6.895 Randomness and Computation March 20, 2006

Lecture 12

Lecturer: Ronitt Rubinfeld *Scribe: Petar Maymounkov*

Generating Random Spamming Trees with Markov Chains (Part I)

This lecture continues the analysis of the algorithm for generating uniform samples from the set of spanning trees of a graph. The algorithm is defined as follows:

- The state of the Markov chain is a rooted spanning tree (of the input graph G) with directed edges, such that all edges point toward the root.
- Begin from an arbitrary rooted tree. If the graph did not have self-loops, add them.
- Repeat several times:
  - Let the root of the current state's tree be v. Pick a random edge (v, w) uniformly at random.
  - Add (v, w) to the tree. This creates a cycle.
  - Delete the single edge outgoing from w, and set w as the new root.

Fact 1 *Two basic properties of this Markov chain are:*

1. *The chain is ergodic, because:*
   - (a) *It is irreducible (left as an exercise)*
   - (b) *It is aperiodic (With the self-loops added)*
2. *The out-degree of each Markov state is equal to the degree of the corresponding tree node in the input graph, plus 1 to account for the self-loops.*
3. *The in-degree of each Markov state is equal to its out-degree.*

Since the in- and out-degrees of the Markov chain are the same, we have that the stationary distribution $\tilde{\pi}$ is such that:

$$\tilde{\pi}_s = \frac{1 + \text{Degree of root at state } s}{z}$$

here z is a normalizing factor:

$$z = (2m + 2n) \cdot (\text{\# of unrooted spanning trees})$$

where m stands for the number of edges in the input graph, and n stands for the number of vertices.

Next we need to verify that when the stationary distribution is reached, each unrooted tree is equally likely to be reached:

$$\begin{aligned}
\mathbf{Pr}[\text{output unrooted tree } T] &= \sum_{v\in V} \mathbf{Pr}[\text{reach state } (T,v)] \\
&= \sum_{v\in V} \frac{\deg(v)+1}{z} \\
&= \frac{2m+2n}{z} \\
&= \frac{1}{\#\text{ unrooted trees}}
\end{aligned}$$

Finally, we need to verify that the Markov chain is "rapid mixing", i.e. that after a small number of Markov transitions (algorithm steps) starting from an arbitrary state the resulting distribution approximates the stationary. To show this, we need to take a digression into "coupling techniques for proving rapid convergence".

Coupling for Markov Chains

For notation, we are going to use $(x, y) \sim w$ and $(x, y) \in_R w$ interchangably to mean that the random variables (x, y) are distributed according to the joint distribution w.

Definition 2 *For distributions u on Ω and v on Ω, a distribution w on the (joint) event space Ω^2 is a **coupling** if:*

$$\begin{aligned}
\forall x \quad & \sum_y w(x,y) = u(x) \\
\forall y \quad & \sum_x w(x,y) = v(y)
\end{aligned}$$

i.e. the marginal distributions on x and y equal u and v, respectively.

Additionally, to measure the "closeness" of two distributions, we use the following:

Definition 3 *The **total variation distance** between u and v is:*

$$\begin{aligned}
d_{TV}(u,v) &= \frac{1}{2}||u-v||_1 \\
&= \frac{1}{2}\sum_{x\in\Omega} |u(x)-v(x)| \\
&= \max_{s\subseteq\Omega} \big\{u(s)-v(s)\big\}
\end{aligned}$$

Lemma 4 (Coupling Lemma) *For any u and v both on a finite event space Ω, a coupling w, and $(x, y) \sim w$, we have that:*

$$d_{TV}(u, v) \leq \mathbf{Pr}[x \neq y]$$

Proof For every $z \in \Omega$, we have $w(z, z) \leq \min\{u(z), v(z)\}$, hence for $(x, y) \sim w$:

$$\begin{aligned}\mathbf{Pr}[x = y] &= \sum_z w(z, z) \\ &\leq \sum_z \min\{u(z), v(z)\}\end{aligned}$$

And therefore:

$$\begin{aligned}\mathbf{Pr}[x \neq y] &\geq 1 - \sum_z \min\{u(z), v(z)\} \\ &= \sum_z \Big(u(z) - \min\{u(z), v(z)\}\Big) \\ &= \sum_{\substack{z \\ v(z)<u(z)}} \Big(u(z) - v(z)\Big) \\ &= \max_S\{u(s) - v(s)\} \\ &= d_{TV}(u, v)\end{aligned}$$

■

Now consider two random walks of t steps X_t and Y_t, starting at states (or distributions over states) a and b respectively, and using the same transition matrix. Also define the joint distribution of X_t and Y_t to be such that when they first meet they stay together.

More specifically now, consider a coupling (X_t, Y_t), where b is chosen according to the Markov chain's stationary distribution $\tilde{\pi}$. The coupling lemma implies that:

$$d_{TV}(X_t, \tilde{\pi}) \leq \mathbf{Pr}[X_t \neq Y_t | X_0 = a \wedge Y_0 \sim \tilde{\pi}]$$

Here we have abused notation by using X_t to also refer to the distribution of the random variable X_t. This inequality suggests that our goal will be to find an appropriate coupling, for which the two distributions X_t and $Y_t = \tilde{\pi}$ become close for small values of t. More formally:

Let T be the expected first coupling time (when the two random processes arrive at the same state), then:

Theorem 5

$$d_{TV}(X_t, \tilde{\pi}) \leq \frac{T}{t}$$

Proof Let τ be the random variable denoting the time when the two random walks meet for the first time. Then:

$$\begin{aligned} d_{TV}(X_t, \tilde{\pi}) &\leq \mathbf{Pr}[X_t \neq Y_t] && \text{by Coupling Lemma} \\ &= \mathbf{Pr}[\tau > t] \\ &\leq T/t && \text{by Markov Inequality} \end{aligned}$$

■

Generating Random Spamming Trees with Markov Chains (Part II)

Now, we define the coupling process for our problem at hand:

1. Run the processes X^t and Y^t independently until just the tree roots collide
2. Force the roots to move identically hereafter

We have:

$$\mathbf{E}[\text{coupling time}] = \mathbf{E}[\text{time for roots to collide}] + \mathbf{E}[\text{time until trees identical}|\text{same roots}]$$

The first term (on the right) is left as an exercise. The second term equals the cover time on the input graph, because once the coupling has walked through all vertices of the input graph, the two trees are sure to be identical. This is because once the two trees share a root, a step from root u to root v ensures that edge (u, v) will be in both trees.

Conductance Preliminaries

Let $G = (V, E)$ be given, and let $(S, \bar{S})$ be a cut, i.e. $S \subseteq V$, $\bar{S} = V \backslash S$. Then we'll make a first attempt at defining the "conductance" of the cut as:

$$\phi(S, \bar{S}) = \frac{|E_{S,\bar{S}}|}{|E_S|}$$

Where:

$$\begin{aligned} E_{S,\bar{S}} &= \{(u, v)|(u, v) \in E \wedge ((u, v) \in S \times \bar{S} \vee (v, u) \in S \times \bar{S})\} \\ E_S &= \{(u, v)|(u, v) \in E \wedge (u \in S \vee v \in S)\} \end{aligned}$$

Then the conductance of the graph G might be defined as:

$$\phi_G = \min_{S:|S|=|V|/2} \phi(S, \bar{S})$$

In the next lecture, we'll look at a more appropriate definition.