Theorem 1. *Assuming that the 3-SAT formula has a satisfying assignment the expected run-time to find a satisfying assignment is* $O(n^{3/2}(4/3)^n)$.

Proof:

Assume that the formula has a satisfying assignment S.

We determine the expected number of times we should repeat Step 1 before we find a satisfying assignment.

Let q be the probability that the algorithm reaches a satisfying assignment in $3n$ steps starting with a truth assignment chosen uniformly at random.

Let q_j be a lower bound on the probability that the algorithm reaches S (or some other satisfying assignment) when it starts with a truth assignment that includes exactly j variables that do not agree with S. Then,

$$q_j \geq \binom{3j}{j}(\frac{2}{3})^j(\frac{1}{3})^{2j}$$

.

When $j > 0$,

$$\binom{3j}{j} = \frac{(3j!)}{j!(2j!)}$$

Use Stirling's Formula:

$$\sqrt{2\pi n}(n/e)^n \leq n! \leq 2\sqrt{2\pi n}(n/e)^n$$

$$\begin{aligned} &\geq \frac{\sqrt{2\pi(3j)}}{\sqrt{4\pi j}\sqrt{2\pi(2j)}}(\frac{3j}{e})^{3j}(\frac{e}{2j})^{2j}(\frac{e}{j})^{j} \\ &= \frac{\sqrt{3}}{8\sqrt{\pi j}}(\frac{27}{4})^{j} \\ &= \frac{c}{\sqrt{j}}(\frac{27}{4})^{j} \end{aligned}$$

where $c = \sqrt{3}/8\sqrt{\pi}$.

Thus, when $j > 0$,

$$
\begin{aligned}
q_j &\geq \binom{3j}{j} (\frac{2}{3})^j (\frac{1}{3})^{2j} \\
&\geq \frac{c}{\sqrt{j}} (\frac{27}{4})^j (\frac{2}{3})^j (\frac{1}{3})^{2j} \\
&\geq \frac{c}{\sqrt{j}} \frac{1}{2^j}.
\end{aligned}
$$

Also, $q_0 = 1$.

$$q \geq$$

$$\sum_{j=0}^{n} \Pr(\text{a random assignment has } j \text{ mismatches with } S).q_j$$

$$\geq \frac{1}{2^n} + \sum_{j=1}^{n} \binom{n}{j} (\frac{1}{2})^n \frac{c}{\sqrt{j}} \frac{1}{2^j}$$

$$\geq \frac{c}{\sqrt{n}} (\frac{1}{2})^n \sum_{j=0}^{n} \binom{n}{j} \frac{1}{2^j} (1)^{n-j}$$

$$= \frac{c}{\sqrt{n}} (\frac{1}{2})^n (\frac{3}{2})^n$$

$$= \frac{c}{\sqrt{n}} (\frac{3}{4})^n.$$

The number of random assignments the algorithm tries is geometrically distributed with parameter q.

The expected number of tries is $1/q$ and the number of steps per try is at most $3n$.

Back to Markov Chains

$P_{i,j}^t$ = probability that starting at i at time 0 we reach j at time t.

$r_{i,j}^t$ = probability that starting at i at time 0 we reach j **first time** at step t.

$f_{i,j} = \sum_{t>0} r_{i,j}^t$ = probability of ever reaching j when starting from i.

$h_{i,j} = \sum_{t>0} t r_{i,j}^t$ = the expected time of reaching j from i.

A state is **transient** iff $f_{i,i} < 1$, else it is **recurrent**.

A Markov chain is **irreducible** iff for all i, j $P^t_{i,j} > 0$ for some $t > 0$.

Lemma 1. *A finite discrete time Markov chain is irreducible iff its underlying graph is strongly connected.*

A state is **periodic** iff there exists an integer d such that $P^t_{i,i} > 0$ only for $t = jd + c$ for some fixed integer c, else the state is **aperiodic**.

A chain is aperiodic iff all its states are aperiodic.

Stationary Distribution

Assume that we start at state 1. The state of the system at time t is given by the distribution

$$(P^t_{1,1}, P^t_{1,2}, ..., P^t_{1,i}, ..., P^t_{1,n}) = P_1(t)$$

In general, if $P_i(t)$ is the distribution of the state of the system at time t, starting at i.

$$P_i(t+1) = P_i(t)\mathcal{P}$$

$$P_i(t+s) = P_i(t)\mathcal{P}^s.$$

A **stationary distribution** of a chain is a distribution

$$\pi = (\pi_1, \pi_2, ..., \pi_i, ..., \pi_n)$$

such that

$$\pi = \pi\mathcal{P}.$$

$$\pi_j = \sum_k \pi_k P_{k,j}$$

Example

In the two containers example:

A stationary probability $\bar{\pi}$ needs to satisfy:

$$\pi\mathcal{P} = \pi$$

or:

$\pi_0 = \frac{1}{2}\pi_0 + \pi_1\frac{1}{2m}$

$\pi_j = \frac{1}{2}\pi_j + \pi_{j-1}\frac{m-j+1}{2m} + \pi_{j+1}\frac{j+1}{2m}$

$\pi_m = \frac{1}{2}\pi_m + \pi_{m-1}\frac{1}{2m}$

Satisfied by

$$\pi_i = \binom{m}{j}(\frac{1}{2})^m$$

The Fundamental Theorem of Markov Chains

Theorem 2. *Any irreducible, finite, aperiodic Markov chain*

1. *has a* **stationary distribution** π*;*

2. *the stationary distribution is* **unique***;*

3. $\pi_i = \lim_{t\to\infty} P^t_{j,i}$

4. $h_{i,i} = \frac{1}{\pi_i}$.

The Fundamental Theorem of (infinite) Markov Chains

A state i is **ergodic** if it is aperiodic, and $h_{i,i} < \infty$.

A Markov chain is **ergodic** if all its states are ergodic.

Theorem 3. *Any irreducible, ergodic Markov chain*

1. *has a* **stationary distribution** π*;*

2. *the stationary distribution is* **unique***;*

3. $\pi_i = \lim_{t\to\infty} P^t_{j,i}$

4. $h_{i,i} = \frac{1}{\pi_i}$.

Random walk in undirected graphs

Given a general undirected graph.

Given that the walk is at vertex v, and v has $d(v)$ neighbors, the walk proceeds to a neighbor of v chosen uniformly at random (i.e. with probability $\frac{1}{d(v)}$).

Let $G = (V, E)$ be an undirected graph. $|V| = n$, $|E| = m$, $d(v)$ is the degree of vertex v.

The state of the Markov chain is a vertex.

The transition probabilities are $P_{u,v} = \frac{1}{d(u)}$ if v a neighbor of u, else 0.

Theorem 4. *If G is not bi-partite the chain is aperiodic.*

Proof. Since the graph is undirected there are walks of length 2 from i to i.

If the graph is not bi-partite it has odd cycles, thus has walks of odd length. $\square$

Theorem 5. *Assume that the graph is undirected, connected, and not bi-partite, then it has a unique stationary distribution* π *where*

$$\pi_u = \frac{d(u)}{2m}.$$

Proof. We need to show that π is a probability distribution and that $\pi = \pi\mathcal{P}$.

Since $\sum_{u \in V} d(u) = 2m$. $\sum_{v \in V} \pi_v = 1$, thus we have a probability distribution.

For all $v \in V$

$$\frac{d(v)}{2m} = \sum_{u \in N(v)} \frac{d(u)}{2m} \frac{1}{d(u)}.$$

□

Expected Return Time

Theorem 6. *In a random walk* $h_{u,u} = \frac{2m}{d(u)}$.

Theorem 7. *Assume that* $(u, v) \in E$ *then*

$$h_{u,v} \leq 2m.$$

Proof.

$$\frac{2m}{d(u)} = h_{u,u} = \frac{1}{d(u)} \sum_{v \in N(u)} [1 + h_{v,u}].$$

$$2m = \sum_{v \in N(u)} [1 + h_{v,u}].$$

Thus,

$$h_{v,u} \leq 2m.$$

□

Definition 1. *The* **cover time** *of a graph G is the maximum expected time to visit all nodes of the graph starting from any vertex.*

Theorem 8. *The cover time of an undirected graph with n vertices and m edges is bounded by $4nm$.*

Proof.

Choose a spanning tree (it has n vertices), fix an Eulerian path on that tree (it pass every edge exactly twice). Let $v_1, v_2, v_3, ...$ be the sequence of vertices in that path.

The cover time is bounded by

$$h_{v_1,v_2} + h_{v_2,v_3} + ... \leq 4nm.$$

□

$s - t$ Connectivity Algorithm

Given an undirected graph $G = (V, E)$ and two vertices $s, t \in V$, is there a path connecting s to t in G?

Algorithm:

1. Start a random walk from s.

2. If reaches t within $4n^3$ steps return **CONNECTED** else return **NOT CONNECTED**.

Theorem 9. *There is a one side error algorithm for $s - t$ connectivity. The algorithm returns the correct answer with probability $\geq 1/2$ and terminates in $O(n^3)$ steps.*

Proof. If there is no path the algorithm is always correct.

If there is a path the probability that the algorithm does not find the path is bounded by

$$\frac{4nm}{4n^3} \leq \frac{1}{2}.$$

□

Deterministic $s - t$ connectivity takes $O(m)$ steps, but requires $O(n)$ **space**.

The randomized solution takes $O(mn)$ steps, requires $O(\log n)$ space.

The randomized algorithm needs to store the last position $O(\log n)$ bits, and count to $O(n^3)$, again $O(\log n)$ bits.