Lecturer: Siu On Chan — *Scribe: Siu On Chan*

Notes 20: Sampling spanning trees by random walk

1. Random spanning trees

Fix a connected graph $G = (V, E)$ on n vertices. A spanning tree in G is an acyclic subgraph of G containing $n - 1$ edges.

We want to sample a spanning tree of G, (nearly) uniformly at random, as follows.

Random walk on spanning trees

> Let T_0 be an arbitrary spanning tree of G
> For $t = 0, 1, 2, \dots$
>   Remove an edge from T_t uniformly at random to obtain F_t
>   Among all spanning trees containing F_t, uniformly pick one as the new T_{t+1}

This is a random walk/Markov chain on an auxiliary weighted graph $\mathcal{T}_G$, whose nodes are spanning trees in G, and two nodes in $\mathcal{T}_G$ are adjacent if they share exactly $n - 2$ edges.

For decades, this random walk was conjectured to mix in polynomial time. It was recently proved by Anari, Liu, Oveis Gharan, and Vinzart.

Theorem 1.1. *The above random walk has eigenvalue gap at least* $1/(n - 1)$.

Eigenvalue gap β is the difference $\lambda_1 - \lambda_2$ between the two largest eigenvalues. By results in Notes12, the lazy version of the random walk mixes in polynomial time. Recall that the lazy random walk mixes in time $O((\log |V(\mathcal{T}_G)|)/\beta)$. Since $|V(\mathcal{T}_G)| \leqslant \binom{\binom{n}{2}}{n-1} \leqslant \binom{n^2}{n-1} \leqslant n^{2(n-1)} = \exp(O(n \log n))$, the lazy random walk mixes in time $O(n^2 \log n)$.

2. Bipartite inclusion graphs

Since the ambient graph G is fixed, we identify a spanning tree T in G with the set of $n-1$ edges in T.

Denote by $Y(n - 1)$ the sets of edges of spanning trees in G. More generally, define

$$Y(k) = \left\{F \in \binom{E}{k} \mid F \text{ is acyclic}\right\} \qquad \text{for } 0 \leqslant k \leqslant n - 1 .$$

Decompose the random walk over spanning trees into two transitions:

- (Down) Go from $T_t \in Y(n - 1)$ to $F_t \in Y(n - 2)$ by removing an edge uniformly at random
- (Up) Then from F_t to $T_{t+1} \in Y(n-1)$ by choosing T_{t+1} uniformly from among all $T_{t+1} \supset F_t$

Down transition corresponds to matrix D_{n-1}, which is $Y(n - 1)$-by-$Y(n - 2)$ (recall that a row probability vector multiplies on the left to a transition matrix). Likewise up transition corresponds to matrix U_{n-1}, which is $Y(n - 2)$-by-$Y(n - 1)$.

$D_{n-1}U_{n-1}$ is the transition matrix for the random walk. We want to bound $\beta(D_{n-1}U_{n-1})$.

We think of up and down transitions as random walk transitions on the auxiliary graph Γ_{n-1}:

Definition 2.1 (Bipartite inclusion graph)**.** For $0 \leqslant k \leqslant n - 1$, Γ_k has vertex set $Y(k) \cup Y(k - 1)$. $F \in Y(k)$ is adjacent to $F' \in Y(k - 1)$ if $F \supset F'$.

We will look at Γ_k for $0 \leqslant k \leqslant n - 1$ later on to apply induction.

In Γ_k, down transition means moving from $F \in Y(k)$ to a random neighbor $F' \in Y(k - 1)$, uniformly from among all k neighbors of F.

We will define up transition matrix U_k shortly, to move from $F' \in Y(k-1)$ to a random neighbor $F \in Y(k)$ from some distribution.

Thanks to the following proposition, $D_{n-1}U_{n-1}$ and $U_{n-1}D_{n-1}$ have the same eigenvalue gap, whenever this gap is at most 1.

Proposition 2.2. *Given any matrices A and B, AB and BA have the same non-zero eigenvalues with the same multiplicities.*

This proposition can be proved by showing AB and BA have essentially the same characteristic polynomials. Look up "Characteristic polynomials" on Wikipedia if interested.

Kaufman and Oppenheim came up with a way to relate $\beta(U_{n-1}D_{n-1})$ to $\beta(D_{n-2}U_{n-2})$. Of course, $\beta(D_{n-2}U_{n-2}) = \beta(U_{n-2}D_{n-2})$. One can then inductively bound $\beta(U_kD_k) = \beta(D_kU_k)$.

Proposition 2.3 (Kaufman–Oppenheim)**.** *$\beta(D_kU_k) \geqslant 1/k$ for $1 \leqslant k \leqslant n-1$.*

This main proposition implies the main theorem.

3. Random walk on bipartite graph

Up and down transitions U_k and D_k in Γ_k are special cases of random walk on bipartite graph.

3.1. **General bipartite graph.** Consider bipartite graph Γ on vertex set $T \cup B$ ("top" and "bottom") with weight μ over its edges.

Further assume μ is a distribution: nonnegative weight on edges summing to 1.

Choosing an edge $(t,b) \in T \times B$ from μ induces marginal distribution π_T on T, and marginal π_B on B.

The marginal probability $\pi_T(t)$ coincides with the degree of $t \in T$ (sum of edge weights incident to t). Similarly $\pi_B(b)$ is the degree of $b \in B$.

The simple random walk on Γ with edge weights μ has transition matrix $\begin{pmatrix} O & D \\ U & O \end{pmatrix}$.

D denotes ("down") transition from T to B; U denotes ("up") transition from B to T.

Distribution $\pi = (\frac{1}{2}\pi_T, \frac{1}{2}\pi_B)$ over $T \cup B$ is stationary for the random walk, because its probability masses are proportional to vertex degrees.

$$\begin{pmatrix} \frac{1}{2}\pi_T^\top & \frac{1}{2}\pi_B^\top \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\pi_T^\top & \frac{1}{2}\pi_B^\top \end{pmatrix} \begin{pmatrix} O & D \\ U & O \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\pi_B^\top U & \frac{1}{2}\pi_T^\top D \end{pmatrix} \ , \text{ so}$$

$$\pi_T^\top D = \pi_B^\top \qquad \text{and} \qquad \pi_B^\top U = \pi_T^\top \ .$$

3.2. **Bipartite inclusion graph Γ_k.** Then $T = Y(k)$ and $B = Y(k-1)$.

Edge weight $\mu = \mu_k$ will be chosen later so that $D = D_k$. That is, random walk from T to B according to μ_k corresponds to uniformly dropping an element from $t \in T$.

We want transition probabilities $\mu_k(t,b)/\pi_T(t)$ from t to any of its neighbor b to be $1/k$, so

$$\mu_k(t,b) = \frac{\pi_T(t)}{k} \qquad \text{for every edge } (t,b) \text{ in } \Gamma_k.$$

U_k will be up transition matrix U in Γ_k, i.e. random walk from B to T according to μ_k.

U_k coincides with this transition: Given $b \in B$, choose neighbor t of b with probability proportional to $\pi_T(t)$. This is because

$$U_k(b,t) = \frac{\mu_k(t,b)}{\pi_B(b)} = \frac{1}{\pi_B(b)}\frac{\pi_T(t)}{k} \propto \pi_T(t) \ .$$

4. n-bipartite inclusion graph

We can visualize using the n-bipartite inclusion graph $\Gamma_{n-1} \cup \cdots \cup \Gamma_1$.

We now define edge weights μ_k's for every layer so that

(A) Every μ_k is a distribution over edges in Γ_k
(B) Down transition in Γ_k according to μ_k is D_k
(C) The top marginal of μ_{n-1} is the uniform over $Y(n-1)$
(D) Top marginal distribution of μ_{k-1} in Γ_{k-1} equals bottom marginal distribution of μ_k in Γ_k (so as to relate $\beta(U_kD_k)$ in Γ_k to $\beta(D_{k-1}U_{k-1})$ in Γ_{k-1})

This is done by the following random process of choosing a path $(F_{n-1}, \ldots, F_0)$ on $\Gamma_{n-1} \cup \cdots \cup \Gamma_1$:

(1) Choose $F_{n-1} \in Y(n-1)$ uniformly at random
(2) For k from $n-1$ to 1, drop a uniformly random edge in F_k to get $F_{k-1} \in Y(k-1)$

F_k in this random path has marginal distribution π_k over $Y(k)$. π_k is both the top marginal of μ_k in Γ_k and bottom marginal of μ_{k+1} in Γ_{k+1}.

The joint marginal of $(F_k, F_{k-1}) \in Y(k) \times Y(k-1)$ induces edge distribution μ_k for Γ_k.

$\pi_k^\top D_k = \pi_{k-1}^\top$ and $\pi_{k-1}^\top U_k = \pi_k^\top$, so both $D_k U_k$ and $U_{k+1} D_{k+1}$ have stationary distribution π_k.

π_{n-1} is uniform over $Y(n-1)$. U_{n-1} defined here agrees with §2 since $U_{n-1}(b,t) \propto \pi_{n-1}(t) \propto 1$.

For some applications, one may want to sample from a nonuniform target distribution π over spanning trees in $Y(n-1)$. For example, G may have edge weights w, and a natural distribution π on spanning trees T would be proportional to the product of edge weights in T, so that $\pi(T) \propto \prod_{e \in T} w(e)$.

We need to change the edge weights μ_k, so that they now satisfy

(C') The top marginal of μ_{n-1} is π

This can be done by changing the first step of the random process of the path $(F_{n-1}, \ldots, F_0)$:

(1') Choose $F_{n-1} \in Y(n-1)$ from π

Step (2) remains the same. Changing initial distribution π in step (1) will affect edge distributions μ_k in Γ_k and marginals π_k on $Y(k)$.

Everything in our analysis still holds given any target distribution π over $Y(n-1)$.

5. Up-walk and down-walk

In Γ_k, layer k of the n-partite inclusion graph, there are two natural two-step walks:

- Up-walk $P_{k-1}^{\wedge} = U_k D_k$ on a weighted graph on vertex set $Y(k-1)$
- Down-walk $P_k^{\vee} = D_k U_k$ on a weighted graph on vertex set $Y(k)$

Focusing on Γ_k, abbreviate

$$T = Y(k) \qquad B = Y(k-1) \qquad \mu = \mu_k$$

5.1. Down-walk.

$P_k^{\vee} = D_k U_k$ induces a random walk on a weighted graph on T.

A step in $P_k^{\vee}$ also corresponds to a length-2 walk in Γ_k, from T to B to T.

$$P_k^{\vee}(t,t') = \begin{cases} \displaystyle\sum_{b \in B,\, b \subset t} \frac{\mu(t,b)}{\pi_k(t)} \frac{\mu(t',b)}{\pi_{k-1}(b)} & t = t' \\ \dfrac{\mu(t,b)}{\pi_k(t)} \dfrac{\mu(t',b)}{\pi_{k-1}(b)} & t \cap t' = b \in B \\ 0 & \text{otherwise} \end{cases}.$$

5.2. Up-walk.

$P_{k-1}^{\wedge} = U_{k-1} D_{k-1}$ corresponds to the random walk on a weighted graph on B.

A step in $P_{k-1}^{\wedge}$ also corresponds to a length-2 walk in Γ_k, from B to T to B.

$$P_{k-1}^{\wedge}(b,b') = \begin{cases} 1/k & b = b' \\ \dfrac{\mu(t,b)}{\pi_{k-1}(b)} \dfrac{\mu(t,b')}{\pi_k(t)} & b \cup b' = t \in T \\ 0 & \text{otherwise} \end{cases}.$$

5.3. Non-lazy up-walk.

$\tilde{P}_{k-1}^{\wedge}$ is the non-lazy version of $P_{k-1}^{\wedge}$.

A step in $\tilde{P}_{k-1}^{\wedge}$ also corresponds to a length-2 *path* in Γ_k, from B to T to a different vertex in B.

$$P_{k-1}^{\wedge} = \frac{1}{k} I + \frac{k-1}{k} \tilde{P}_{k-1}^{\wedge}\,.$$

π_{k-1} is stationary for both $P_{k-1}^{\wedge}$ and $\tilde{P}_{k-1}^{\wedge}$, as a common left-eigenvector with eigenvalue 1.

6. Spectral comparison

We now upperbound the spectrum of $\tilde{P}_k^{\wedge}$ by the spectrum of $P_k^{\vee}$. It is more convenient to first transform each of them into a symmetric matrix with the same spectrum.

$\tilde{P}_k^{\wedge}$ and $P_k^{\vee}$ share a common stationary distribution π_k. These transitions are both of the form $P(b,b') = A_P(b,b')/\pi_k(b)$ for symmetric A_P, so $P = \Pi^{-1}A_P$, where $\Pi = \mathrm{Diag}(\pi_k)$. They both represent a random walk with adjacency matrix $A_P = \Pi P$ and common degree matrix Π.

Definition 6.1. Given matrices P, Q, Π, if $\Pi P, \Pi Q$ are symmetric, we write

$$P \preccurlyeq_{\Pi} Q \qquad \Longleftrightarrow \qquad \Pi P \preccurlyeq \Pi Q \ .$$

$P \preccurlyeq_{\Pi} Q$ is equivalent to $\Pi^{1/2}P\Pi^{-1/2} \preccurlyeq \Pi^{1/2}Q\Pi^{-1/2}$, if Π is symmetric and $\Pi \succ 0$.

Since P is similar to the symmetric matrix $\mathcal{A}_P = \Pi^{1/2}P\Pi^{-1/2}$ (the normalized adjacency matrix), they have the same spectra.

In our application, P and Q often represent transitions with a common stationary distribution (the main diagonal of Π). $\mathcal{A}_P$ and $\mathcal{A}_Q$ will have the same top eigenspace (spanned by $\mathbb{1}$). Applying Courant–Fishcer to the orthogonal subspace,

$$P \preccurlyeq_{\Pi} Q \qquad \Longleftrightarrow \qquad \mathcal{A}_P \preccurlyeq \mathcal{A}_Q \qquad \Longrightarrow$$
$$\lambda_2(\mathcal{A}_P) = \sup_{x\perp\mathbb{1},\ \|x\|=1} x^\top \mathcal{A}_P x \leqslant \sup_{x\perp\mathbb{1},\ \|x\|=1} x^\top \mathcal{A}_Q x = \lambda_2(\mathcal{A}_Q)\ ,$$

and hence $\lambda_2(P) = \lambda_2(\mathcal{A}_P) \leqslant \lambda_2(\mathcal{A}_Q) = \lambda_2(Q)$.

Proposition 6.2. *For $1 \leqslant k \leqslant n-2$, let $\Pi = \mathrm{Diag}(\pi_k)$. Then $\tilde{P}_k^{\wedge} \preccurlyeq_{\Pi} P_k^{\vee}$.*

This proposition is proved in the next section. We now show how it implies Proposition 2.3.

Proof of Proposition 2.3. $\lambda_1(P_k^{\vee}) = 1$. We prove by induction that $\lambda_2(P_k^{\vee}) \leqslant 1 - \dfrac{1}{k} = \dfrac{k-1}{k}$.

When $k=1$, $P_k^{\vee}$ has rank 1, so $\lambda_2(P_k^{\vee}) = 0 \leqslant \dfrac{k-1}{k}$.

For $k > 1$, $\tilde{P}_{k-1}^{\wedge} \preccurlyeq_{\Pi} P_{k-1}^{\vee}$ by Proposition 6.2, so $\lambda_2(\tilde{P}_{k-1}^{\wedge}) \leqslant \lambda_2(P_{k-1}^{\vee}) \leqslant \dfrac{k-2}{k-1}$.

Also $\lambda_2(P_{k-1}^{\wedge}) = \dfrac{1}{d} + \dfrac{k-1}{k}\lambda_2(\tilde{P}_{k-1}^{\wedge}) \leqslant \dfrac{k-1}{k}$.

$P_{k-1}^{\wedge} = U_k D_k$ and $P_k^{\vee} = D_k U_k$ share the same non-zero eigenvalues, thus $\lambda_2(P_k^{\vee}) \leqslant \dfrac{k-1}{k}$. □

7. Garland's method

We now discuss Proposition 6.2 that bounds $\tilde{P}_k^{\wedge}$ by $P_k^{\vee}$.

Focusing on adjacent layers $\Gamma_{k+1} \cup \Gamma_k$, abbreviate

$$T = Y(k+1) \qquad M = Y(k) \qquad B = Y(k-1) \qquad (\text{“top” “middle” “bottom”})$$

Garland method decomposes transitions of $P_k^{\vee}$ and $\tilde{P}_k^{\wedge}$ into unions of subgraphs.

7.1. **Down-walk.** A step in $P_k^{\vee} = D_k U_k$ represents a length-2 walk (m, b, m') from M to B to M. Decompose $P_k^{\vee} = \sum_{b\in B} P_b^{\vee}$ based on the bottom vertex b of (m, b, m').

$P_b^{\vee}$ also corresponds to transitions on a weighted clique (with self-loops) over $S_b \subseteq M$, where

$$S_b = \{m \in M \mid m \supset b\}\ .$$

7.2. **Non-lazy up-walk.** A step in $\tilde{P}_k^{\wedge}$ represents a length-2 path (m, t, m') from M to T to a different vertex in M. Decompose $\tilde{P}_k^{\wedge} = \sum_{b\in B} \tilde{P}_b^{\wedge}$ based on the common intersection $b = m \cap m'$ of this path.

$\tilde{P}_b^{\wedge}$ also corresponds to transitions on the weighted graph $H_b = (S_b, E_b)$ over S_b with edge set

$$E_b = \{(m, m') \in S_b \times S_b \mid m \cup m' \in T\}\ .$$

7.3. **Spectra.** The down-walk and non-lazy up-walk are now decomposed into subgraphs on S_b, over various $b \in B$.

For $P_b^\vee$, the subgraph (with adjacency matrix $\Pi P_b^\vee$) is a clique with self-loops.

For $\tilde{P}_b^\wedge$, the subgraph H_b (with adjacency matrix $\Pi \tilde{P}_b^\wedge$) has its edges determined by $Y(k+1)$.

Kaufman and Oppenheim proved that Proposition 6.2 still holds "when restricted to these two subgraphs on S_b".

Lemma 7.1. *For $1 \leqslant k \leqslant n-2$ and $b \in Y(k-1)$, let $\Pi = \mathrm{Diag}(\pi_k)$. Then $\tilde{P}_b^\wedge \preccurlyeq_\Pi P_b^\vee$.*

This lemma is proved in the next lecture. Summing over $b \in Y(k-1)$ yields Proposition 6.2.

For $P_b^\vee$: It has rank 1 since $\Pi P_b^\vee = \mu_b \mu_b^\top / \pi_{k-1}(b)$ (μ_b is the vector of edge weights incident to b).

Therefore $\lambda_2(P_b^\vee) = \cdots = \lambda_n(P_b^\vee) = 0$, and $\Pi P_b^\vee$ is a clique that mixes perfectly in one step.

For $\tilde{P}_b^\wedge$: Turns out $\tilde{P}_b^\wedge$ and $P_b^\vee$ share the same top eigenvector with the same eigenvalue.

Therefore $\tilde{P}_b^\wedge \preccurlyeq_\Pi P_b^\vee$ is equivalent to $\lambda_2(\tilde{P}_b^\wedge) \leqslant 0$.

In particular $\tilde{P}_b^\wedge$ must be a weighted expander.

8. Variations

Suppose you want apply the same random walk algorithm to a different setting, such as

- Uniformly sample a path of length d in a graph G; or
- Uniformly sample a clique of size d in a graph G.

Does the same random walk mix quickly in these settings?

All the constructions still make sense in those settings ($(k+1)$-partite inclusion graph, up- and down-walks, Garland's decomposition). But Lemma 7.1 may or may not hold.

Turns out $P_b^\vee$ from the previous section will still has rank 1 and represents a perfectly mixing weighted clique with $\lambda_2(P_b^\vee) = \cdots = \lambda_n(P_b^\vee) = 0$.

But now $\tilde{P}_b^\wedge$ may not represent weighted expanders and may violate $\lambda_2(\tilde{P}_b^\wedge) \leqslant 0$, because the subgraphs H_b depend crucially on $Y(k+1)$ (and also indirectly on distribution π at the top layer).

They might only satisfy the weaker bound $\lambda_2(\tilde{P}_b^\wedge) \leqslant \alpha$ for some $0 \leqslant \alpha \leqslant 1$, depending on the situation. So $\tilde{P}_b^\wedge$ are only α-expanders. This α upperbounds the error when you approximate the clique of $P_b^\vee$ by the α-expander of $\tilde{P}_b^\wedge$. We will discuss this general situation in the next lecture.

If you can only show $\lambda_2(\tilde{P}_b^\wedge) \leqslant \alpha$ for all $b \in Y(k)$ and $1 \leqslant k \leqslant d-1$, then the conclusion of Proposition 6.2 is weakened to be $\tilde{P}_k^\wedge \preccurlyeq_\Pi P_k^\vee + \alpha I$. And in the proof of Proposition 2.3, you get the weaker bound $\lambda_2(P_k^\vee) \leqslant 1 - \frac{1}{k} + \alpha(k-1)$, since the error accumulates in the induction. This might still give you something useful if α is tiny (at most roughly $1/k^2$).

With a more careful analysis, Alev et al can get a nontrivial upperboundon on $\lambda_2(P_k^\vee)$ even when $\alpha = \Theta(1/k)$.

9. Matroid

One general situation in which Lemma 7.1 holds (and hence fast mixing) is when the set system $Y(0) \cup \cdots \cup Y(d)$ is a matroid.

A matroid is a family $\mathcal{I}$ of subsets over a ground set U that is:

(1) Nonempty: $\mathcal{I} \neq \emptyset$
(2) Downward closed: If $A \in \mathcal{I}$ and $B \subseteq A$, then $B \in \mathcal{I}$
(3) Exchangable: If $A, B \in \mathcal{I}$ and $|A| > |B|$, then there is $e \in A \setminus B$ such that $B \cup \{e\} \in \mathcal{I}$

Turns out all maximal $A \in \mathcal{I}$ have the same size d (called the rank of the matroid).

The family of acyclic edges in a graph G is an example of a matroid. The ground set U is the set E of edges in G. A subset $F \subseteq U$ belongs to $\mathcal{I}$ if F is acyclic. Maximal $F \in \mathcal{I}$ are spanning trees in G. This matroid has rank $n-1$, where n is the number of vertices in G. Fast mixing over spanning trees is thus a special case.