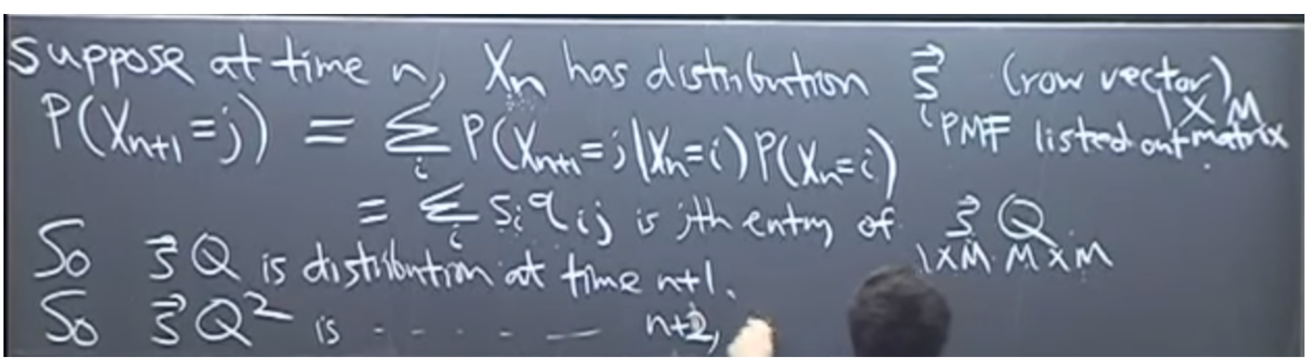
markov chains

- a stochastic process is one in which the random variables evolve over time. we can model as discrete or continuous time and space. continuous space just means that the random variables are discrete or continuous. an example model of a stochastic process is something like $P(X_n = a | X_{n-1} = i_{n-1}, X_{n-1} = i_{n-1} ... X_0 = i_0)$.
- A markov process is one in which $P(X_n = a | X_{n-1} = i_{n-1}, X_{n-1} = i_{n-1} ... X_0 = i_0) = P(X_n = a | X_{n-1})$. This means that the distribution of the immediate future value is only dependent on the present value.
- That is, the past and the future are conditionally independent, given the present. Note this does *not* mean that the future is independent of the past.
- the above is also called the markov assumption. whether that is a good assumption or not for a given problem completely depends on the problem, obviously. as always we will study what happens if this assumption is true.
- We will tack on another assumption that $P(X_{n+1} = j | X_n = i) = q_{ij}$. This means that the probability is independent of n, the time step. It only depends on the current and future value. This assumption is called the homogenous property, which means independent of time. We will only be studying these cases. You can see how this homogenous markov chain can be written as a state machine with transition probabilities, along with a transition probability matrix.
- This is called the first order markov model where the distribution of the future depends on only 1 step back. if you think about something that depends on n steps back, that is a n^{th} order markov assumption.
- markov chain monte carlo is where you encode the information so that the markov chain's resultant distribution converges to the actual underlying distribution of the process we are interested in. once we've encoded it this way (if we can), we can use all the results and tools that come from markov chains. "this has caused a revolution in scientific computing" - joe blitzstein.
- note that as we see in gilbert strang's linear algebra course, matrices and associated operations (especially eigenvalues) are useful for markov processes.

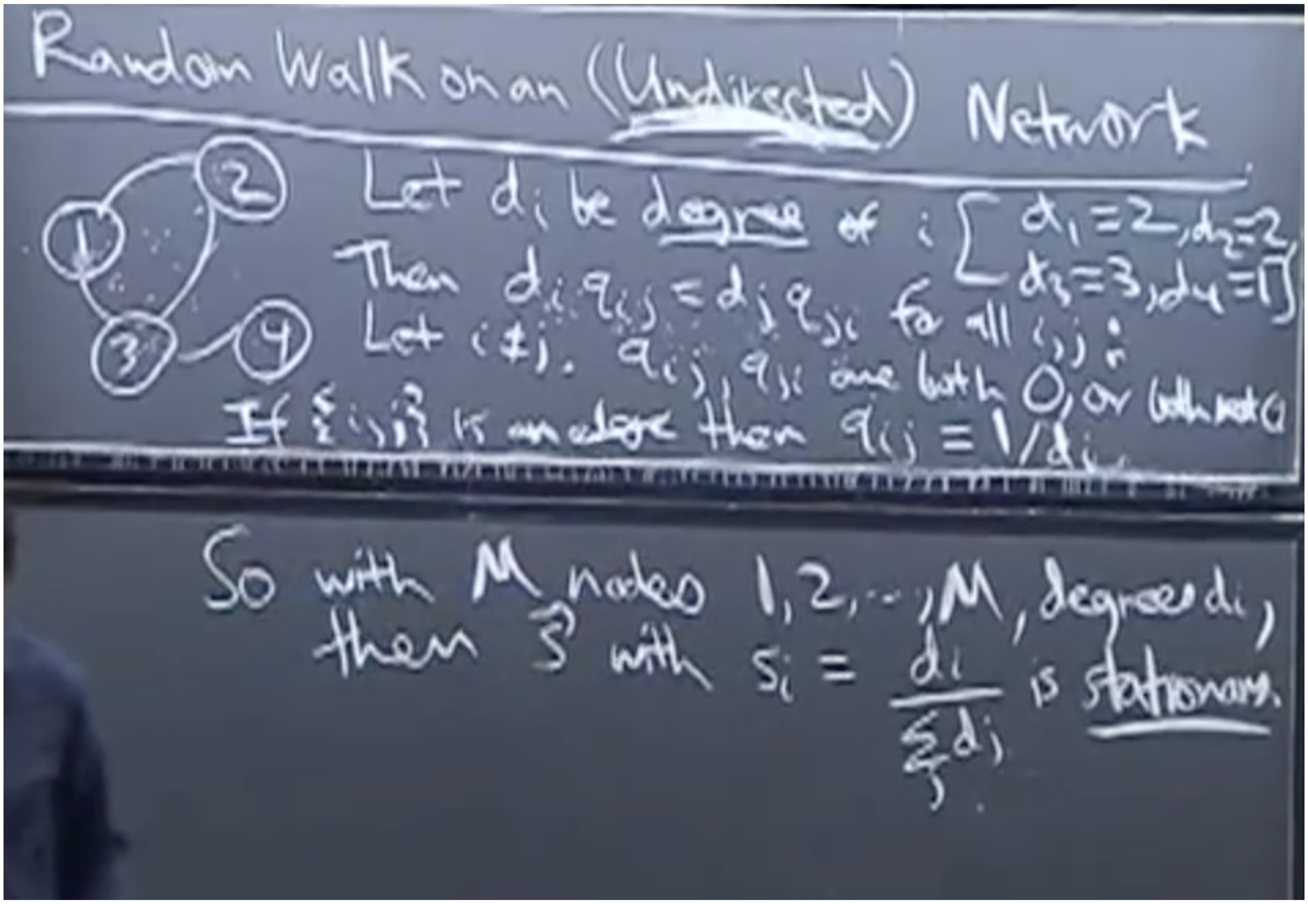
eg.



- A stationary distribution is one where $\vec{s}Q = \vec{s}$, where $\vec{s}$ is a given state's probabilities (state's pmf), and Q is the transition matrix. This should remind us of the linear algebra matrix eigenvalue/eigenvectors notion of stationary. Note that $\vec{s}Q = \vec{s}$ is also similar to how a eigenvalue/eigenvector equation looks.

- The intuition for this name "stationary" is that over time the state distribution doesn't change at all since every time step simulated forward gives us the same state probabilities. so if the chain starts out with the stationary distribution, it is going to have the stationary distribution forever.
- Some interesting questions we can ask of a distribution:
  - does a stationary distribution exist?
  - is it unique?
  - does the chain converge to $\vec{s}$?
  - if it exists, how can we compute it?

  To answer the above questions, solving matrix equations comes into question.
- some vocab to invent a way to talk about markov chains:
  - a chain is irreducible, if its possible (with non-0 probability) to go from anywhere to anywhere. otherwise its called a reducible chain.
  - a state is recurrent if, starting there, the chain has probability 1 of returning to that state. otherwise, its called a transient state. note that here we mean probability 1 of returning to that state *eventually,* which means it can only be 0 or 1 probability (think why).
  - a state is called absorbing if once it gets to that state, it never leaves.
  - we call a state periodic if every n (some constant) steps, we keep returning to that state.
- Let $\vec{s}$ be the belief vector (that is, our belief probabilities of which state we are currently in). Let Q be the transition matrix where q_{ij} is the probability of going from state $i \rightarrow j$. $\vec{s}Q$ gives the probability belief state after 1 transition (you can verify this). $\vec{s}Q^n$ gives the belief state after n transitions.
- if $\vec{s}Q = \vec{s}$, then $\vec{s}$ is called a stationary state. If we just take the transponse on both sides, we get $Q^T s^T = s^T$. Now we can see how this is the same as eigenvalue/eigenvector equation and all of the theory we've learnt there will transfer over here. Note that for stationary state to exist, we will need a eigenvalue of 1, and a non-zero elements' eigenvector whose elements add up to 1. that will be a stationary state. joe states a bunch of theorems about these stationary states without proving them.
- A state is reversible if for all i, j, $s_i q_{ij} = s_j q_{ij}$.
- note that we have already seen that $\sum_i s_i^{current} q_{ij} = s_j^{next}$
- for a reversible state, $\sum_i s_i^{current} q_{ij} = \sum_i s_j^{current} q_{ji} = s_j^{current} \sum_i q_{ji} = s_j^{current} = s_j^{next}$. this shows that if its a reversible state, it is a stationary state.
- a (uniformly) random walk (going randomly from state to state at each time step, picking the state to go to uniformly randomly). random walk on an undirected network is a markov chain, obviously.
- for this setup, we can find the stationary state pretty easily, using the property that reversible states are stationary.



- note that s_i is proportional to the degree (number of edges) of the state.
- Google pagerank was built on an idea like this → the importance of each web page is the sum of the importances of the page that links to it times the edge weight, with edge weight calculated as $\frac{1}{no.\ outgoing\ links}$.
- They do a trick to allow random teleportation with a small probability. this means that the web graph is a irreducible markov chain. joe stated (but didn't prove) a theorem that said these irreducible ones always have a stationary state and that any probability belief converges to that. they let the chain run for a long time instead of finding exact solution. and now they have the stationary probability belief, which they use for page rank.