Markov Chains and Random Walks from « Randomized Algorithms» by Rajeev Motwani and Prabhakar Raghavan

$G = (V, E)$ n vertices m edges

$v \in V$. $\Gamma(v)$ denotes the set of neighbours of v in G

A random walk on G is the following process, which occurs in a sequence of discrete steps. starting at a vertex v_0, we proceed at the first step to a randomly chosen neighbour of v_0. At the second step, we proceed to a randomly chosen neighbour of v_1, and so on.

{6.1 A 2-SAT EXAMPLE}

algorithm: we start with an arbitrary assignment of values to the literals. As long as there is a clause that is unsatisfied, we modify the current assignment as follows: we choose an arbitrary unsatisfied clause, and pick one of the two literals in it uniformly at random; the new assignment is obtained by complementing the value of the chosen literal.

Theorem 6.1 The expected number of steps for the above 2-SAT algorithm to find a satisfying assignment is $O(n^2)$. proof?

{6.2 Markov Chains}

A Markov chain $\mathcal{M}$ is a discrete-time stochastic process defined over a set of states S in terms of a matrix P of transition probabilities.

The set S is either finite or countably infinite.
The transition probability matrix P has one row and one column for each state in S.
The Markov chain is in one state at any time, making state-transitions at discrete time-steps $t = 1, 2, \ldots$ The entry P_{ij} in the transition probability matrix is the probability that the next state will be j, given that the current state is i.

Thus, for all $i, j \in S$, we have $0 \le P_{ij} \le 1$, and $\sum_j P_{ij} = 1$.

memorylessness property: the future behaviour of a Markov chain depends only on its current state, and not on how it arrived at the present state.
We will denote by X_t the state of the Markov chain at time t; thus, the sequence $\{X_t\}$ specifies the history or the evolution of the Markov chain

$$\Pr[X_{t+1} = j \mid X_0 = i_0, X_1 = i_1, \ldots, X_t = i] = \Pr[X_{t+1} = j \mid X_t = i] = P_{ij}.$$

For states $i, j \in S$, define the t-step transition probability as $P_{ij}^{(t)} = Pr[X_t = j \mid X_0 = i]$.

Given an initial state $X_0 = i$, the probability that the first transition into state j occurs at time t is denoted by $r_{ij}^{(t)}$ and is given by

$$r_{ij}^{(t)} = Pr[X_t = j \text{, and, for } 1 \le s \le t-1, X_s \ne j \mid X_0 = i].$$

Also, for $X_0 = i$, the probability that there is a visit (transition into) state j at some time $t > 0$ is denoted by f_{ij}, and is given by

$$f_{ij} = \sum_{t>0} r_{ij}^{(t)}$$

Finally, the expected number of time steps to reach state j starting from state i is denoted by h_{ij} and is given by

$$h_{ij} = \sum_{t>0} t\, r_{ij}^{(t)}$$

Definition 6.1 A state i for which $f_{ii} < 1$ (and hence $h_{ii} = \infty$) is said to be transient, and one for which $f_{ii} = 1$ is said to be persisent. Those persistent states i for which $h_{ii} = \infty$ are said to be null persistent and those for which $h_{ii} \ne \infty$ are said to be non-null persistent.

Definition 6.2 A strong component of a directed graph G is a maximal subgraph C of G such that for any pair of vertices i and j in the vertex set of C, there is a directed path from i to j, as well as a directed path from j to i.

Definition 6.3 A strong component C is said to be a final strong component if there is no edge going from a vertex in C to a vertex not in C.

A state is persistent if and only if it lies in a final strong component

Definition 6.4: A Markov chain is said to be irreducible whenever its underlying graph consists of a single strong component.

Definition 6.5: Define $q^{(t)} = (q_1^{(t)}, q_2^{(t)}, \ldots, q_n^{(t)})$, the state probability vector (also called the distribution of the chain at time t), to be the row vector whose ith component is the probability that the chain is in state i at time t.

It is easy to check that $q^{(t+1)} = q^{(t)} P$, so $q^{(t)} = q^{(0)} P^t$.

It follows that a Markov chain's behaviour for all time is specified by its initial distribution $q^{(0)}$ and its transition matrix P.

Definition 6.6: A *stationary distribution* for the Markov chain with transition matrix P is a probability distribution π such that $\pi = \pi P$.

Definition 6.7: ? The *periodicity* of a state i is the maximum integer T for which there exists an initial distribution $q^{(0)}$ and positive integer a such that, for all t, if at time t we have $q_i^{(t)} > 0$, then t belongs to the arithmetic progression $\{a + Ti \mid i \geq 0\}$. A state is said to be *periodic* if it has periodicity greater than 1, and is said to be *aperiodic* otherwise. A Markov chain in which every state is aperiodic is known as an aperiodic Markov chain.

Definition 6.8: An *ergodic* state is one that is aperiodic and non-null persistent.

Definition 6.9: An *ergodic* Markov chain is one which all states are ergodic.

Theorem 6.2 (Fundamental Theorem of Markov Chains):
Any irreducible, finite, and aperiodic Markov chain has the following properties

1. All states are ergodic
2. There is a unique stationary distribution π such that, for $1 \leq i \leq n$, $\pi_i > 0$.
3. For $1 \leq i \leq n$, $f_{ii} = 1$ and $h_{ii} = \frac{1}{\pi_i}$
4. Let $N(i,t)$ be the number of times the Markov chain visits state i in t steps.

   Then $$\lim_{t \to \infty} \frac{N(i,t)}{t} = \pi_i.$$

6.3. Random Walks on Graphs

$G = (V, E)$ be a connected, non-bipartite, undirected graph $|V| = n$, $|E| = m$.
It induces a Markov chain M_G as follows: the states of M_G are the vertices of G and for any two vertices $u, v \in V$

$$P_{uv} = \begin{cases} \frac{1}{d(u)} & \text{if } (u,v) \in E \\ 0 & \text{otherwise} \end{cases}$$

$d(w)$ is the degree of vertex w. G is connected $\Rightarrow$ M_G is irreducible.
For a connected, undirected graph G, the periodicity of the states in M_G is the greatest common divisor of the length of all closed walks in G.

a closed walk is any walk that starts and ends at the same vertex

G is undirected $\Rightarrow$ there are closed walks of length 2.
G is non-bipartite $\Rightarrow$ it has odd cycles that give closed walks of odd length. } $\Rightarrow$ gcd of closed walks is 1.

$\Rightarrow M_G$ is aperiodic. $\xRightarrow{M_G \text{ is finite}}$ M_G has a unique stationary distribution π.

Lemma 6.3 For all $v \in V$, $\pi_v = \frac{d(v)}{2m}$

Proof: Let $[\pi P]_v$ denote the component corresponding to vertex v in the probability vector πP. Then

$$
\begin{aligned}
[\pi P]_v &= \sum_u \pi_u P_{uv} \\
&= \sum_{(u,v)\in E} \frac{d(u)}{2m} \times \frac{1}{d(u)} \\
&= \sum_{(u,v)\in E} \frac{1}{2m} \\
&= \frac{d(v)}{2m}
\end{aligned}
$$

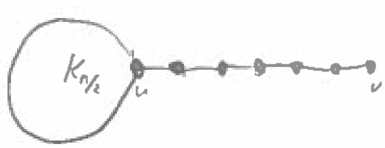
Lemma 6.4 For all $v \in V$, $h_{vv} = 1/\pi_v = 2m/d(v)$.

Definition 6.10 The hitting time h_{uv} (sometimes called the mean first passage time) is the expected number of steps in a random walk that starts at u and ends upon first reaching v.

Definition 6.11 We define C_{uv}, the commute time between u and v, to be $C_{uv} = h_{uv} + h_{vu} = C_{vu}$
This is expected time for a random walk starting at u to return to u after at least one visit to v.

Definition 6.12 Let $C_u(G)$ denote the expected length of a walk that starts at u and ends upon visiting every vertex in G at least once. The cover time of G, denoted $C(G)$ is defined by $C(G) = \max_u C_u(G)$.

Example 6.1 n-vertex lollipop graph L_n. This graph consists of a clique on $n/2$ vertices, and a path on the remaining vertices. There is a vertex u in the clique to which the path is attached; let v denote the other end of the path.



L_n has cover time $\Theta(n^3)$

chain $\Theta(n^2)$ cover time

K_n $\Theta(n \log n)$ cover time.

h_{uv} is $\Theta(n^3)$, whereas h_{vu} is $\Theta(n^2)$

Lemma 6.5 For any edge $(u,v) \in E$, $h_{uv} + h_{vu} \leq 2m$.

Proof. consider a new Markov chain on the edges of G.

The current state = (the edge most recently traversed, the direction of this traversal)
replace each undirected edge by two oppositely directed edges, the directed edges form the state space, —— $2m$ states in this new Markov chain.
The transition matrix Q for this Markov chain has non-zero entry

$$Q_{(u,v),(v,w)} = P_{vw} = 1/d(v).$$

corresponding to an edge (v,w). This matrix is doubly stochastic, meaning that not only do the rows sum to one, but the columns sum to one as well.
Uniform distribution on the edges is stationary for this Markov chain. so the stationary probability of each directed edge is $1/2m$.
By part(3) of Theorem 6.2, we can conclude that the expected time between successive traversals of the directed edge (u,v) is $2m$

Consider $h_{uv} + h_{vu}$, the expected time for a walk starting from u to visit v and return to u. Conditioned on the event that the initial entry into u was via the directed edge (v,u), we conclude that the expected time to go from there to v and then to u along (v,u) is $2m$. The memoryless property allows us to remove the conditioning, the expected time back to u is at most $2m$ □

{ 6.5 Cover Times }

Theorem 6.8: $C(G) \leq 2m(n-1)$

Proof: let T be a spanning tree of G. There is a traversal of T, visiting vertices $v_0, v_1, \ldots, v_{2n-2} = v_0$ that traverses each edge exactly once in each direction. Further, every vertex of G appears at least once in the sequence $v_0, v_1, \ldots, v_{2n-2}$. Consider a random walk that starts at v_0 and terminates upon returning to v_0, having visited the vertices prescribed by the traversal. Since this walk has visited every vertex in G, an upper bound on the expected length of this walk is an upper bound on $C_{v_0}(G)$.

$$C_{v_0}(G) \leq \sum_{j=0}^{2n-3} h_{v_j, v_{j+1}} = \sum_{(v,w) \in T} C_{vw}$$

Since v_j, v_{j+1} are adjacent for all j, we have by Lemma 6.5 that

$$C_{v_j, v_{j+1}} \leq 2m$$

Since there are $n-1$ vertices in T. $C_{v_0}(G) \leq 2m(n-1)$
This upper bound holds no matter which vertex we designate to be v_0. $C(G) \leq 2m(n-1)$ □

{6.4. Electrical Networks}

undirected graph G, $N(G)$ be the electrical network:

for every edge in E, it has one ohm resistance between corresponding nodes in $N(G)$

R_{uv} denotes the effective resistance between the corresponding nodes in $N(G)$

Theorem 6.6 For any two vertices u and v in G, the commute time $C_{uv} = 2m R_{uv}$.

Corollary 6.7 In any n-vertex graph, and for all vertices u and v, $C_{uv} < n^3$.

Theorem 6.9 $m R(G) \le C(G) \le 2e^3 m R(G) \ln n + n$

$R(G) = \max_{u,v \in V} R_{uv}$ — the resistance of G

Lemma 6.10 Suppose that g contains p edge-disjoint paths of length at most l from s to t. Then $R_{st} \le l/p$.

{6.6 Graph Connectivity}

A connected component of G is a (maximal) subset of vertices in which every pair of vertices is connected

The undirected s-t connectivity (USTCON):

Given an undirected graph G and two vertices s and t in G, decide whether s and t are in the same connected component.

A probablistic log-space Turing machine for a language L is a probablistic Turing machine using space $O(\log n)$ on instances of size n, and running in time polynomial in n.

RLP: A language A is in RLP if there exists a probablistic log-space Turing machine M such that on any input x,

$$\Pr[M \text{ accepts } x] \begin{cases} \ge 1/2 & x \in A \\ 0 & x \notin A \end{cases}$$

$O(\log n)$ refers to the workspace of the Turing machine

Theorem 6.11 $\text{USTCON} \subset \text{RLP}$

Proof. Simulates a simple random walk of length $2n^3$, starting from s.
By theorem 6.6, $h_{st} \le n^3$. By Markov inequality, if s is in the same component of G as s, the probability that it is not visited in a random walk of $2n^3$ steps starting from s is at most $1/2$. The Turing machine uses its workspace to count up to $2n^3$, and keep track of its position in the graph during the walk; both require space $O(\log n)$ $\square$.

Randomized Algorithms and Probablistic Analysis by Anna Karlin

{Coupling}

Definition] Given a Markov chain on Ω, a coupling is a Markov chain on $\Omega \times \Omega$ defining a stochastic process (X_t, Y_t) so that

1. X_t and Y_t alone are faithful copies of the original Markov chain; that is, there are initial conditions X_0' and Y_0' on the original Markov chain defining stochastic processes X_t' and Y_t' so that

$$Pr((X_t, Y_t) \in A \times \Omega) = P_r(X_t' \in A)$$

$$Pr((X_t, Y_t) \in \Omega \times A) = P_r(Y_t' \in A)$$

for all $A \subset \Omega$.

2. If $X_t = Y_t$ then $X_{t+1} = Y_{t+1}$ pointwise.

Lemma 8 (Coupling Lemma).

Let $Z_t = (X_t, Y_t)$ be a coupling where $Y_0 = \pi$ and $X_0 = X$. where X is some arbitrary distribution. Suppose there exist a T so that

$$Pr(X_T \neq Y_T \mid X_0 = X) \leq \epsilon$$

Then the mixing time starting at X is bounded by T. that is $\tau_X(\epsilon)$

Proof Suppose X starts at some arbitrary X_0. For all $A \subseteq \Omega$.

$$
\begin{aligned}
P_r(X_T \in A) &= Pr(X_T = Y_T \cap Y_T \in A) + P_r(X_T \neq Y_T \cap X_T \in A) \\
&\geq P_r(X_T = Y_T \cap Y_T \in A) \\
&= 1 - P_r(X_T \neq Y_T \cup Y_T \notin A) \\
&\geq 1 - P_r(Y_T \notin A) - P_r(X_T \neq Y_T) \geq \pi(A) - \epsilon
\end{aligned}
$$

Similarly, $P_r(X_T \notin A) = P_r(X_T \in A^c) \geq \pi(A^c) - \epsilon$

so $P_r(X_T \in A) \leq \pi(A) + \epsilon$

The total variation distance between two distributions D_1 and D_2 on the same sample space Ω is defined as.

$$\|D_1 - D_2\| = \frac{1}{2} \sum_{x \in \Omega} |D_1(x) - D_2(x)| = \max_{A \subset \Omega} |D_1(A) - D_2(A)|$$

mixing time $\tau(\epsilon) = \min\{t \mid \|P^t - \pi\|_{TV} \leq \epsilon\}$.