

SPECTRAL GRAPH THEORY (CONT)

Recap

- ① Our goal was to maximize the local variance $\mathbb{E}[f] = \langle f, Lf \rangle$ & we subjected this to the constraint $\|f\|_2^2 = 1$ (to guarantee the existence of a maxima)

We proved that for a ϕ maximizing $\mathbb{E}[f]$ with $\|\phi\|_2^2 = 1$, $L\phi = \lambda\phi$ for some $\lambda \in \mathbb{R}$

- ② Constructively, we saw the theorem (spectral):
Given graph G , $\forall$ orthonormal functions $\mathbb{1} = \phi_0, \phi_1, \dots, \phi_{n-1}, \phi_n : V \rightarrow \mathbb{R}$ and real numbers $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-2} \leq \lambda_{n-1} \leq 2$ such that
 $L\phi_i = \lambda_i \phi_i \quad \forall i$

$\{\phi_0, \phi_1, \dots, \phi_{n-1}\}$ is an orthonormal basis for $\{f: V \rightarrow \mathbb{R}\}$.

- ③ For $S \subseteq V$, conductance of S is
$$\Phi(S) = \Pr_{u \sim v} [v \notin S | u \in S] = \frac{\mathbb{E}[\mathbb{1}_S]}{\text{Vol}(S)} \Big|_{\mathbb{1} \in \mathbb{1}_S}$$

The "sparsest cut" problem is to determine the min-conductance: $\Phi_G = \min_{S: 0 < \text{Vol}(S) \leq \frac{1}{2}} \{\Phi(S)\}$ (Prof. Akash's lecture)

Cheeger's inequality (Chalk & talk): $\frac{1}{2} \lambda_1 \leq \Phi_G \leq \sqrt{2 \lambda_1}$

- ④ Intuitively, this means λ_1 "large" $\Rightarrow$ "fast mixing" of random walk.

But λ_1 "large" $\Leftrightarrow K_1 = 1 - \lambda_1$ "far from" 1 $\Leftrightarrow |K_1| \leq 1 - \epsilon$
(Exception: Bipartite graph) $\forall i \in \{1, \dots, n-1\}$

Recall: $K = I - L$ has eigenvectors $\underline{1} = \phi_0, \phi_1, \dots, \phi_{n-1}$ & eigenvalues $1 = \kappa_0 \geq \kappa_1 \geq \dots \geq \kappa_{n-1} = -1$. ($\kappa_0 = 1 - \lambda_0$)

The main theorem that we will prove today is:

Theorem (Convergence): Let $|\kappa_1| \leq 1 - \epsilon$ & $\epsilon > 0$ ($0 < \epsilon < 1$)

Then for any worst-case distribution p_0 on V

If $u_0 \sim p_0$ and $u_0 \rightarrow u_1 \rightarrow \dots \rightarrow u_t$ (t -random walk)
and

$t \geq \text{constant} \cdot \frac{\ln(n)}{\epsilon}$, then p_t (distribution of u_t)

is "very close" to π

↓
will define this today.
& then prove this theorem!

Remark: ① For bipartite graph G (for which we know convergence doesn't occur), $\kappa_{n-1} = -1$. But $\epsilon > 0$ in above theorem guarantees that $|\kappa_{n-1}|$ has to be bounded away from 1 $\Rightarrow$ not bipartite.

② Exercise: Show that G is connected iff and only iff $\lambda_1 > 0$ (i.e. $\kappa_1 < 1$)

∴ $\epsilon < 1$ in the above theorem ensures that G is not disconnected (in which case convergence again does not occur).

How to measure distance between two prob distributions?

There are many ways to measure distance b/w prob. distributions: Total variation distance, KL Divergence, Hellinger distribution, Bhattacharya distance, chi-squared divergence, etc. (Each useful in different fields, if interested you can read more to see what each of them try to capture & where all they are used)

Let's focus on our need:

For any two probability distributions f & π

"f is close to π " if $\frac{f(u)}{\pi(u)} \approx 1 \quad \forall u \in V$
 $\uparrow$
 "close to"

Defn (Chi-Squared Divergence)

For $f(u) = \frac{g(u)}{\pi(u)}$ (captures relative density of g w.r.t. π)
($f: V \rightarrow \mathbb{R}^{\geq 0}$), we define the chi-squared divergence:

$$d_{\chi^2}(\beta, \pi) = \mathbb{E}_{u \sim \pi} [(f(u) - 1)^2]$$

(clearly $d_{\mathcal{H}}(\beta, \pi) = 0 \Leftrightarrow \beta \equiv \pi$)

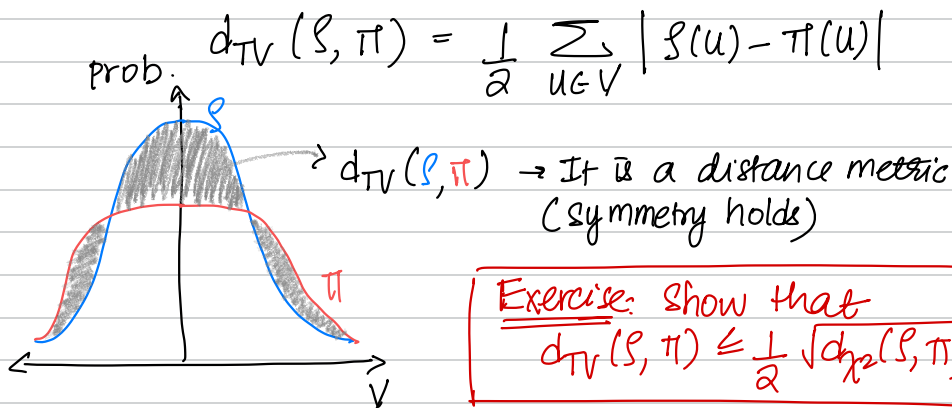
Remarks: (i) It is not a distance metric since it is not symmetric. Hence called χ^2 -divergence instead of calling it a distance.)

i.e., $d_{\mathcal{H}}(\mathcal{I}, \Pi) \neq d_{\mathcal{H}}(\Pi, \mathcal{I})$ (think intuitively why?)

↳ Exercise: Prove this by taking some specific β & π .

- ② There is a more natural distance metric that may come to your mind (similar to ℓ_1 -norm):

Total Variation Distance :



Exercise: Show that

$$d_{TV}(\rho, \pi) \leq \frac{1}{2} \sqrt{\chi^2(\rho, \pi)}$$

But, we will focus on chi-squared divergence.
(because of the exercise you can see that closeness in chi-squared divergence $\Rightarrow$ closeness in TV-distance)

Why not other distances? $\rightarrow$ they are considered in different contexts (beyond scope of this lecture)

- ③ For any probability distribution ρ , for $f: V \rightarrow \mathbb{R}^{\geq 0}$ defined as $f(u) = \frac{\rho(u)}{\pi(u)}$

(Recall this $\leftarrow$ $= f(0)$) $\mathbb{E}_{u \sim \pi} [f(u)] = 1$

Proof: $\mathbb{E}_{u \sim \pi} [f(u)] = \sum_{u \in V} \pi(u) \cdot \left(\frac{\rho(u)}{\pi(u)} \right) = \sum_{u \in V} \rho(u) = 1$ ■

- ④ (Corollary of ③) : $\chi^2(\rho, \pi) = \text{Var}_{u \sim \pi} [f(u)]$

Proof: $\chi^2(\rho, \pi) = \mathbb{E}_{u \sim \pi} [(f(u) - 1)^2] = \mathbb{E}_{u \sim \pi} [(f(u) - \mathbb{E}[f(u)])^2]$

$$= \text{Var}_{u \sim \pi} [f(u)] \quad (\text{by defn})$$

We also know from last lecture:

$$\chi^2(S, \pi) = \text{Var}_{u \sim \pi} [f(u)] = \sum_{0 \leq i \leq n} \hat{f}(i)^2$$

Proof of the Convergence Theorem

Idea:

We want to prove that starting at S_0 , taking t -steps from $u_0 \sim S_0$ will lead to distributions S_t that will be close to π (in chi-squared divergence).

Instead of tracking $S_0 \xrightarrow{1} S_1 \xrightarrow{2} \dots \xrightarrow{t} S_t$, we want to look at $f_0 \rightarrow f_1 \rightarrow \dots \rightarrow f_t$, where $f_t(u) = \frac{S_t(u)}{\pi(u)}$ (the "densities") $\forall u \in V$.

Then we want to show: For large t , $(\chi^2(S_t, \pi) =) \text{Var}_{u \sim \pi} [f_t(u)]$ is small.

Proof: Start with any worst case S_0 . Let $f_0(u) = \frac{S_0(u)}{\pi(u)} \forall u$.

Then, by the spectral theorem, $f_0: V \rightarrow \mathbb{R}$ can be expressed as

$$f_0 = 1 + \hat{f}_0(1)\phi_1 + \hat{f}_0(2)\phi_2 + \dots + \hat{f}_0(n-1)\phi_{n-1}$$

Now, we need to find what f_1 is?

$$(f_1(u) = \frac{S_1(u)}{\pi(u)})$$

Claim: $f_1 = K f_0$.

Proof of claim: $K f_0(u) = \mathbb{E}_{v \sim u} [f_0(v)]$

$$\Rightarrow K f_0(u) = \sum_{v: u \rightarrow v} K_{uv} \cdot \frac{S_0(v)}{\pi(v)}$$

edges between u & v
deg(u)

(Recall: K_{uv} is prob that v is picked as random neighbor of u .)
deg(v)
2|E|

$$= \sum_{v: u \rightarrow v} \frac{\text{\# edges b/w } u \text{ & } v}{\deg(v)} \cdot \frac{2|E|}{\deg(u)} \cdot S_0(v)$$

(Why? Recall K is self-adjoint)
($\because$ we saw $S_1 = S_0 \cdot K$)

$$= \sum_{v: v \rightarrow u} K_{vu} \cdot \frac{S_0(v)}{\pi(u)}$$

$$= \frac{S_1(u)}{\pi(u)} = f_1(u)$$

$$\begin{aligned} \because f_1 &= K f_0 \\ &= \mathbb{1} + \hat{f}_0(1) K \phi_1 + \hat{f}_0(2) K \phi_2 + \dots + \hat{f}_0(n-1) K \phi_{n-1} \\ &= \mathbb{1} + \hat{f}_0(1) \kappa_1 \phi_1 + \hat{f}_0(2) \kappa_2 \phi_2 + \dots + \hat{f}_0(n-1) \kappa_{n-1} \phi_{n-1} \end{aligned}$$

Similarly, we get

$$\begin{aligned} f_t &= K^t f_0 \\ &= \mathbb{1} + \kappa_1^t \hat{f}_0(1) \phi_1 + \kappa_2^t \hat{f}_0(2) \phi_2 + \dots + \kappa_{n-1}^t \hat{f}_0(n-1) \phi_{n-1} \\ \Rightarrow \hat{f}_t(i) &= \kappa_i^t \cdot \hat{f}_0(i) \quad \forall i=1, 2, \dots, n-1 \quad \rightarrow \textcircled{1} \end{aligned}$$

By hypothesis of the theorem, $|\kappa_i| \leq 1-\epsilon \quad \forall i \geq 1$ $\rightarrow \textcircled{2}$

$$\chi^2(S_t, \pi) = \text{Var}_{u \sim \pi} [f_t(u)]$$

$$= \sum_{0 \leq i \leq n} \hat{f}_t(i)^2 = \sum_{i=1}^{n-1} \kappa_i^{2t} \cdot \hat{f}_0(i)^2 \quad (\text{By } \textcircled{1})$$

$$\leq \max_{i=1, \dots, n-1} \kappa_i^{2t} \cdot (\hat{f}_0(1)^2 + \hat{f}_0(2)^2 + \dots + \hat{f}_0(n-1)^2)$$

$$\text{By } \textcircled{2}, \max_{1 \leq i \leq n-1} \kappa_i^{2t} \leq (1-\epsilon)^{2t}$$

we saw that this is $\chi^2(S_0, \pi)$

$$\therefore d_{\chi^2}(S_t, \pi) \leq (1-\epsilon)^{2t} \cdot d_{\chi^2}(S_0, \pi) \leq \exp(-2t\epsilon) \cdot d_{\chi^2}(S_0, \pi) \rightarrow \textcircled{2}$$

→ Upper bounding $d_{\chi^2}(S_0, \pi)$

Claim: for any distribution S_0 on V ,

$$d_{\chi^2}(S_0, \pi) \leq \min_{u \in V} \frac{1}{\pi(u)}$$

Intuition: the largest value of $d_{\chi^2}(S_0, \pi)$ is attained for S_0^* that outputs a particular vertex u^* with probability 1.

For such S_0^* : $f(u) = \begin{cases} \frac{1}{\pi(u^*)}, & \text{if } u = u^* \\ 0, & \text{otherwise} \end{cases}$

$$\begin{aligned} d_{\chi^2}(S_0^*, \pi) &= \text{Var}[f] \leq \mathbb{E}[f(u)^2] \\ &= \pi(u^*) \cdot \frac{1}{\pi(u^*)^2} = \frac{1}{\pi(u^*)} \end{aligned}$$

Proof → Exercise

(Hint $G(p) = d_{\chi^2}(p, \pi)$ (for any dist. p) is a convex function.)

Using above claim in eq $\textcircled{2}$, we get

$$d_{\chi^2}(S_t, \pi) \leq \exp(-2t\epsilon) \cdot \min_{u \in V} \left\{ \frac{1}{\pi(u)} \right\}$$

Examples: ① If G is d -regular, $\pi(u) = \frac{1}{n} \forall u$

$$\Rightarrow \frac{1}{\pi(u)} = n \quad \forall u$$

$$\therefore d_{\chi^2}(S_t, \pi) \leq \exp(-2t\epsilon) \cdot n$$

$\therefore \text{For } t \geq \underset{\substack{\uparrow \\ \text{some constant}}}{C \cdot \frac{\ln(n)}}{\epsilon}, d_{\chi^2}(\mathcal{P}_t, \Pi) \leq \frac{1}{n^{2c-1}}$
(very small)

This proves the theorem for regular graphs. ■

In general for any graph G

$$d_{\chi^2}(\mathcal{P}_t, \Pi) \leq \exp(-2t\epsilon) \cdot 2 \cdot |E| \leq \exp(-2t\epsilon) \cdot n^2$$

$\left(\begin{array}{l} \because \deg(u) \geq 1 \forall u \\ \frac{1}{\Gamma(u)} = \frac{2|E|}{\deg(u)} \leq 2|E| \end{array} \right)$

$\therefore \text{If } t \geq \underset{\substack{\uparrow \\ \text{(some constant)} \\ c > 1}}{C \cdot \frac{\ln(n)}}{\epsilon}, d_{\chi^2}(\mathcal{P}_t, \Pi) \leq \frac{1}{n^{2c-2}}$
■
