Randomness and Computation 2023
Tutorial 6 (week 8)
Issue date: Wednesday, 1th November, 2023.

1. **Sampling from Independent Sets of a graph**. A independent set (IS) of a graph $G = (V, E)$ is a set of vertices belonging to V for which no edge $e \in E$ connects two vertices of the set. We want to construct a Markov chain that allows to sample uniformly at random from the ensemble of independent sets of G, where G is assumed to be composed of a single connected component. Let's defined the following Markov chain X_t over the space Ω_{IS} of independent set (Section 12.6 of Mitzenmacher and Upfal). Let's X_0 be a trivial IS, for example the empty set. Then at each step we:

   - Select an edge $e = (u, v) \in E$ from the graph uniformly at random.
   - We proceed as follows:
     - (M1): with probability $1/3$ set $X_{t+1} = X_t - \{u, v\}$
     - (M2): with probability $1/3$ let $Y = (X_t - \{u\}) \cup \{v\}$. If Y is an IS, then $X_{t+1} = Y$; otherwise $X_{t+1} = X_t$ (M2).
     - (M3): with probability $1/3$ let $Y = (X_t - \{v\}) \cup \{u\}$. If Y is an IS, then $X_{t+1} = Y$; otherwise $X_{t+1} = X_t$ (M3).

   (a) Explain the effect of the three action that have $1/3$ probability in the MC.

   (b) Show that the MC is irreversible and aperiodic.

   (c) Use detail balance to show that it converges to the uniform distribution over Ω_{IS}.

2. **An s-t connectivity algorithm using $O(\log n)$ space.** We are given an undirected graph $G = (V, E)$ and two vertices s and t in G. Let $|V| = n$ and $|E| = m$. One can easily find a connection between s and t in linear time using standard breath-first search or depth-first search, however, requiring $\Omega(n)$ space. Consider the following algorithm:

   - Start a random walk from s
   - If the walk reaches t within $4n^3$ steps, return that there is a path.
   - Otherwise return no path exists.

   (a) Discuss why the algorithm needs only $O(\log n)$ work memory space, assuming the graph G is given as input on read-only memory.

   (b) Proof that the expected time to reach x from y noted as $h_{x,y}$, satisfies the inequality $h_{x,y} < 2m$, exploiting the relation between the expected time to return and the stationary distribution $h_{i,i} = 1/\pi(i)$ (Theorem 7.7 of Mitzenmacher and Upfal).

   (c) The cover time of G is the maximum over all $v \in V$ of the expected time to visit all nodes starting from v. Show that the cover time is bounded by $4nm$.

   (d) Prove that the algorithm only errs by returning there is no path when there is indeed one with probability smaller than 1/2.

(e) Discuss how to reduce that probability to a very small constant δ.

Raul Garcia-Patron Sanchez