Markov Chains 3

1 USTCON: Reachability in undirected graphs

We are given an undirected graph G on n vertices, via access to the adjacency list representation, and the names of two vertices s and t. The problem is to decide whether there is a path between s and t. This problem is called USTCON (undirected s-t connectivity). Standard algorithms like BFS and DFS use $\Omega(n)$ space to solve this problem. Our goal is to solve USTCON using much less space.

We will see how to solve this problem in (i) $O(\log^2 n)$ space deterministically, (ii) $O(\log n)$ space via a randomized algorithm. In 2004, Omer Reingold gave a deterministic $O(\log n)$ algorithm for this problem. Note that the label of a single vertex requires $\Theta(\log n)$ space to store, thus this is best possible.

Proposition 1 *We can solve USTCON using $O(\log^2 n)$ space.*

Proof We run the following recursive procedure with $D = n - 1$.

$Check(u, v, D)$: Is $d(u, v) \leq D$?

- If $D = 1$, check if u, v are adjacent.
- For each $w \in V \setminus \{u, v\}$, do:
- Return $Check(u, w, \lfloor D/2 \rfloor)$ AND $Check(w, v, \lceil D/2 \rceil)$.

The depth of recursion is $O(\log n)$ and each vertex label needs $O(\log n)$ space.
■

Proposition 2 *We can solve USTCON in randomized logspace.*

Proof Do a **random walk from** s for $200n|E|$ steps. If we reach t in this walk, accept, otherwise reject.

Claim 1 *If s, t belong to the same connected component, then the expected number of steps from s to reach t is at most $2n|E|$.*

Therefore, with probability at least 99/100, the algorithm decides correctly. The space used is the space needed for one vertex label and to count the number of steps, which is $O(\log n)$. ■

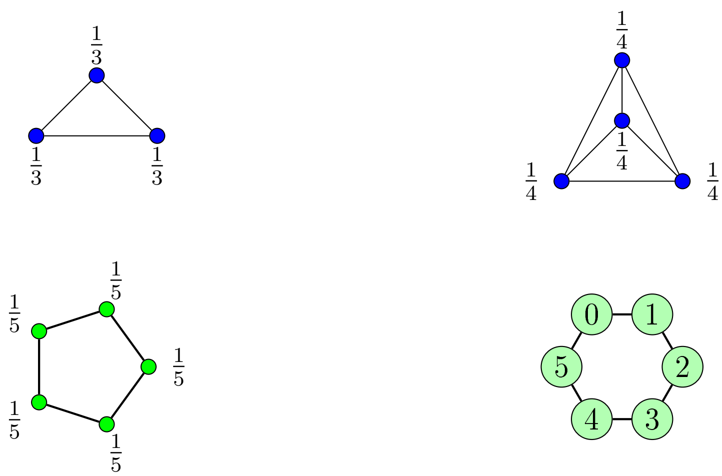
Thus, what remains to be proven is the claim on the expected time to reach one vertex from another. We shall do this in the remaining sections.

2 Stationary distributions

We define the stationary distribution on a Markov chain as $\pi^* = lim_{t\to\infty}\pi_t = \pi^* = lim_{t\to\infty}\pi_0 M^t$. Which Markov chains have stationary distributions and how can we find it? The answer to the second question can be found from the fact that the stationary distribution must satisfy $\pi^* = \pi^* \cdot M$. Thus, π^* must be a left-eigenvector of M (or a right eigenvector of M^T), and further the sum of the entries of π^* must equal one. For the case of random walks on undirected graphs, there is a simpler answer.

2.1 Periodicity

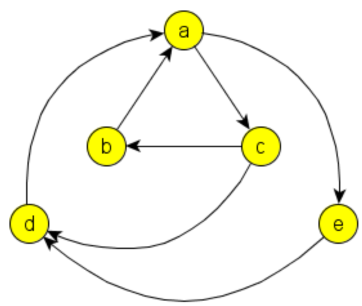
Consider a random walk on an undirected graph. In the following examples, we may guess what happens to the probability distribution as t, the number of time-steps, goes to inifinity.



In the first three cases, the distribution converges towards a uniform distribution on the vertices, irrespective of the initial distribution. However, in the last example, there is no stationary distribution.

The reason for this is that if the random walk beings at vertex zero, then at even time-steps, it can only be in positions zero, two or four; and in odd time-steps in positions one, three or five.

When a state of a Markov chain can only be visited at time-steps that avoid some congruence class (modulo some integer), then we say that such a state is **periodic**. A Markov chain is periodic if it has at least one periodic state, and aperiodic otherwise. Periodicity can thus be an obstruction to the existence of stationary probability.

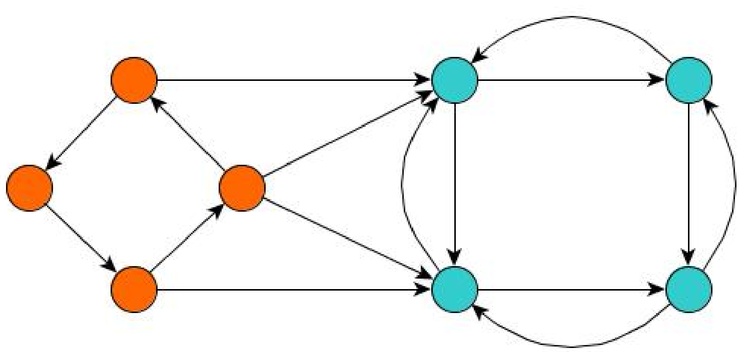


For undirected graphs, bipartiteness is the deciding criteriion for periodicity.

Claim 2 *A random walk starting at a fixed vertex on an undirected graph G is periodic if and only if G is bipartite.*

2.2 Reducibility

The other issue we deal with is reducibility. Consider the following example.



Here, the set of blue vertices forms a strongly connected component, and is a leaf in the underlying DAG. Once the Markov chain enters such a leaf component, it must stay there; thus it suffices to study stationary distributions on irreducible Markov chains, that is when the underlying directed graph is strongly connected. Otherwise, the Markov chain is said to be reducible.

We can now state a sufficient condition for the existence of staioary distributions: the following result, which is a special case of the fundamental theorem of Markov chains.

Theorem 3 *If a Markov chain is finite, aperiodic and irreducible, then:*

1. *It has a unique stationary distribution* $\pi^* = (\pi^*(1), \ldots, \pi^*(n))$*, which is independent of the initial distribution* π_0.
2. *The expected number of time-steps to return to a vertex* v*, starting at* v *is* $h_{v,v} = \dfrac{1}{\pi^*(v)}$.

We conclude this section with an exact expression for the stationary distribution for random walks on undirected graphs.

Proposition 4 *Let* $G = (V, E)$ *be finite, undirected, connected, and non-bipartite. Then for a random walk on* G*, we have:*

$$\pi^*(v) = \frac{deg(v)}{2|E|}.$$

Proof For every vertex v, we have:

$$\pi^*(v) = \sum_{u \in N(v)} \pi^*(u) \cdot \frac{1}{deg(u)}.$$

We can check by substitution that the above equation is satisfied by the given expression, along with $\sum_{v \in V} \pi^*(v) = 1$. From the fundamental theorem of Markov chains, this must therefore be the unique stationary distribution.

For an alternative way, we write:

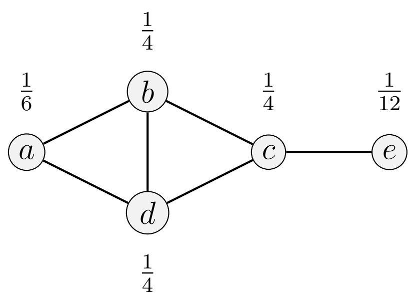
$$\frac{\pi^*(v)}{deg(v)} = \frac{1}{deg(v)} \sum_{u \in N(v)} \frac{\pi^*(u)}{deg(u)}.$$

Now, substituting $f(v) = \dfrac{\pi^*(v)}{deg(v)}$, we get:

$$f(v) = \frac{1}{deg(v)} \sum_{u \in N(v)} f(u).$$

We now observe that f must be constant: consider a vertex v for which $f(v)$ is minimum. Then every neighbor u of v must satisfy $f(u) = f(v)$ (if some value is larger, then the average-of-neighbors property is violated). Repeating this argument and using the connectedness of G proves that f is constant. Finally, the constant of proportionality must be equal to the sum of the degrees of all the vertices, which proves the result. ■

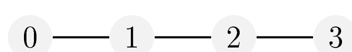
An example calculation is shown below, with the stationary probabilities marked for each vertex.



3 Hitting times

We define $h_{u,v} = E[min\{r \geq 1\} | X_{t+r} = v | X_t = u]$. In words, if $u \neq v$, then $h_{u,v}$ is equal to $T_{u,v}$, the expected time to reach v from u. The quantity $h_{u,u}$ is the expected time to return to u, and is thus non-zero, whereas $T_{u,u} = 0$.

An example:



Here, $h_{1,1} = \dfrac{1}{2}(h_{0,1} + h_{2,1}) = \dfrac{1+3}{2} = 2$.

The following claim follows from Theorem 3 and Proposition 4.

Claim 3 *For an undirected graph G, and a vertex $v \in V(G)$, we have:* $h_{v,v} = \frac{2|E|}{deg(v)}$.

Proposition 5 *Let G be a connected, undirected graph.*

- *If u, v are adjacent vertices of G, then $h_{u,v} < 2|E|$.*
- *If u, v are arbitrary vertices of G, then $h_{u,v} < 2n|E|$.*

Proof We have: $h_{v,v} = \frac{2|E|}{d(v)}$. Also, $h_{v,v} = \frac{1}{d(v)} \sum_{u \in N(v)} (1 + h_{u,v})$. Thus, $2|E| \geq 1 + h_{u,v}$. This proves the first part.

Now, let u, v be arbitrary vertices and $uv_1v_2 \ldots v_kv$ be a $u - v$ path in G. Then we have:

$$h_{u,v} \leq h_{u,v_1} + h_{v_1,v_2} + \ldots + h_{v_k,v}.$$

Thus $h_{u,v} \leq (n-1)max\{h_{u,v_1}, \ldots, h_{v_k,v}\} \leq (n-1)2|E| < 2n|E|$, which proves the second part. ■

By using the same idea in the above proof, we can bound the expected time to visit all vertices, starting at v, and returning to v.

Cover time: The cover time of a vertex v, denoted by $c(v)$ is the expected time to visit all vertices, starting from v. The cover time of G is defined as $c(G) = max_v\{c(v)\}$.

Then we have: $c(G) \leq 4(n-1)|E| < 2n^3$.

This can be proved by considering a spanning tree and a traversal from v and applying the idea from the previous proof.