Chapter V Higher order Cheeger inequalities

§1 Motivation, basic definitions and properties

Important task in data analysis: Clustering

Given: data points $x_1, \dots, x_n$, notion of similarity $s_{ij} \geq 0$ between all pairs of data points x_i, x_j

Goal: Divide data points into groups ("clusters") s.t.

- points in same group are similar
- points in different groups not similar

Similarity graph: weighted graph $G = (V, E, w)$

$V = \{x_1, \dots, x_n\}$,

$E = \{\{x_i, x_j\} : s_{ij} > 0\}$

$w(x_i, x_j) = s_{ij}$

$\longrightarrow$ Find partition $V = U_1 \dot\cup \dots \dot\cup U_r$:

- $w(x_i, x_j)$ large if $x_i, x_j \in U_t$
- $w(x_i, x_j)$ small if $\{x_i, x_j\} \in \delta(U_t)$

RECALL: Edge expansion ratio / Cheeger constant (Chapter III)

$$h(G) = \min_{\substack{U \subseteq V, \\ U \neq \emptyset, V}} \frac{|\delta(U)|}{\min\{|U|, |V \setminus U|\}}$$

for unweighted graph G. (weights are 0 or 1)

$\longrightarrow$ Partition $V = U \,\dot\cup\, V \setminus U$ with few edges in $\delta(U)$

Definition: $G = (V, E, w)$ weighted graph

- $$\phi_G(U) = \frac{w(\delta(U))}{\sum_{u \in U} \sum_{e \ni v} w(e)}$$ "expansion" of $U \subseteq V$

- $$\rho_G(r) = \min_{\substack{V = U_1 \dot\cup \dots \dot\cup U_r, \\ U_i \neq 0}} \max_{i = 1, \dots, r} \phi_G(U_i)$$

Here, we focus on

k-regular graphs with 0/1-weights:

Then: $$\rho_G(2) = \min_{\substack{V = U_1 \dot\cup U_2, \\ U_i \neq \emptyset}} \max_{i=1,2} \frac{|\delta(U_i)|}{k \cdot |U_i|} = \frac{1}{k} h(G)$$

By Cheeger's inequality: $[k = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ eigenvalues of $A]$

$$\frac{k - \lambda_2}{2} \leq k \cdot \rho_G(2) \leq \sqrt{2k(k - \lambda_2)}$$

So, λ_2 connected to partitioning into two clusters.

Already saw: Multiplicity of k as eigenvalue
= number of connected compontents

$\rightsquigarrow$ λ_r connected to partitioning into r clusters?

Goal of this chapter: "Higher order Cheeger's inequality"

Thm: $G = (V, E)$ k-regular graph without loops, $r \leq n$

Then: $\exists$ partition $V = U_1 \dot\cup U_2 \dot\cup \dots \dot\cup U_r$ s.t.

$$\phi_G(U_i) \leq c \cdot r^3 \sqrt{\tfrac{1}{k}(k - \lambda_r)} \quad \text{for } i = 1, \dots, r,$$

for some constant c independent of r and $|V|$.

So: $\quad k \cdot \rho_G(r) \leq c \cdot r^3 \sqrt{k(k-\lambda_r)}$

Remarks:

- Corresponding statement true for weighted, non-regular graphs (& same proof works), instead of $\Delta = kI - A$, use $D - A$, D diagonal matrix with $D_{v,v} = \sum_{e \ni v} w(e)$).

- Proof similar to what is actually done in practice: Use different clustering algorithm (not "k-means") that is more easily analyzed.

- Also have $\frac{k - \lambda_r}{2} \leq k \cdot \rho_G(r)$ ($\rightarrow$ exercise)

§2 Low diameter partitioning in $\mathbb{R}^r$

In proof of higher order Cheeger inequality:

Map graph to $\mathbb{R}^r$ and cluster points in $\mathbb{R}^r$.

$\rightarrow$ Use randomized algorithm that guarantees <u>low (Euclidean) diameter</u>:

<u>INPUT</u>: $P \subseteq \mathbb{R}^r$, $|P| = n$, $\delta > 0$

<u>OUTPUT</u>: Partition of P with diam(clusters) $\leq \delta$

<u>ALGORITHM</u>:
- Balls of radius $\frac{\delta}{2}$ around points in P: B_x, $x \in P$.
- Choose p uniformly at random from $M = \bigcup_{x \in P} B_x$:

$$\Pr[p \in A] = \frac{\operatorname{vol}[A \cap M]}{\operatorname{vol}[M]} \quad (\text{for } A \text{ measurable})$$

- Form cluster:

$$C = \{x \in P : \|x - p\| \leq \tfrac{\delta}{2},\ x \text{ not in any previous cluster}\}$$

Repeat until every point in some cluster.

<u>Lemma</u>: $\Pr[\text{Get a partition after } N \text{ rounds}] \geq 1 - n\left(\frac{n-1}{n}\right)^N \xrightarrow[N \to \infty]{} 1$ (for n fixed)

<u>Proof</u>: $\longrightarrow$ exercises

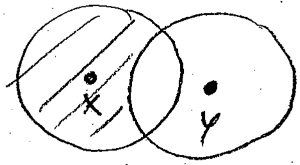
$\rightarrow \Pr[\text{Get a partition in a finite number of steps}] = 1$

<u>Lemma</u>: For $x, y \in P$ fixed, $x \neq y$:

$$\Pr[x, y \text{ in different clusters}] \leq 2 \cdot \|x-y\| \cdot \frac{\sqrt{r}}{\delta}$$

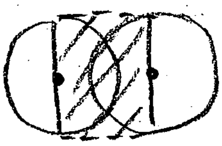
<u>Proof:</u>

If x, y not in any cluster yet:



x lands in cluster and y not if $p \in B_x \setminus B_y$.

- $\mathrm{vol}(B_x \setminus B_y) = \underbrace{\mathrm{vol}(B_x \cup B_y)}_{\leq \mathrm{vol}(B^r_{\delta/2}) + \mathrm{vol}(C)} - \mathrm{vol}(B_y) \leq \mathrm{vol}(C) = \|x-y\| \cdot \mathrm{vol}(B^{r-1}_{\delta/2})$



↑ cylinder C of height $\|x-y\|$

- $\Pr[x, y \text{ in different cluster}, x \text{ chosen in an earlier round than } y]$

$$= \sum_{j=1}^{\infty} \frac{\mathrm{vol}(B_x \setminus B_y)}{\mathrm{vol}(M)} \cdot \prod_{i=1}^{j-1} \frac{\mathrm{vol}(M \setminus (B_x \cup B_y))}{\mathrm{vol}(M)}$$

$$= \frac{\mathrm{vol}(B_x \setminus B_y)}{\mathrm{vol}(M)} \cdot \underbrace{\sum_{j=0}^{\infty} \left(\frac{\mathrm{vol}(M \setminus (B_x \cup B_y))}{\mathrm{vol}(M)} \right)^j}_{= \frac{1}{1 - \frac{\mathrm{vol}(M \setminus (B_x \cup B_y))}{\mathrm{vol}(M)}} = \frac{\mathrm{vol}(M)}{\mathrm{vol}(B_x \cup B_y)}}$$

$$= \frac{\mathrm{vol}(B_x \setminus B_y)}{\mathrm{vol}(B_x \cup B_y)}$$

can be shown...

$$\leq \|x-y\| \cdot \frac{\mathrm{vol}(B^{r-1}_{\delta/2})}{\mathrm{vol}(B^r_{\delta/2})} \leq \|x-y\| \cdot \frac{\sqrt{r}}{\delta}$$

□

§3 Proof of the higher order Cheeger's inequality

Idea: Use spectral embedding of G in $\mathbb{R}^r$:

- Choose f_i, $i=1,\dots,n$, ONB of eigenvectors of A: $\quad Af_i = \lambda_i f_i, \quad \|f_i\|^2 = 1$.
- $$F: V \to \mathbb{R}^r$$
  $$v \mapsto (f_1(v), \dots, f_r(v))$$

Relation of F to eigenvalues:

Definition: Rayleigh quotient of F (w.r.t. G):

$$R_G(F) := \frac{\sum_{\{u,v\}\in E} \|F(u)-F(v)\|^2}{\sum_{v\in V} \|F(v)\|^2}$$

Lemma: $R_G(F) \le k - \lambda_r$

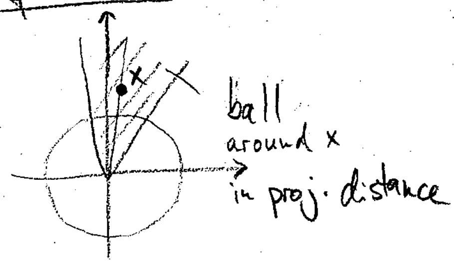
Proof:
$$\frac{\sum_{\{u,v\}\in E} \|F(u)-F(v)\|^2}{\sum_{v\in V} \|F(v)\|^2} = \frac{\sum_{i=1}^{r}\sum_{\{u,v\}\in E} (f_i(u)-f_i(v))^2}{\sum_{i=1}^{r}\sum_{v\in V} f_i(v)^2}$$

Recall: • $\Delta = kI - A = d^* d$, d: simplicial coboundary operator, defined by incidence matrix

• $(\Delta f, f) = (df, df) = \sum_{\{u,v\}\in E} (f(u)-f(v))^2$

So: $$R_G(F) = \frac{\sum_{i=1}^{r}(\Delta f_i, f_i)}{\sum_{i=1}^{r}(f_i,f_i)} = \frac{1}{r}\sum_{i=1}^{r}(k-\lambda_i) \le k - \lambda_r$$ $\square$

Definition: $d_F(u,v) = \left\| \frac{F(u)}{\|F(u)\|} - \frac{F(v)}{\|F(v)\|} \right\|$ "radial projection distance"



Lemma: $\forall\, \delta \in (0,1],\ S \subseteq V$:

$$\operatorname{diam}(S, d_F) \le \delta \implies \sum_{v \in S} \|F(v)\|^2 \le \frac{1}{1-\delta^2}$$

Proof: For any $x \in \mathbb{R}^r$, have:

$$\sum_{v \in V} \|F(v)\|^2 \left\langle x, \frac{F(v)}{\|F(v)\|} \right\rangle^2 = \sum_{v \in V} \langle x, F(v) \rangle^2$$

$$= \sum_{v \in V} \sum_{i,j} x_i x_j f_i(v) f_j(v)$$

$$= \sum_i x_i \sum_j x_j \underbrace{\sum_{v \in V} f_i(v) f_j(v)}_{= \langle f_i, f_j \rangle = \delta_{ij}}$$

$$= \sum_i x_i^2 = \|x\|^2$$

For $u \in S$ consider $x = \frac{F(u)}{\|F(u)\|}$:

$$1 = \sum_{v \in V} \|F(v)\|^2 \left\langle \frac{F(u)}{\|F(u)\|}, \frac{F(v)}{\|F(v)\|} \right\rangle^2$$

$$\ge \sum_{v \in S} \|F(v)\|^2 \underbrace{\left\langle \frac{F(u)}{\|F(u)\|}, \frac{F(v)}{\|F(v)\|} \right\rangle^2}_{\substack{\ge 1 - \left\| \frac{F(u)}{\|F(u)\|} - \frac{F(v)}{\|F(v)\|} \right\|^2 \ (\to \text{exercises}) \\ \ge 1-\delta^2 \text{ for } u,v \in S}} \ge (1-\delta^2) \sum_{v \in S} \|F(v)\|^2$$

$\square$

Now, combine spectral embedding and randomized partitioning algorithm:

Lemma: $\forall \delta \in (0,1)$ $\exists$ partition $V = S_1 \dot\cup S_2 \dot\cup \dots \dot\cup S_m$:

① $\sum_{v\in S_i} \|F(v)\|^2 \le \frac{1}{1-\delta^2}$ for $i=1,\dots,m$

② For $\tau \in [0, \max\{\|F(v)\|^2 : v\in V\})$ chosen uniformly at random:

$$\frac{\mathbb{E}\left[\sum_{i=1}^m |\partial(\hat{S}_i)|\right]}{\mathbb{E}\left[\sum_{i=1}^m |\hat{S}_i|\right]} \le C \cdot \frac{\sqrt{r}}{\delta} \sqrt{k(k-\lambda_r)}$$

where $\hat{S} := \{v\in S : \tau \le \|F(v)\|^2\}$ for $S \subseteq V$.

PROOF: For any partition $V = S_1 \dot\cup \dots \dot\cup S_m$:

- $\mathbb{E}[|\hat{S}_i|] = \mathbb{E}\left[\sum_{v\in S_i} \mathbb{1}_{(0,\|F(v)\|^2)}\right] = \sum_{v\in S_i} \mathbb{E}[\mathbb{1}_{(0,\|F(v)\|^2)}]$
$$= \sum_{v\in S_i} \frac{1}{M} \|F(v)\|^2$$

- $\mathbb{E}\left[\sum_{i=1}^m |\hat{S}_i|\right] = \frac{1}{M} \cdot \sum_{v\in V} \|F(v)\|^2$

Now, consider points $\left\{\frac{F(v)}{\|F(v)\|} : v\in V\right\} \subseteq \mathbb{R}^r$:

$\rightsquigarrow$ Random partition $V = S_1 \dot\cup \dots \dot\cup S_m$ s.t.

- $\operatorname{diam}(S_i, d_F) \le \delta$ for $i=1,\dots,m$
- $\Pr[u,v \text{ in different clusters}] \le 2 \cdot \frac{\sqrt{r}}{\delta} d_F(u,v)$

By previous lemma: Any such partition satisfies ①.
Want to show: There is one that satisfies ②.

Strategy

- Use: $\mathbb{E}[\text{random variable}] \leq a \Rightarrow$ Random variable $\leq a$ for some sample

- Will show, for $X = \mathbb{E}_\tau\left[\sum_{i=1}^{m} |\delta(\hat{S}_i)|\right]$, $Y = \mathbb{E}_\tau\left[\sum_{i=1}^{m} |\hat{S}_i|\right]$:

$$\underbrace{\frac{\mathbb{E}_{\{S_i\}}[X]}{\mathbb{E}_{\{S_i\}}[Y]}}_{\neq\ \mathbb{E}_{\{S_i\}}\left[\frac{X}{Y}\right]} \leq c\,\frac{\sqrt{r}}{\delta}\sqrt{k(k-d_r)}$$

$\neq \mathbb{E}_{\{S_i\}}\left[\frac{X}{Y}\right]$, but:

$$\Rightarrow \mathbb{E}_{\{S_i\}}\left[X - c\,\frac{\sqrt{r}}{\delta}\sqrt{k(k-d_r)}\cdot Y\right] \leq 0$$

$\Rightarrow \exists$ partition $V = S_1 \dot\cup \ldots \dot\cup S_m$ given by randomized algorithm:

$$X - c\,\frac{\sqrt{r}}{\delta}\sqrt{k(k-d_r)}\ Y \leq 0$$

Have seen:

$$\mathbb{E}_{\{S_i\}}[Y] = \mathbb{E}_{\{S_i\}}\left[\frac{1}{M}\cdot\sum_{v\in V}\|F(v)\|^2\right] = \frac{1}{M}\cdot\sum_{v\in V}\|F(v)\|^2$$

Bounding $\mathbb{E}_{\{S_i\}}[X]$:

Fix $\{u,v\}\in E$ with $\|F(u)\|^2 \leq \|F(v)\|^2$:

$$\Pr_{\{S_i\},\tau}\left[\{u,v\}\in \bigcup_{i=1}^{m}\delta(\hat{S}_i)\right]$$

$$\leq \Pr_{\{S_i\}}[u,v \text{ in different clusters}]\cdot \Pr_\tau\left[\|F(u)\|^2 \geq \tau \text{ or } \|F(v)\|^2 \geq \tau\right]$$

$$+ \Pr_{\{S_i\}}[u,v \text{ in same cluster}]\cdot \Pr_\tau\left[\tau \in [\|F(u)\|^2, \|F(v)\|^2]\right]$$

$$\text{First summand} \le 2 \cdot \frac{\sqrt{r}}{\delta} d_F(u,v) \cdot \underbrace{\left(\frac{1}{M}\|F(u)\|^2 + \frac{1}{M}\|F(v)\|^2\right)}_{\le \frac{1}{M}\left(\|F(u)\| + \|F(v)\|\right)^2}$$

$$\le \frac{1}{M}\left(\|F(u)\| + \|F(v)\|\right) 2 \frac{\sqrt{r}}{\delta} \underbrace{\left\| \frac{F(u)}{\|F(u)\|} - \frac{F(v)}{\|F(v)\|} \right\| \left(\|F(u)\| + \|F(v)\|\right)}_{\le 2\|F(u) - F(v)\|}$$

$(\rightarrow \text{exercise})$

$$\text{Second summand} \le 1 \cdot \frac{1}{M}\left(\|F(v)\|^2 - \|F(u)\|^2\right) = \frac{1}{M}\left(\|F(u)\| + \|F(v)\|\right)\underbrace{\left(\|F(u)\| - \|F(v)\|\right)}_{\le \|F(u) - F(v)\|}$$

triangle inequ.

So:

$$\Pr_{\{S_i\},\tau}\left[\{u,v\} \in \bigcup_{i=1}^m \delta(\hat{S}_i)\right] \le \frac{1}{M}\left(\|F(u)\| + \|F(v)\|\right)\left(\|F(u) - F(v)\|\right)\left(\frac{4\sqrt{r}}{\delta} + 1\right)$$

Because $\sum_{i=1}^m |\delta(\hat{S}_i)| \le 2 \left| \bigcup_{i=1}^m \delta(\hat{S}_i) \right| = 2 \cdot \sum_{\{u,v\} \in E} \mathbb{1}_{\{\{u,v\} \in \bigcup_i \delta(\hat{S}_i)\}}$

$$\mathbb{E}_{\{S_i\}} \mathbb{E}_\tau \left[\sum_{i=1}^m |\delta(\hat{S}_i)|\right] \le 2 \cdot \sum_{\{u,v\} \in E} \Pr\left[\{u,v\} \in \bigcup_{i=1}^m \delta(\hat{S}_i)\right]$$

$$\le \frac{2}{M}\left(4\frac{\sqrt{r}}{\delta} + 1\right) \underbrace{\sum_{\{u,v\} \in E} \left(\|F(u)\| + \|F(v)\|\right) \|F(u) - F(v)\|}$$

$$\underset{CS}{\le} \underbrace{\sqrt{\sum_{\{u,v\} \in E} \underbrace{\left(\|F(u)\| + \|F(v)\|\right)^2}_{\le 2\left(\|F(u)\|^2 + \|F(v)\|^2\right)}}}_{\le 2 \sum_{v \in V} k \cdot \|F(v)\|^2} \sqrt{\sum_{\{u,v\} \in E} \|F(u) - F(v)\|^2}$$

We get:

$$\frac{\mathbb{E}_{\{S_i\}}[X]}{\mathbb{E}_{\{S_i\}}[Y]} \leq 2\left(4\frac{\sqrt{r}}{\delta}+1\right)\sqrt{2k}\underbrace{\sqrt{\frac{\sum_{\{u,v\}\in E}\|F(u)-F(v)\|^2}{\sum_{v\in V}\|F(v)\|^2}}}_{=\mathcal{R}_G(F)\,\leq\, k-\lambda_r}$$

$\square$

<u>Proof</u> of the Higher-order Cheeger's inequality:

For $\delta=\sqrt{\frac{1}{8r+1}}$ choose a partition $V=S_1\dot\cup\ldots\dot\cup S_m$ as in the previous lemma.

<u>Problem</u>: Could have many sets S_i with $\sum_{v\in S_i}\|F(v)\|^2$ small.

$\rightarrow$ Iteratively merge sets with $\sum_{v\in S_i}\|F(v)\|^2$ smallest such that

$$\frac{1}{4}\leq\sum_{v\in S}\|F(v)\|^2\leq\frac{1}{1-\delta^2}=1+\frac{1}{8r}\text{ for the new set } S.$$

If this is no longer possible:

$T_1,\ldots,T_s$ - new sets, ordered by size of $\sum_{v\in T_i}\|F(v)\|^2$

$$s':=\min\left\{i\in\{1,\ldots,s\}:\sum_{v\in T_i}\|F(v)\|^2\geq\frac{1}{4}\right\}$$

Then: $\displaystyle\sum_{i>s'}\sum_{v\in T_i}\|F(v)\|^2<\frac{1}{4}$

$$\text{So: } s'\left(1+\frac{1}{8r}\right)\geq\sum_{i=1}^{s'}\sum_{v\in T_i}\|F(v)\|^2>\underbrace{\sum_{v\in V}\|F(v)\|^2}_{=\sum_{i=1}^{r}\|f_i\|^2=r}-\frac{1}{4}=r-\frac{1}{4}.$$

$$\Longrightarrow s'\geq\frac{r-\frac{1}{4}}{1+\frac{1}{8r}}=r-\underbrace{\frac{3r}{8r+1}}_{<1}\Longrightarrow \underline{\underline{s'\geq r}}$$

So, have $\geq r$ sets T_i with

$$\frac{1}{4} \leq \sum_{v \in T_i} \|F(v)\|^2 \leq 1 + \frac{1}{8r}.$$

Now, show that for fixed i there is $\tau \in (0, M)$, $M = \max_{v \in V} \|F(v)\|^2$:

$$\frac{|\delta(\hat{T}_i)|}{|\hat{T}_i|} \leq \tilde{c} \cdot r^2 \cdot \sqrt{k(k-\lambda_r)} \qquad (*)$$

For τ chosen u.a.r. in $(0, M)$:

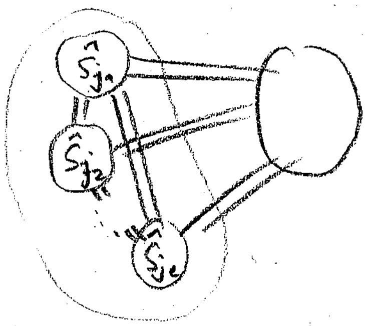
- $\mathbb{E}_\tau[|\hat{T}_i|] = \frac{1}{M} \sum_{v \in T_i} \|F(v)\|^2$

  $\Longrightarrow \frac{1}{M} \cdot \frac{1}{4} \leq \mathbb{E}_\tau[|\hat{T}_i|] \leq \frac{1}{M}\left(1 + \frac{1}{8r}\right)$

- Suppose $T_i = \bigcup_{j \in I} S_j$. Then $\hat{T}_i = \bigcup_{j \in I} \hat{S}_j$

  $$|\delta(\hat{T}_i)| \leq \sum_{j \in I} |\delta(\hat{S}_j)| \leq \sum_{j=1}^{m} |\delta(\hat{S}_j)|$$

$\hat{T}_i$



So: $\mathbb{E}_\tau[|\delta(\hat{T}_i)|] \leq c \cdot \frac{\sqrt{r}}{\delta} \sqrt{k(k-\lambda_r)} \cdot \underbrace{\mathbb{E}_\tau\left[\sum_{j=1}^{m} |\hat{S}_j|\right]}$

$$= \frac{1}{M} \cdot \sum_{j=1}^{m} \sum_{v \in S_j} \|F(v)\|^2$$

$$= \frac{1}{M} \cdot r$$

- Together: $\dfrac{\mathbb{E}_\tau[|\delta(\hat{T}_i)|]}{\mathbb{E}_\tau[|\hat{T}_i|]} \leq 4c \cdot \dfrac{r^{3/2}}{\delta} \sqrt{k(k-\lambda_r)} \underset{\underset{\delta = \sqrt{\frac{1}{8r+1}}}{\uparrow}}{=} \tilde{c} \cdot r^2 \sqrt{k(k-\lambda_r)}$

As in previous proof, this implies that there is $\tau \in (0, M)$ with $(*)$.

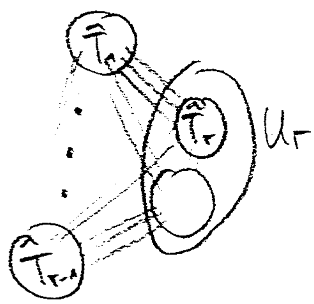
<u>Last problem:</u> $\hat{T}_1, \ldots, \hat{T}_r$ pairwise disjoint, but might not form a partition.

Reorder s.t. $|\hat{T}_1| \leq |\hat{T}_2| \leq \ldots \leq |\hat{T}_r|$

Let $U_i = \hat{T}_i$ for $i = 1, \ldots, r-1$ and $U_r = V \setminus \bigcup_{i=1}^{r-1} U_i$.

$$\frac{|\delta(U_r)|}{|U_r|} \leq \frac{\sum_{i=1}^{r} |\delta(\hat{T}_i)|}{|U_r|} \underset{|\hat{T}_i| \leq |U_r|,\ i=1,\ldots,r}{\leq} \sum_{i=1}^{r} \frac{|\delta(\hat{T}_i)|}{|\hat{T}_i|} \leq r \cdot \tilde{c} \cdot r^2 \sqrt{k(k-\lambda_r)}$$

$\square$



<u>Remark:</u>

- Higher order Cheeger's inequality from

  Lee, Gharan, Trevisan –

  "Multiway spectral partitioning and higher-order Cheeger inequalities"

- Randomized partitioning algorithm from

  Charikar, Chekuri, Goel, Guha, Plotkin –

  "Approximating a finite metric by a small number of tree metrics"