are observing follows a Markov chain. MCMC means we synthetically construct our own markov chain, that will converge to a distribution that we are interested in.

Example

Suppose we are trying to simulate some complicated system and the computations are too hard to do everything analytically and explicitly. There are some extremely clever ways to construct a Markov Chain synthetically, that will converge to the thing we are interested in.

We program our Markov chain on the computer, run the chain for a long time and then use those result to study the distribution that we are interested in.

✓ Suppose at time n, X_n has distribution $\vec{s}$ (row vector) (PMF since discrete) $(1 \times M)$ matrix

$P(X_{n+1}=j) = ?$

(i.e., PMF at time $n+1$, one step in future) PMF listed out at time n

Solu: $$P(X_{n+1}=j) = \sum_i P(X_{n+1}=j \mid X_n=i)\, P(X_n=i) \quad \text{(from LOTP)}$$

$$= \sum_i q_{ij}\, s_i \text{ is the } j\text{th entry of } \underset{(1\times M)(M\times M)}{\vec{s}Q}$$

or $\sum_i s_i q_{ij}$

So, $\vec{s}Q$ is the distribution at time $n+1$.

Therefore, $\vec{s}Q^2$ is the distribution at time $n+2$,

$\vec{s}Q^3$ is the distribution at time $n+3$, ...

(3 steps in the future)

$$P(X_{n+1}=j \mid X_n=i) = q_{ij}$$

$$P(X_{n+2}=j \mid X_n=i) = \sum_k P(X_{n+2}=j \mid X_{n+1}=k, X_n=i)\, P(X_{n+1}=k \mid X_n=i)$$

$$= \sum_k P(X_{n+2}=j \mid X_{n+1}=k)\, P(X_{n+1}=k \mid X_n=i)$$

(A/c to Markov Property)

$$= \sum_k q_{kj}\, q_{ik}$$

$$= \sum_k q_{ik}\, q_{kj} \text{ is } (i,j) \text{ entry of } Q^2.$$

$\vdots$

Similarly,

$$P(X_{n+m}=j \mid X_n=i) = (i,j) \text{ entry of } Q^m.$$

Stationary distribution (Also, called a steady stay or long line or equilibrium)

$\vec{s}$ (prob. vector $1 \times M$) is stationary for the Markov chain that we are considering, if

$$\underset{1\times M}{\vec{s}}\ \underset{M\times M}{Q} = \vec{s}.$$

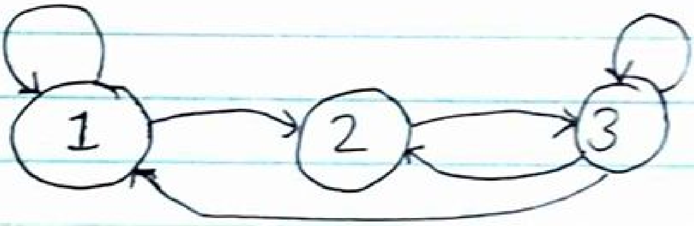
Intuition - If the chain at time n follows the distribution $\vec{s}$, then $\vec{s}Q$ is the distribution one step later in time. So, if $\vec{s}Q$ equals $\vec{s}$, it says if it starts out according to $\vec{s}$, ~~the~~ one step forward in time, it's still following $\vec{s}$. That's why, it's called stationary because it didn't change. And, we go another step forward in time, the distribution is still $\vec{s}$, and so on.

So, if the chain starts out of the stationary distribution, it will have the stationary distribution forever. It doesn't change. That's why it's called stationary.

<u>Questions:</u> (1) Does a stationary distribution exist?
Can we solve the equation $\vec{s}Q = \vec{s}$?
(2) Is it unique?
(3) Does the chain converge to $\vec{s}$ in some sense?
(4) How can we compute it?

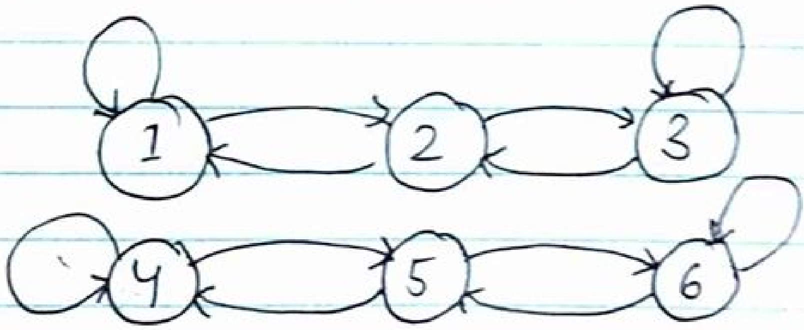
Lecture 32: Markov chains continued

(1)



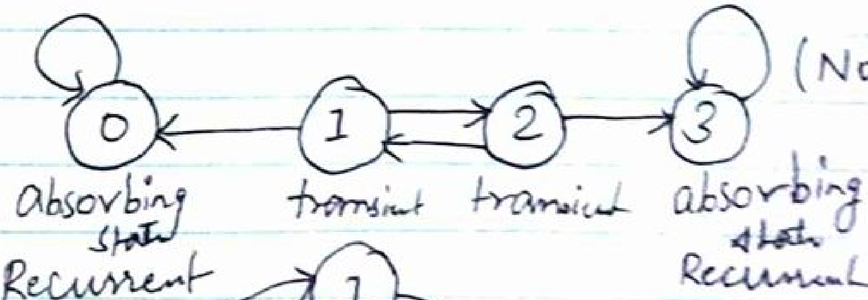
(irreducible)
All states are recurrent

(2)



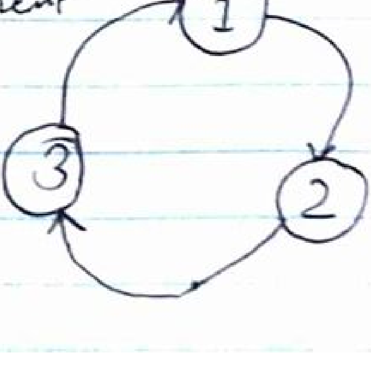
(Not irreducible)

(3)



(Not irreducible)
(gambler's Ruin drawn as Markov chain)

(4)



(irreducible)

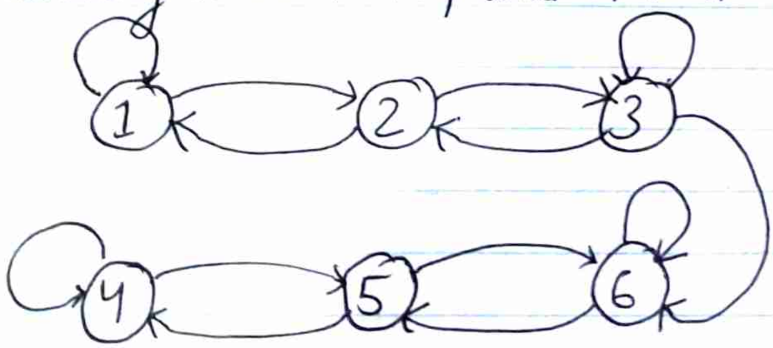
periodic chain
Recurrent

A Chain is irreducible if possible (with positive prob.) to get from anywhere to anywhere, not necessary in one step but in some finite number of steps.

A state is recurrent if starting there, chain has prob. 1 of returning to that state.
Otherwise, transient.

Note: In the irreducible case, if there's a finite number of states, all the states are ~~goining~~ going to be Recurrent.

Let's say we start at a given state, wander around, get back to the same starting state. Since, it's Markov, then we no longer care about the whole pass history. It's the same problem again. So, it's prob. 1 that it will come back again, and it will probably will come back again. So, if it's recurrent, it will come back infinitely often. On the other hand, if it's transient then it might come back again and again for a while but eventually it will stop and it will never go back again.

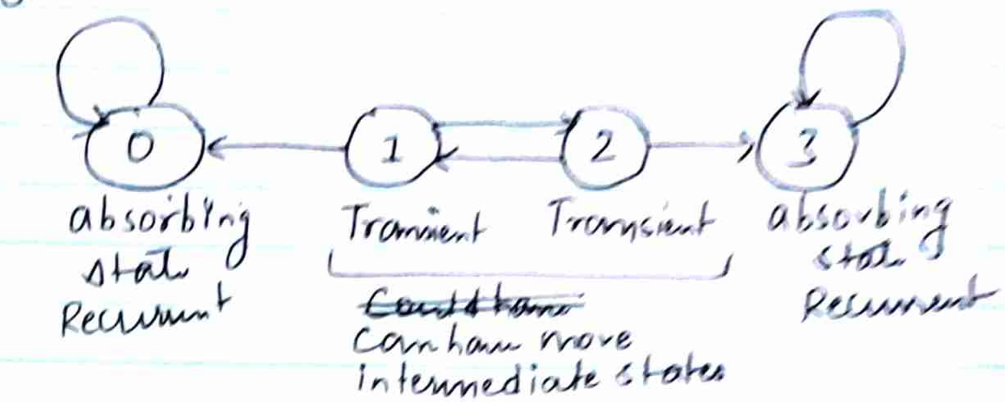


Reducible
or not irreducible

But, now states 1, 2 and 3 have become transient. Let's suppose we start at 1, 2 or 3, we may wander around in the upper chain for years and years but eventually we will transition from state 3 to 6 in the lower chain. Then, there is no turning back in the upper chain.

So, if we add edge from state 2 to 6, then states 1, 2 and 3 will be transient ~~while states~~ 4, 5 and 6 ~~would~~ will be recurrent.

Gambler's ruin drawn as Markov chain



Above is a gambler's ruin problem visualized as a Markov chain, which is saying how much money does gambler A have at a certain time? Wanders around at some point and then it eventually ends with either gambler ~~is~~ A is bankrupt and he stays bankrupt forever in the problem or gambler A has all the money. Gambler B is bankrupt and then it stays that way forever.

Stationary Distribution

$\vec{s}$ (prob. row vector) is stationary for a chain with transition matrix Q if

$$\vec{s}Q = \vec{s}$$

Note :- If pick a random state that's distributed according to $\vec{s}$, where $\vec{s}$ is not necessarily stationary yet. ~~and then~~ Then, $\vec{s}Q$ says

What's the probability distribution over states one step later.

$SQ^2 \longrightarrow$ prob. distribution over states two steps later.

<u>Theorem</u>: for any irreducible Markov chain (with finitely many states):

(1) A stationary distribution $\vec{S}$ exists.

(2) It's unique. $\longrightarrow$ there exists a unique stationary distribution even if there is a trillion states.

(3) $S_i = 1/r_i$,

where r_i is average return time (i.e. how many steps does it take to return to i if the chain starts at i.

expected value

One way to think of stationary distribution intuitively is that it's the long run fraction of time of being in a certain state.

So, think of S_i as if we run the chain for long, long time and we say what fraction of times was it inhabiting state i. That's going to converge to S_i under some mild conditions. So, S_i is the long run fraction of time at state i.

<u>Example</u>

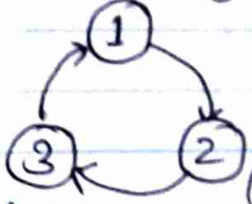
$S_i = 1/10$

If the chain is at i, $1/10$ th of the time in long run, it says if we start at state i then on average it'll take 10 steps to get back to state i.

(4) If Q^m is strictly positive for some m, (no periodicity)
→ also implies irreducibility

(m step transition probability, i.e., probability of going from somewhere to somewhere in exactly m steps)

We can find one value of m, such that we don't have any 0 in ~~this~~ matrix (Q^m), that will rule out any kind of periodicity problem.



transition matrix 3x3

(periodic)

In this, we can never find one power of the transition matrix here where all entries are positive.
(0 is neither negative nor positive)

then $P(X_n = i) \longrightarrow s_i$ as $n \to \infty$.

It means, no matter what the initial condition is, in the long run (as $n \to \infty$), the distribution (PMF) at time n converges to the stationary distribution.

In matrix terms,

$\vec{t}Q^n \longrightarrow \vec{s}$ as $n \to \infty$, where $\vec{t}$ is just any probability vector, not necessarily the stationary one. (Initially, we choose a random state where the probability is given by t. If we want to be deterministic, then just make t = 1, and everything else 0.)

→ It's just saying, start up the chain however we want, then let long time elapse and it converges to the stationary distribution.

The only difficulty with this theorem is that it doesnot give us much of a clue for how to compute it.
We can use (3) to compute it but then we have to find r_i which is also a difficult problem.

Reversible Markov chains

refers to time, hence also called time reversible

Defn. Markov chain with transition matrix $Q = (q_{ij})$ is reversible if there is a prob. vector $\vec{S}$ such that

$$\boxed{S_i q_{ij} = S_j q_{ji}} \text{ for all states } i, j.$$

Theorem: If reversible with respect to $\vec{S}$, then $\vec{S}$ is stationary.

Intuition: Reversible is also called time reversible. The reversibility refers to time. It says that if we start up the chain with distribution $\vec{S}$ and then imagine like recording a video (imagine one of ~~these fu~~ the previous 4 pictures/chains and we kind of videotaping ~~#~~ a particle which is wandering around in this process), then reversibility says if ~~you~~ we took that tape and played it backwards in time, we reversed it and we showed that to someone else, they would not know whether it was going forwards or backwards in time. It looks the same.
So, $S_i q_{ij} = S_j q_{ji}$ is saying if we run time forwards or backwards, it looks the same.

In physics, $\boxed{s_i q_{ij} = s_i q_{ji}}$ is called detailed balance, synonym for reversibility

Let $s_i q_{ij} = s_i q_{ji}$ for all i,j, show $\vec{s}Q = \vec{s}$.

$(s_1 s_2 \dots s_M)Q$

proof:

$$\sum_i s_i q_{ij} = \sum_i s_j q_{ji}$$

($\because$ $s_i q_{ij} = s_i q_{ji}$, so the sums are equal)

(definition of matrix multiplication) $= s_j \sum_i q_{ji}$ = prob. of going from state j to state i, summed over all states i. This sum is just 1, because it has to go somewhere. (each row of Q has to add up to 1)

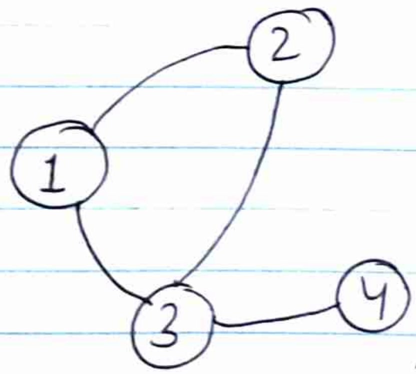
$= s_j$

$$\boxed{\text{So, } \vec{s}Q = \vec{s}}$$

Hence, reversibility implies stationarity.

Example (Reversible Markov chain) Random walk on an (undirected) network.

irreducible



Random walk just says imagine that we start out at one of these states and wherever we are, look at all of our available edges and we pick one randomly with equal probabilities.

Let d_i be degree of i. [$d_1=2$, $d_2=2$, $d_3=3$, $d_4=1$]

Then,

$d_i q_{ij} = d_j q_{ji}$ for all i,j.

Let $i \neq j$.

q_{ij}, q_{ji} are both 0, or both not 0.

If $\{i,j\}$ is an edge, then

$$q_{ij} = 1/d_i$$

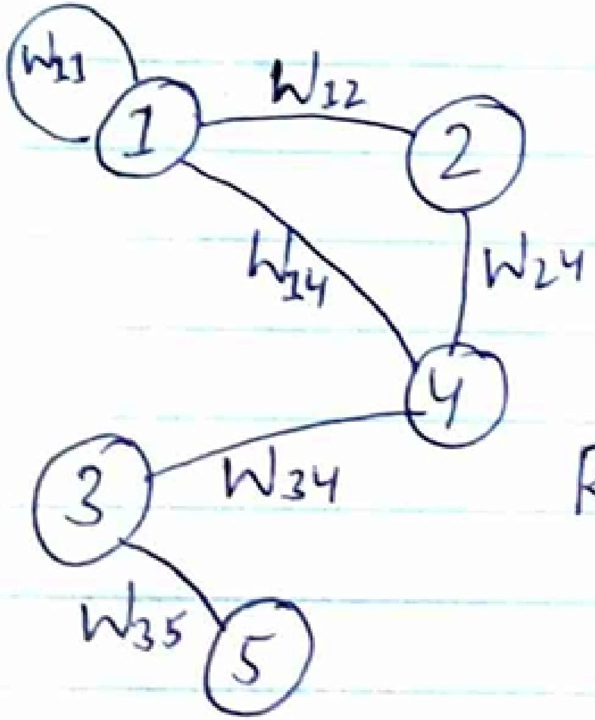
$$\therefore \cancel{d_i} \times \frac{1}{\cancel{d_i}} = \cancel{d_j} \times \frac{1}{\cancel{d_j}} \quad [d_i q_{ij} = d_j q_{ji}]$$

$$\Rightarrow 1 = 1 \text{ (always true)}$$

So, with M nodes $1, 2, \ldots, M$, degree d_i, then $\vec{S}$ with $S_i = \dfrac{d_i}{\sum_j d_j}$ is stationary.

in the long run, it will spend more time in the states of higher degree

Lecture 33: Markov chains Continued Further



$w_{ij} \geq 0$ "weight",
$w_{ij} = 0$ if no edge $\{i, j\}$
$w_{ij} = w_{ji}$ (assumption) [Symmetry]
Undirected network

Random walk: from state i, go to j with prob. $\propto w_{ij}$.
(proportionality symbol)

(So, we look at all the available choices and choose ~~what p~~ with probability proportional to the weights we see.)

If $\{i,j\}$ is an edge, then $q_{ij} = \dfrac{w_{ij}}{\sum_k w_{ik}}$.

Note: $\left(\sum_k w_{ik}\right) q_{ij} = w_{ij}$
$= w_{ji}$
$= \left(\sum_k w_{jk}\right) q_{ji}$ (Reversibility Equation)

$\Rightarrow s_i \propto \sum_k w_{ik}$ (Stationary distribution)

Stationary probability of state i

this is proportionality, but if we want to know what it's equal to, we would just divide by sum of $\sum_k w_{ik}$ over all i.

Above is completely general reversible Markov chain. So, this is actually a complete generalization, in the sense that if we have any reversible Markov chain, we can actually interpret in this way.

Conversely: Any reversible chain is of above form!

Let $w_{ij} = s_i q_{ij}$ (Assuming we already have the s and q) and we want to define the weights such that the Markov chain just is this (above) particular random walk.

$= s_j q_{ji}$ (since, it's reversible)

$P(X_{n+1} = j \mid X_n = i) = w_{ij} / \sum_k w_{ik} = \dfrac{s_i q_{ij}}{\sum_k s_i q_{ik}}$

X process is random walk on above network

transition prob. we have

$$= \frac{\cancel{S_i} q_{ij}}{S_i \sum_K q_{iK}} = \frac{q_{ij}}{\sum_K q_{iK}}$$

Assuming S_i's are positive
(It's not a very interesting stationary chain if some stationary probs. are 0, then we should have just restricted the space and not included them anyway)

$\sum_K q_{iK} = 1$ (since it's a Markov chain, so we have to go from i to somewhere)

$$= q_{ij}$$

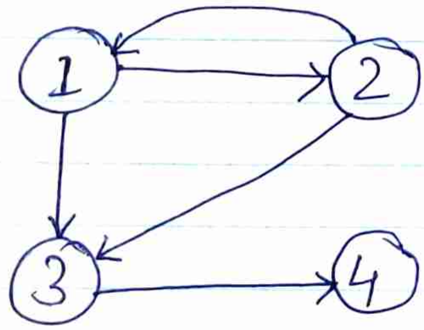
$$\text{or, } P(X_{n+1} = j \mid X_n = i) = q_{ij}.$$

Non-reversible example

Google PageRank → based exactly on a Markov chain and it happens to be a non-reversible chain

WWW

(Assume that the WWW has only 4 pages)



Importance of page based on number of pages linking to it and their importance.

S_j (score of jth webpage) $= \sum_i S_i q_{ij}$,

$\vec{S} = \vec{S} Q$
(vector of all the page ranks)

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

Sum of each row sum to 1 in Markov chain

(We assume that from page 4, we can go anywhere on the web with equal prob. since we want to have MC here)

Which says $\vec{s}$ is stationary distribution of random web-surfing chain.

This has a natural interpretation in terms of stationary distributions as well, i.e., intuitively we think of the stationary distribution as giving the long-run probabilities of being in different states.

$\vec{s}$ is normalized.

In terms of the interpretation of stationary distribution is the long run fraction of time being in a certain state, what $s_j = \sum_i s_i q_{ij}$, i.e., $\vec{s} = \vec{s}Q$ is saying is that, if we imagine just randomly surfing the web for ages and ages..., and in the long run this says that the importance of a page is the long run fraction of time that we spend at that page.

"The pages that are more important, we'll find ourself spending more time there in the long run."

$$G = \alpha Q + (1-\alpha)\frac{J}{M}$$

$\frac{J}{M}$ → valid Markov transition matrix

Where $M = \#$ pages

Q is $(M \times M)$ matrix

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

$0 < \alpha < 1$

G: Google chain

αQ: follow the link

$(1-\alpha)$: we do what is called teleportation, don't follow the link structure anymore, go to uniformly random page

Guarantees irreducibility, no zeroes in transition matrix G

$\alpha = 0.85$ (original paper)

85% of time following random links, 15% of time teleporting to random pages.

Use convergence to stionarity!

Let $\underbrace{\vec{t}}_{\text{PMF at time 0 written as row vector}}$ be initial probability vector that adds up to 1.

$\vec{t}G$ = dist. after 1 step.

$\vec{t}G^2$ = dist. after 2 steps, ...

$$\vec{t}G = \alpha \underbrace{\vec{t}Q}_{\substack{\text{Very sparse} \\ \text{(Mostly 0's)}}} + (1-\alpha)\underbrace{\vec{t}J/M}$$

$\vec{t}J = (1\,1\,1 \ldots 1) \longrightarrow$ row vector

$$(\vec{t}G)G = tG^2, \vec{t}G^3, \ldots, \vec{t}G^n, \ldots$$

If $n \to \infty$, it will converge to the stationary distribution. The stationary distribution is the page rank vector.