# Advanced Algorithms

## Spectral methods and algorithms

尹一通　栗师　**刘景铖**

# Recap

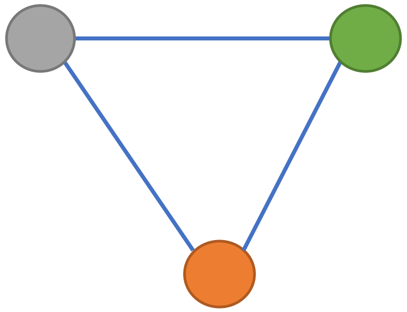
**Previous lecture**:
Random walks on undirected graphs

- Fundamental theorem of Markov chains
- Spectral analysis
- Mixing time
- Random sampling

**What next?**
Random walks on undirected graphs

- From sampling to counting and MCMC
- Expander graphs and random walks

# Coupling for Graph Coloring



- Start with any $k$-coloring $\sigma$
- Pick a vertex $v$ and a color $c$ uniformly at random, recolor $v$ with $c$ if it is legal; otherwise do nothing

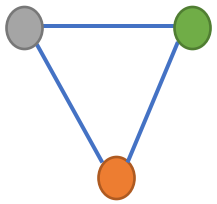
Say we have two arbitrary copies of the Markov chain, $(X_t)$ and $(Y_t)$

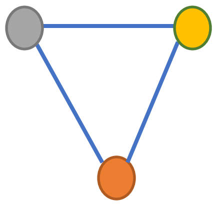
At each step, we let them choose the same vertex $v$ and same color $c$

Let $d_t = \text{number of vertices } X_t \text{ disagree with } Y_t$

Unlike the previous example, $d_t$ can increase now

We need to consider Good Moves that decrease $d_t$, and balance them with Bad Moves that increase $d_t$





Start with any $k$-coloring $\sigma$
Pick a vertex $v$ and a color $c$ u.a.r.,
recolor $v$ with $c$ if legal

# Coupling for Graph Coloring

Say we have two arbitrary copies of the Markov chain, $(X_t)$ and $(Y_t)$

At each step, we let them choose the same vertex $v$ and same color $c$

Let $d_t = \text{number of vertices } X_t \text{ disagree with } Y_t$

Good Moves that decrease $d_t$:

If we chose a disagreeing vertex $v$, and color $c$ does not appear in the neighborhood of $v$ in $X_t$ or $Y_t$, this is a good move

Because we can safely recolor a disagreeing vertex $v$ with color $c$, and they agree from then on

Let $g_t$ be the number of good moves (among all possible $kn$ choices)

There are $d_t$ vertices to choose from, and each disagreeing vertex has a neighborhood of at most $\Delta$ colors in either process, so each disagreeing vertex has $k - 2\Delta$ "safe colors"

$$g_t \geq d_t(k - 2\Delta)$$

Start with any $k$-coloring $\sigma$
Pick a vertex $v$ and a color $c$ u.a.r.,
recolor $v$ with $c$ if legal

# Coupling for Graph Coloring

Say we have two arbitrary copies of the Markov chain, $(X_t)$ and $(Y_t)$

At each step, we let them choose the same vertex $v$ and same color $c$

Let $d_t = \text{number of vertices } X_t \text{ disagree with } Y_t$

Bad Moves that increase $d_t$: a legal move in one process but not the other

This happens when (and only when) the chosen color $c$ is already the color of some neighbor of $v$ in one process but not the other

In other words, $v$ must be a neighbor of some disagreeing vertex $u$, and $c$ must be the color of $u$ in either $X_t$ or $Y_t$

Let $b_t$ be the number of bad moves (among all possible $kn$ choices)

There are $d_t$ choices of disagreeing vertex $u$, then $\Delta$ choices for $v$, then 2 for $X_t$ or $Y_t$

$$b_t \leq 2\Delta d_t$$

Start with any $k$-coloring $\sigma$
Pick a vertex $v$ and a color $c$ u.a.r.,
recolor $v$ with $c$ if legal

# Coupling for Graph Coloring

Say we have two arbitrary copies of the Markov chain, $(X_t)$ and $(Y_t)$
At each step, we let them choose the same vertex $v$ and same color $c$
Let $d_t = \text{number of vertices } X_t \text{ disagree with } Y_t$

Combined: $\mathbb{E}[d_{t+1}|d_t] = d_t + \frac{b_t - g_t}{kn} \leq d_t + d_t \frac{4\Delta - k}{kn} \leq d_t \left(1 - \frac{1}{kn}\right)$

Since $d_0 \leq n$, we have $\mathbb{E}[d_t|d_0] \leq 1/e$ for $t = 2k\, n \ln n$. Thus,

$$d_{TV}(p_t, \pi) \leq \Pr_{(X_t, Y_t) \sim \mu}[X_t \neq Y_t] \leq \Pr[d_t > 0 | X_0, Y_0] = \Pr[d_t \geq 1 | X_0, Y_0] \leq \mathbb{E}[d_t | d_0] \leq 1/e$$

This concludes that the $\epsilon$-mixing time is $O\left(nk \log \frac{n}{\epsilon}\right)$

To improve this to $k \geq 2\Delta + 1$, one tries to pair bad moves in $(X_t)$ but blocked in $(Y_t)$, with bad moves in $(Y_t)$ but blocked in $(X_t)$

# From sampling to counting

Now that we have a Markov chain that outputs a $k$-coloring $\sigma$ almost uniformly at random from all proper colorings after $O\left(nk \log \frac{n}{\epsilon}\right)$ steps

Can we estimate the total number of proper colorings?

This task is known as ***approximate counting***

For many natural concrete problems

$\boldsymbol{ApproxCount} \equiv \boldsymbol{ApproxSample} \equiv \boldsymbol{UniformSample} \subset \boldsymbol{ExactCount}$

The first three are in $\mathrm{BPP}^{\mathrm{NP}}$, while ***ExactCount*** is #P

# From sampling to counting

Denote the number of proper colorings of a graph $G$ by $Z_G$
We start by finding an arbitrary proper coloring $\sigma$ in $G$
Then, we reveal the colors in $\sigma$ one by one
We count how many proper colorings are consistent with the revealed colors

Let $Z_i$ be the number of proper colorings $\tau$ such that
in the first $i$ coordinates, $\tau$ agrees with $\sigma$

Notice that $Z_0 = Z_G, Z_n = 1$, and

$$Z_G = \frac{Z_0}{Z_1} \cdot \frac{Z_1}{Z_2} \cdots \frac{Z_{n-1}}{Z_n}$$

# From sampling to counting

Let $Z_i$ be the number of proper colorings $\tau$ such that

in the first $i$ coordinates, $\tau$ agrees with $\sigma$

Notice that $Z_0 = Z_G$, $Z_n = 1$, and

$$Z_G = \frac{Z_0}{Z_1} \cdot \frac{Z_1}{Z_2} \cdots \frac{Z_{n-1}}{Z_n}$$

Suppose we estimate each ratio within $\left(1 \pm \frac{\epsilon}{2n}\right) \cdot \frac{Z_{i+1}}{Z_i}$ except with prob. $\frac{\delta}{n}$

Then multiplying them all together gives $(1 \pm \epsilon) \cdot Z_G$ except with prob. $\delta$

# From sampling to counting

Let $Z_i$ be the number of proper colorings $\tau$ such that: in the first $i$ coordinates, $\tau$ agrees with $\sigma$

Let $\pi_i$ be the uniform distribution of proper colorings $\tau$ such that: in the first $i$ coordinates, $\tau$ agrees with $\sigma$

Recall that $Z_0 = Z_G$, $Z_n = 1$, and

$$Z_G = \frac{Z_0}{Z_1} \cdot \frac{Z_1}{Z_2} \cdots \frac{Z_{n-1}}{Z_n}$$

How do we estimate each ratio $\frac{Z_{i+1}}{Z_i}$?

We run a Markov chain that samples from $\pi_i$, and use Monte Carlo method to estimate how many are counted in $Z_{i+1}$

**Markov chain**: in the first $i$ coordinates, we fix the colorings as in $\sigma$, and only update the remaining $n - i$ coordinates

**Monte Carlo**: given a sample $\tau$, we check if $\tau_{i+1} = \sigma_{i+1}$

Sampling from $\pi_i$ is an unbiased estimator for the ratio:

$$E_{\tau \sim \pi_i}\,[[\tau_{i+1} = \sigma_{i+1}]] = \frac{Z_{i+1}}{Z_i}$$

Sampling from a rapidly mixing Markov chain $p_t$ only introduces a small bias (recall the def. of TV distance):

$$\left| E_{\tau \sim p_t}\,[[\tau_{i+1} = \sigma_{i+1}]] - E_{\tau \sim \pi_i}\,[[\tau_{i+1} = \sigma_{i+1}]] \right| \le d_{TV}(p_t, \pi_i)$$

$$d_{TV}(p_t, \pi) = \max_{S \subseteq [n]} |p_t(S) - \pi(S)|$$

# From sampling to counting

We run a Markov chain that samples from $\pi_i$, and use Monte Carlo method to estimate how many are counted in $Z_{i+1}$

**Markov chain**: in the first $i$ coordinates, we fix the colorings as in $\sigma$, and only update the remaining

**Monte Carlo**: given a sample $\tau$, we check if $\tau_{i+1} = \sigma_{i+1}$

Sampling from $\pi_i$ is an unbiased estimator for the ratio:

$$E_{\tau\sim\pi_i}\,[[\tau_{i+1} = \sigma_{i+1}]] = \frac{Z_{i+1}}{Z_i}$$

Variance can also be bounded because $\frac{Z_{i+1}}{Z_i}$ is strictly between (0,1): $\frac{\Delta}{k(k-\Delta)} \le \frac{Z_{i+1}}{Z_i} \le \frac{1}{k-\Delta}$

It suffices to take the average over $\text{poly}\left(n,\frac{1}{\epsilon},\frac{1}{\delta}\right)$ samples

Then apply Chebyshev's inequality

Upperbound for the ratio follows from having many colors available
lowerbound from bounding the prob. that any neighbors take the same color

See Chapter 14.4 of LPW book

# Expander Graphs

- Combinatorial: graphs with good expansion
- Probabilistic: graphs in which random walks mix rapidly
- Algebraic: graphs with large spectral gap

Let $G$ be a $d$-regular graph, and let $d = \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq -d$ be the spectrum of its adjacency matrix.

We will be interested in the spectral radius, given by

$$\alpha \coloneqq \max\{\alpha_2, |\alpha_n|\}$$

If $\alpha$ is much smaller than $d$, we have good **spectral expansion**.

There are many nice properties associated with expander graphs

Among others, say if we want more than one sample in MCMC, do we have to resample entirely?

**Cauchy-Schwarz inequality**:
$\langle u, v\rangle \le \sqrt{\langle u, u\rangle} \cdot \sqrt{\langle v, v\rangle}$

# Expander Mixing lemma

Intuitively, an expander can be seen as an approximation to the complete graph, because edges are distributed evenly

**Induced edges**: $E(S,T) \coloneqq \{(u,v): u \in S, v \in T, uv \in E\}$
We also allow non-disjoint $S, T$, in which case an edge can be counted twice.

**Expander Mixing lemma**

Let $G$ be a $d$-regular graph with $n$ vertices. If the spectral radius of G is $\alpha$, then for every $S \subseteq [n], T \subseteq [n]$, we have

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \le \alpha\sqrt{|S||T|}.$$

Proof: Note that $E(S,T) = \chi_S^T A \chi_T$. Let $\chi_S = \sum_i a_i v_i, \chi_T = \sum_i b_i v_i$, where $\{v_i\}$ is an orthonormal basis for $A$, with eigenvalues $\{\alpha_i\}$.

$$E(S,T) = \frac{d|S||T|}{n} + \sum_{i\ge 2} \alpha_i a_i b_i \,.$$

By Cauchy-Schwarz,

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \le \alpha\|a\|_2\|b\|_2 = \alpha\|\chi_S\|_2\|\chi_T\|_2 = \alpha\sqrt{|S||T|}$$

# Expander Mixing lemma

Intuition: Expander mixing lemma tells us that a spectral expander looks like a random graph.

**Exercise**: Let G be a $d$-regular graph with spectral radius $\alpha$. Show that the size of the maximum independent set of G is at most $\frac{\alpha n}{d}$.

Use this result to conclude that the chromatic number is at least $\frac{d}{\alpha}$ .

# Converse to Expander Mixing lemma

(By Bilu and Linial)
Suppose that for every $S \subseteq [n], T \subseteq [n]$ with $S \cap T = \emptyset$, we have

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \leq \alpha\sqrt{|S||T|}.$$

Then all but the largest eigenvalue of $A$ in absolute value is at most

$$O\left(\alpha\left(1 + \log\frac{d}{a}\right)\right)$$

- Proof is based on LP duality
- Would be nice to see an analog of Trevisan's Cheeger's rounding proof

# Existence of expanders

- Complete graphs are obviously the best expanders in terms of “expansion” (in all three notions of “expansion”)
- What’s interesting is the existence of sparse expanders: e.g. $d$-regular expanders for constant $d$

- A random $d$-regular graph is a (combinatorial) expander with high probability
- However, deterministic and explicit construction of expanders seems to be much harder to come up with

# Alon-Boppana Bound

- For $d$-regular graphs, how small can the spectral radius be?
- Ramanujan graphs: graphs whose spectral radius are at most $2\sqrt{d-1}$

**Alon-Boppana Bound**

Let G be a $d$-regular graph with n vertices, and $\alpha_2$ be the second largest eigenvalue of its adjacency matrix. Then

$$\alpha_2 \geq 2\sqrt{d-1} - \frac{2\sqrt{d-1}-1}{\lfloor \mathrm{diam(G)}/2 \rfloor}$$

# Alon-Boppana Bound

<u>An easy lower bound on spectral radius</u>

Let G be a $d$-regular graph with n vertices, and $\alpha$ be its spectral radius. Then $\alpha \geq \sqrt{d} \cdot \sqrt{\frac{n-d}{n-1}}$.

Proof: Consider $\mathrm{Tr}(A^2)$. Counting length-2 walks we have

$$\mathrm{Tr}(A^2) \geq nd$$

On the other hand, $\mathrm{Tr}(A^2) = \sum_i \alpha_i^2 \leq d^2 + (n-1)\alpha^2$.

Combined, we have $\alpha \geq \sqrt{d} \cdot \sqrt{\frac{n-d}{n-1}}$.

For the Alon-Boppana bound, one may consider $\mathrm{Tr}\left(A^{2k}\right)$.

Trace method/trick

# Random walks in expanders

- We knew that it mixes rapidly, in time $O\left(\frac{\log n}{1-\epsilon}\right)$ for $\alpha = \epsilon d$.
- Perhaps surprisingly, not just the final vertex is close to the uniform distribution, but the entire sequence of walks looks like a sequence of independent samples for many applications.

- In fact, expander random walks can fool many test functions:

  ***Expander random walks: a Fourier-analytic approach*, by Cohen, Peri and Ta-Shma**

# Probability amplification

Say you have a randomized algorithm that fails with probability $\beta$

To boost success probability, we can run it multiple times until it succeed

Run independently for t rounds, the failure probability becomes $\beta^t$

Q: Can we save randomness while still achieving the same probability amplification?

Imagine a random walk on the $N = 2^n$ random bits

There is a set $B$ of size $\beta N$ that we try to escape from (or avoid)

We want that the escape probability close to $\beta^t$

Q: Can we use a sparse expander instead of a complete graph for the random walk?

# Hitting property of expander walks

Let G be a $d$-regular graph with n vertices, $\alpha = \epsilon d$ be its spectral radius and $B$ be a set of size at most $\beta n$.

Then, starting from a uniformly random vertex, the probability that a t-step random walk has never escaped from $B$, denoted by $P(B, t)$, is at most $(\beta + \epsilon)^t$.

$\Pr[X_0 \in B, X_1 \in B, X_2 \in B, \dots, X_t \in B]$

Remarks before a proof:

- Compare this to a sequence of independent samples.
- Expander mixing lemma is like $t = 2$: Note that $\varphi(S) = \Pr(X_2 \notin S \mid X_1 \sim \pi_S)$
- Bound can be strengthened → see Chapter 4 of ***Pseudorandomness*, by Vadhan**
- Applications to error reduction for randomized algorithms
  - Instead of using $kt$ bits of randomness, only need $k + O(t \log d)$
  - for one-sided error, escaping the bad set of “random bits”
  - for two-sided error, a Chernoff type bound can also be shown → then take the majority of the answers

# Hitting property of expander walks

$$\Pi_B = \begin{matrix} & B & V\backslash B \\ B & I_B & 0 \\ V\backslash B & 0 & 0 \end{matrix}$$

$$\Pi_B \Pi_B = \Pi_B$$

$W = \frac{1}{d} A$

$u = \frac{1}{n}\vec{1}$

Proof. Observe that $P(B,t) = \|(\Pi_B W)^t \Pi_B u\|_1$

To see this, notice that $\Pr[X_0 \in B] = \|\Pi_B u\|_1$

$$\Pr[X_0 \in B, X_1 \in B] = \|\Pi_B W \Pi_B u\|_1$$

And so on and so forth.

Suppose that we can show $\forall f: f$ is a probability distribution, we have

$$\|\Pi_B W \Pi_B f\|_2 \leq (\beta + \epsilon)\|f\|_2$$

Then,

$$\begin{aligned} \|(\Pi_B W)^t \Pi_B u\|_1 &\leq \sqrt{n}\|(\Pi_B W)^t \Pi_B u\|_2 \\ &= \sqrt{n}\|(\Pi_B W \Pi_B)^t u\|_2 \\ &\leq \sqrt{n}(\beta + \epsilon)^t \|u\|_2 \\ &= (\beta + \epsilon)^t \end{aligned}$$

**Cauchy-Schwarz inequality**:

$$\langle u, v\rangle \leq \sqrt{\langle u, u\rangle} \cdot \sqrt{\langle v, v\rangle}$$

# Hitting property of expander walks

$$\Pi_B = \begin{matrix} & B & V\backslash B \\ B & I_B & 0 \\ V\backslash B & 0 & 0 \end{matrix}$$

$W = \frac{1}{d}A$ has $\lambda_2(W^\top W) = \epsilon^2$

Proof (cont'd):It remains to show $\forall f$: $f$ is a probability distribution,

$$\|\Pi_B W \Pi_B f\|_2 \leq (\beta + \epsilon)\|f\|_2$$

Without loss of generality, we can assume $f$ is supported only on $B$.

$$\|\Pi_B W \Pi_B f\|_2 = \|\Pi_B W f\|_2 = \|\Pi_B W(u + v)\|_2 \leq \|\Pi_B u\|_2 + \|\Pi_B W v\|_2$$

$u = \frac{1}{n}\vec{1}$, so $\frac{\langle f,u\rangle}{\langle u,u\rangle}u = u$, then $v \perp \vec{1}$

Next, $\|\Pi_B W v\|_2 \leq \|Wv\|_2 \leq \epsilon\|v\|_2 \leq \epsilon\|f\|_2$.

On the other hand, $\|\Pi_B u\|_2 = \sqrt{\frac{\beta}{n}} \leq \beta\|f\|_2$,

where last inequality follows from Cauchy-Schwarz:

$$1 = \|f\|_1 = \langle 1_B, f\rangle \leq \sqrt{\beta n}\|f\|_2$$

Combined together, we have $\|\Pi_B W \Pi_B f\|_2 \leq (\beta + \epsilon)\|f\|_2$ as desired.

**Cauchy-Schwarz inequality**:

$$\langle u, v\rangle \leq \sqrt{\langle u, u\rangle} \cdot \sqrt{\langle v, v\rangle}$$

# Hitting property of expander

To get a tail bound, consider $P(S,t) = \left\|\Pi_{Z_t} W \Pi_{Z_{t-1}} W \dots \Pi_{Z_1} u\right\|_1$ where $S = (Z_t, Z_{t-1}, \dots, Z_1)$ indicates whether $Z_i \in \{B, \bar{B}\}$

Proof. Observe that $P(B,t) = \|(\Pi_B W)^t \Pi_B u\|_1$

Suppose that we can show $\forall f$: $f$ is a probability distribution, we have $\|\Pi_B W \Pi_B f\|_2 \le (\beta + \epsilon)\|f\|_2$. Then,

$$\|(\Pi_B W)^t \Pi_B u\|_1 \le \sqrt{n}\|(\Pi_B W)^t \Pi_B u\|_2 = \sqrt{n}\|(\Pi_B W \Pi_B)^t u\|_2 \le \sqrt{n}(\beta + \epsilon)^t \|u\|_2 = (\beta + \epsilon)^t$$

It remains to show $\forall f$: $f$ is a probability distribution,

$$\|\Pi_B W \Pi_B f\|_2 \le (\beta + \epsilon)\|f\|_2$$

Without loss of generality, we can assume $f$ is supported only on $B$.

$$\|\Pi_B W \Pi_B f\|_2 = \|\Pi_B W f\|_2 = \|\Pi_B W (u + v)\|_2 \le \|\Pi_B u\|_2 + \|\Pi_B W v\|_2$$

Next, $\|\Pi_B W v\|_2 \le \|W v\|_2 \le \epsilon \|v\|_2 \le \epsilon \|f\|_2$.

On the other hand, $\|\Pi_B u\|_2 = \sqrt{\frac{\beta}{n}} \le \beta \|f\|_2$,

The last inequality follows from Cauchy-Schwarz:

$$1 = \|f\|_1 = \langle 1_B, f \rangle \le \sqrt{\beta n}\|f\|_2$$

Combined together, we have $\|\Pi_B W \Pi_B f\|_2 \le (\beta + \epsilon)\|f\|_2$ as desired.