Lecture 29

Recall:

- P transition matrix is similar to symm. normalised adj. matrix.

$$P = D^{-1}A = D^{-1/2}\tilde{A}D^{1/2}, \quad \tilde{A} = D^{-1/2}AD^{-1/2}$$

Spectrum $1 = \alpha_1 \geq \cdots \geq \alpha_n \geq -1$

Lazy random walk: $\tilde{P} = P/2 + I/2$

$$1 = \tilde{\alpha}_1 \geq \cdots \geq \tilde{\alpha}_n \geq 0$$

- random walk on any connected non-bipartite graph

$\rightsquigarrow$ stationary distribution $\pi = \frac{d}{1^T d}$

- mixing rate $\mu = \limsup_{t\to\infty} \max_{i,j} \left| P^t(i,j) - \pi(j) \right|^{1/t}$

Thm: $\left| P^t(i,j) - \pi(j) \right| \leq \sqrt{\frac{\pi(j)}{\pi(i)}} \cdot \mu^t$, where $\mu = \max\{|\alpha_2|, |\alpha_n|\}$

$\Rightarrow$ mixing rate $\leq \mu$.

A more quantitative version, mixing time, measures how many steps needed to get close to stationary dist.

• What is a good measure of distance btw distributions?

Natural candidate: Euclidean ℓ_2-norm $\|x\|_2 = \left(\sum x_i^2\right)^{1/2}$

Not ideal here.

Ex 1 Consider $S = [n/2]$

$x = \frac{2}{n} 1_S$

$y = \frac{2}{n} 1_{S^c}$

	S		S^c	V
	$1 \quad \cdots$	$n/2$	$n/2+1 \quad \cdots$	n
x	$\frac{2}{n} \; \frac{2}{n} \; \cdots$	$\frac{2}{n}$	$0 \; 0 \; 0$	0
y	$0 \; 0 \; \cdots$	0	$\frac{2}{n} \; \frac{2}{n} \; \cdots$	$\frac{2}{n}$

$$\|x-y\|_2 = \sqrt{\sum \left(\frac{2}{n}\right)^2} = \frac{2}{\sqrt{n}} \longrightarrow 0 \quad \text{as } n \to \infty$$

but x, y as distr. are very different.

• Better one: a scaled ℓ_1-norm.

Def: The total variation distance between x, y is

$$\|x-y\|_{TV} = \max_{S \subseteq V} \left| \sum_{v \in S} x(v) - \sum_{v \in S} y(v) \right|$$

max. difference of probabilities of events w.r.t. x & y

Ex $\|x-y\|_{TV} = \frac{1}{2} \|x-y\|_1$

In the above example, $\|x-y\|_{TV} = 1$.

distributions w/ disjoint support have distance 1.

Def: A random walk mixes at time t if

$$\|P_t - \pi\|_{TV} < \frac{1}{4} \quad \cdots\cdots \quad (*)$$

Call such t the mixing time.

useful equiv. form

$$\forall i, \quad |P_t(i) - \pi(i)| \le \frac{\pi(i)}{2}$$

Rmk: The constant $1/4$ is not that important. Any small constant will do. $\not\exists$ $(*)$ means that

for most of $i \in V$, $\frac{\pi(i)}{2} \le P_t(i) \le 2\pi(i)$

Consider the simpler d-regular $G=(V,E)$ case

do lazy random walk, no need to worry about α_n.

Recall $1 = \alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_n \Leftarrow \hat{A}$

$$\boxed{\nu_2 = 1 - \alpha_2 > 0} \qquad \tilde{L} = N = I - \hat{A} = I - D^{-1/2} A D^{-1/2}$$

Assume $\exists$ spectral gap.

$$\boxed{\left|P^t(i,j) - \pi(j)\right| \le \sqrt{\frac{d(j)}{d(i)}}\, \mu^t}$$

$\mu = \alpha_2 = 1 - \nu_2$

$\left(\text{WANT } \le \frac{\pi(j)}{2}\right)$

want $\le \frac{d(j)}{2 \cdot \sum_{i=1} d(i)}$

$\sum_{i=1} d(i) = 2e(G)$

$$e^{-\nu_2 t} \approx (1-\nu_2)^t$$

$$\le \sqrt{\frac{d(j)}{d(i)}} \cdot e^{-\nu_2 t}$$

$$\Longleftrightarrow \quad e^{\gamma_2 t} \geq \frac{4e(G)}{\sqrt{d(i)\,d(j)}}$$

$$\Rightarrow \quad t \geq \frac{1}{\gamma_2} \log \boxed{\frac{4e(G)}{\sqrt{d(i)d(j)}}}$$

d-reg $2n$

$$= \frac{1}{\gamma_2} \underline{\log(2n)}.$$

Rmk: Sometimes, the $\log n$ term can be avoided as we use α_2 to bound all α_i, $i \geq 2$.

If there is eigenvalue decay, then we can take advantage to improve the bound.

EXAMPLE. Q_d $\quad n = 2^d$, $\quad \gamma_2 = \frac{2}{d}$

The above bound on mixing time yields

$$t = O\left(\frac{\log n}{\gamma_2}\right) = O(d^2)$$

But the mixing time for hypercube is known to be $\Theta(d \log d)$.

d-reg

$$\phi(G) = \min_{|S| \leq n/2} \frac{|\partial S|}{d|S|}$$

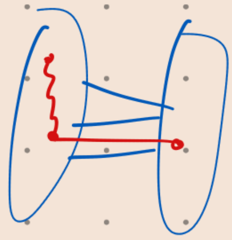
— Conductance & mixing time —

$$\frac{1}{\phi(G)} \leq t_{mix} \leq \frac{\log n}{\phi(G)^2}$$

Lovász–Simonovits 90.

natural

$$\mathbb{E}(\#\text{steps to cross}) \sim \frac{1}{\mathbb{P}(\text{cross})} = \frac{1}{\phi(G)}$$



S S^c

sparse cut

$\phi(S, S^c)$
$= \phi(G)$

This offers a probabilistic way to define an expander: if the random walk on it is fast mixing.

Application Make use of the following fact
"random walk on expanders resembles independent sampling".

From computational aspect, it's used in error reduction for probabilistic algorithm.

Suppose we have a prob. alg. A.

$$A : \overset{\text{uses}}{K \text{ random bits}} \longrightarrow \begin{cases} \text{right answer} & \overset{\text{w/ prob.}}{2/3} \\ \text{wrong} & 1/3 \end{cases}$$

To boost the prob., we can run it t times and take the majority answer.

$\mathbb{P}\left(\text{we get the wrong answer}\right) \le (1-c)^t, \quad c>0.$

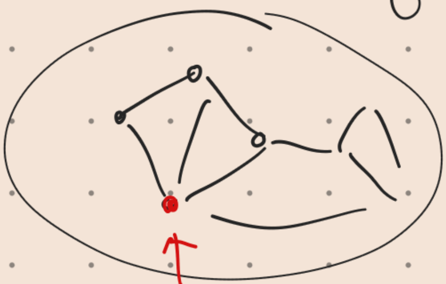
Cost : $t \cdot k$ random bits.

More economic : $k + O(t)$

initial k random bits

$G = (V, E)$ $\left\{ \begin{array}{l} \text{constant deg expander} \\ V = \{0,1\}^k \end{array} \right.$

G



random starting vertex v_0

do t steps random walk.

$\Rightarrow$ t random vxs (t strings of random k-bit).

— Plan —

Two classes $\left\{ \begin{array}{l} \text{- concentration of measure} \\ \quad \text{- pf ----- on a slice} \\ \quad \text{- application: coding theory} \\ \text{- Johnson-Lindenstrauss lemma} \\ \quad \vdots \end{array} \right.$

Concentration of measure

a.k.a. large deviation inequalities.

Basic: Markov: X non-negative r.v.

1st moment $\nearrow$ $\forall a > 0$

$$\Rightarrow \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}$$

2nd moment

Chebyshev: X r.v., w/ finite $\mathbb{E}X = \mu$, $\mathrm{Var} X = \sigma^2$

$$\Rightarrow \forall k > 0, \quad P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- Start w/ simplest setup.

$$X = \begin{cases} 1 & \frac{1}{2} \\ -1 & \frac{1}{2} \end{cases} \quad \text{prob}$$

Consider sum of Rademacher r.v. X_i

$$S_n = X_1 + \cdots + X_n \quad \text{indep.}$$

Trivially: $-n \leq S_n \leq n$

Typically we shall see that S concentrates sharply within a window of width $O(\sqrt{n})$

Intuition: it is very rare that all (indep) r.v. X_i team up to go in the same direction

For instance: $\mathbb{P}(S_n = n) = \mathbb{P}(\text{all } X_i = 1) = 2^{-n}$

exponentially unlikely.

General phenomenon: assuming $\begin{cases} \text{boundedness} \\ \text{suff. independence} \end{cases}$

$\Rightarrow$ concentration of measure.

$\rightsquigarrow$ usually of sub-gaussian nature

i.e. $\mathbb{P}(\lambda \underline{\underline{\sigma}}$ away from mean $\mu) \le c_1 \cdot e^{-c_2 \lambda^2}$.

standard deviation