MaxCut on dense graphs

MaxCut Recall MaxCut defined by the following quadratic programming:

$$\text{Maximize} \sum_{(i,j)\in E} (x_i + x_j - x_i x_j)$$
$$\text{s.t.} \quad x_i \in \{0,1\} \quad \forall i \in V$$

Dense graph. $|E| \geq \frac{1}{100}|V|^2$ (100 is an arbitrary constant)

Goal of this lecture Via Sherali-Adams LP hierachy, a $(1-\varepsilon)$-approx. alg. for maxcut on dense graph in poly(n) time $\forall$ constant ε.

— a polynomial time approximation scheme (PTAS).

SA for MaxCut. Problem: $x_i x_j$ not linear.

Soln. → Introduce variables $p^{(i,j)}_{(0,0)}$, $p^{(i,j)}_{(0,1)}$, $p^{(i,j)}_{(1,0)}$, $p^{(i,j)}_{(1,1)}$ with constraints

a) $p^{(i,j)}_{(0,0)}, p^{(i,j)}_{(0,1)}, \dots \in [0,1]$, b) $\sum_{a,b} p^{(i,j)}_{(a,b)} = 1$ — to form a distrib.

→ $x_i x_j$ can be represented by $p^{(i,j)}_{(1,1)} = \mathbb{E}_{x_i,x_j\sim p}\, x_i x_j$

x_i can be represented by $p^{(i,j)}_{(1,0)} + p^{(i,j)}_{(1,1)} = \mathbb{E}_{x_i\sim p}\, x_i$

→ Consistency check: $p^{(i)}_a = p^{(i,j)}_{(a,0)} + p^{(i,j)}_{(a,1)} \quad \forall j$.

In general, extend p to sets of distributions on every set of ℓ variables (level-ℓ)

→ $p^S_{\sigma: S\to\{0,1\}} \in [0,1] \quad |S| \leq \ell$

→ $\sum_\sigma p^S_\sigma = 1 \quad \forall S$.

→ consistency check: $p^{S\cup\{i\}}_{\sigma\cup\{x_i\to 0\}} + p^{S\cup\{i\}}_{\sigma\cup\{x_i\to 1\}} = p^S_\sigma \quad \forall |S| < \ell$

objective: Maximize $\sum_{(i,j)\in E} \mathbb{E}_{(x_i,x_j)\sim p} (x_i + x_j - x_i x_j)$.

Observation need $n^{O(\ell)}$ time to solve level-ℓ SA.

Let's focus on level-2 first. Doesn't work well — always set $p^{(i)}_{(0)} = p^{(i)}_{(1)} = 1/2$

$\forall \{i,j\}$ set the distribution as

$$\begin{array}{c|cc} j\downarrow \ \backslash\ i\rightarrow & 0 & 1 \\ \hline 0 & 0 & 1/2 \\ 1 & 1/2 & 0 \end{array}$$

*) passes the consistency check

*) every edge (i,j) is on the cut with prob. 1.

<u>Intuition on why this fails.</u> x_i & x_j are super correlated: always have $x_i = 1 - x_j$

or. $\mathbb{E}_p x_i x_j = 0$ $\quad (\mathbb{E}_p x_i)(\mathbb{E}_p x_j) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ — very different.

If x_i & x_j are less correlated, maybe we can do better?

<u>Question.</u> How to measure correlation?

<u>Soln.</u> Here we use mutual information $I(x_i; x_j)$ — for easier analysis later.

High mutual information $\Leftrightarrow$ High correlation.

<u>Claim 1</u> $\quad \left| \mathbb{E} x_i x_j - (\mathbb{E} x_i)(\mathbb{E} x_j) \right| \leq \sqrt{2 I(x_i; x_j)}$

<u>Proof.</u> $\quad \text{LHS} \leq \underbrace{\Delta}_{\text{total variation distance}}((x_i, x_j),\ x_i \times x_j) \leq \sqrt{2}\, \underbrace{h}_{\text{Hellinger distance}}((x_i, x_j),\ x_i \times x_j)$

$$\leq \sqrt{2} \cdot \sqrt{\underbrace{KL}_{\text{KL divergence.}}((x_i, x_j) \| x_i \times x_j)} = \sqrt{2} \cdot \sqrt{I(x_i; x_j)}$$

<u>Def.</u> Call an LP (SA) solution α-independent if $\operatorname{avg}_{i,j} I(x_i; x_j) \leq \alpha$

<u>Claim 2.</u> Let Round be the value of independent rounding (set each x_i to 1 w.p. $p^{(i)}_{(1)}$ independently). ~~On a dense graph, we have~~ Given an α-indep soln, we have

$$\mathbb{E}\,\text{Round} \geq LP - n^2 \cdot \sqrt{2\alpha}$$

<u>Proof</u>

$$\mathbb{E}\,\text{Round} = \sum_{(i,j) \in E} \left(\mathbb{E}(x_i + x_j) - \cancel{x_i x_j}\, (\mathbb{E} x_i)(\mathbb{E} x_j) \right)$$

$$\geq \sum_{(i,j) \in E} \left[\mathbb{E}(x_i + x_j - x_i x_j) - \left| \mathbb{E} x_i x_j - (\mathbb{E} x_i)(\mathbb{E} x_j) \right| \right]$$

$$= LP - \sum_{(i,j) \in E} \left| \mathbb{E} x_i x_j - (\mathbb{E} x_i)(\mathbb{E} x_j) \right|$$

$$\geq LP - \sum_{(i,j) \in E} \sqrt{2 I(x_i; x_j)} \qquad \text{(Claim 1)}$$

$$\geq LP - \sqrt{|E|} \cdot \sqrt{\sum_{(i,j) \in E} 2 I(x_i; x_j)} \qquad \text{(Cauchy-Schwartz)}$$

$$\geq LP - n \cdot \sqrt{\sum_{i,j} 2 I(x_i; x_j)}$$

$$\geq LP - n \cdot \sqrt{n^2 \cdot \operatorname{avg}_{i,j} 2 I(x_i; x_j)} \geq LP - n^2 \cdot \sqrt{2\alpha}$$

<u>Corollary 3.</u> It's a $(1 - 200 \cdot \sqrt{2\alpha})$-approx. since $LP \geq \frac{|E|}{2} \geq \frac{n^2}{200}$ on dense graphs.

What to do if the SA soln ~~has~~ high average correlation/mutual information?

Idea Fix and condition on a (randomly chosen) variable, high mutual information

$\Rightarrow$ large reduction on the entropy of remaining variables.

Recall $H(x_i) - \underbrace{H(x_i|x_j)}_{\mathbb{E}_{b_j \sim p_j} H(x_i|x_j=b_j)} = I(x_i;x_j)$

Take average over all i,j:

$$\operatorname{avg}_i H(x_i) - \operatorname{avg}_j \mathbb{E}_{b_j\sim p_j} \operatorname{avg}_i H(x_i|x_j=b_j) = \operatorname{avg}_{i,j} I(x_i;x_j) \qquad (*)$$

LHS: drop (expected) of the average entropy if we randomly choose and fix a random variable, and condition on it.

The conditioning procedure.

Input: a level-ℓ SA solution $\{p_\sigma^S : |S| \le \ell, \sigma: S \to \{0,1\}\}$

Output: a level-$(\ell-1)$ SA solution.

How: pick a random r.v. x_u, sample $b_u \sim p^{(u)}$, let $\{q_\sigma^S\}$ ($\to$ output.) be the probability distrib. obtained by conditioning on $\{p_\sigma^{S\cup\{x_u\}} | x_u = b_u\}$

Remark. Easy to check $\{q_\sigma^S : |S| \le \ell-1\}$ satisfies the consistency requirement.

Let's rewrite $(*)$: $\underbrace{\operatorname{avg}_i H(x_i)}_{\text{average entropy}} - \underbrace{\mathbb{E}_j \mathbb{E}_{b_j\sim p_j} \operatorname{avg}_i H(x_i|x_j=b_j)}_{\text{expected average entropy after one-step conditioning}} = \operatorname{avg}_{i,j} I(x_i;x_j) \quad (**)$

Condition both sides on a random x_k:

$$\underbrace{\mathbb{E}_k \mathbb{E}_{b_k\sim p_k} \operatorname{avg}_i H(x_i|x_k=b_k)}_{\text{expected avg. entropy after 1-step conditioning}} - \underbrace{\mathbb{E}_{j,k} \mathbb{E}_{b_j,b_k} \operatorname{avg}_i H(x_i|x_j=b_j, x_k=b_k)}_{\text{expected avg. entropy after two-step conditioning}} = \underbrace{\mathbb{E}_k \mathbb{E}_{b_k\sim p_k} \operatorname{avg}_{i,j} I(x_i;x_j|x_k=b_k)}_{\text{expected avg. M.I. after 1-step conditioning}}$$

In general:

$$\mathbb{E}_{k_1,\dots,k_{t-1}} \mathbb{E}_{b_1,\dots,b_{t-1}} \operatorname{avg}_i H(x_i|x_{k_1},\dots,x_{k_{t-1}}) - \mathbb{E}_{k_1,\dots,k_t} \mathbb{E}_{b_1,\dots,b_t} \operatorname{avg}_i H(x_i|x_{k_1},\dots,x_{k_t})$$
$$= \mathbb{E}_{k_1,\dots,k_{t-1}} \mathbb{E}_{b_1,\dots,b_{t-1}} \operatorname{avg}_{i,j} I(x_i;x_j|x_{k_1},\dots,x_{k_{t-1}}) \qquad (***)$$

Remark Expected entropy drop at t-th step conditioning is the expected average M.I. after $(t-1)$ steps of conditioning.

Plan. Keep conditioning on a randomly chosen variable until avg. MI is small (becomes α-independent solution). This process can't last for long as the avg. entropy has to stay non-negative.

Claim 4 There exists $0 \le t < T$, s.t. after t-step conditioning, the expected avg. MI. $\le \frac{1}{T}$.

Proof Add (***) up for $t = 1, 2, \cdots, T$.

LHS forms a telescoping sum, $= \operatorname{avg}_i H(x_i) - \mathop{\mathbb{E}}_{k_1, \cdots, k_{T-1}} \mathop{\mathbb{E}}_{b_1, \cdots, b_{T-1}} \operatorname{avg}_i H(x_i \mid x_{k_1}, \cdots, x_{k_{T-1}})$

$\le \operatorname{avg}_i H(x_i) \le 1$

$$\text{RHS} = \sum_{t=1}^{T} \mathop{\mathbb{E}}_{k_1, \cdots, k_{t-1}} \mathop{\mathbb{E}}_{b_1, \cdots, b_{t-1}} \operatorname{avg}_{i,j} I(x_i ; x_j \mid x_{k_1}, \cdots, x_{k_{t-1}}) = \text{LHS} \le 1$$

Therefore, $\exists\, t \le T : \mathop{\mathbb{E}}_{k_1, \cdots, k_{t-1}} \mathop{\mathbb{E}}_{b_1, \cdots, b_{t-1}} \operatorname{avg}_{i,j} I(x_i ; x_j \mid x_{k_1}, \cdots, x_{k_{t-1}}) \le 1/T$

Remark Using at most $(T-1)$ steps of conditioning, can obtain $1/T$-independent solution. In order to perform $(T-1)$ steps of conditioning, need to start with level-$(T+1)$ SA soln.

What about objective value (cut size)?

Claim 5. Expected obj. value remains the same after 1-step conditioning.

Proof.
$$\mathop{\mathbb{E}}_i \mathop{\mathbb{E}}_{b_i} [\text{obj} \mid x_i = b_i] = \mathop{\mathbb{E}}_i \Big[\Pr[x_i = 1] \cdot \mathbb{E}[\text{obj} \mid x_i = 1] + \Pr[x_i = 0] \cdot \mathbb{E}[\text{obj} \mid x_i = 0] \Big]$$
$$= \mathop{\mathbb{E}}_i \mathbb{E}[\text{obj}] = \mathbb{E}[\text{obj}]$$

Corollary 6 Expected obj. value remains the same after any steps of conditioning.

Put everything together.

Alg. 1.) Solve $T = 2C^2/\varepsilon^2$-level SA for maxcut.

2.) Choose $t < T$ so that after t-step conditioning, the expected M.I. $\le 1/T$

3.) Perform t steps of conditioning to the solution.

4.) Perform independent rounding (i.e. round each x_i to 1 w.p. $p_1^{(i)}$) to the soln. obtained by conditioning.

Remark. Expect $\frac{1}{T}$-indep. soln. after 3), and expect obj. value $= LP$. Via claim 2, Expect Rounding value $\ge LP - n^2 \cancel{\times} \sqrt{2/T} \ge OPT - O(\varepsilon n^2)$

Theorem 7. Expected integral soln. $\geq LP - n^2 \cdot \sqrt{2/T} \geq OPT - \varepsilon n^2/c$

Proof. Let μ — r.v. denote the obj. value of soln. after t-step conditioning.

α — r.v. denote the avg. M.I. of soln. after t-step conditioning.

We have: $\mu = LP$ (Corollary 6) ; $\alpha \leq \frac{1}{T}$ (Claim 4 & step 2))

By Claim 2. $\mathbb{E}[\text{Indep. Rounding Value}] \geq \mathbb{E}[\mu - n^2 \cdot \sqrt{2\alpha}]$

$$= \mathbb{E}\mu - n^2 \cdot \mathbb{E}\sqrt{2\alpha}$$

$$\geq \mathbb{E}\mu - n^2 \sqrt{\mathbb{E}\, 2\alpha}$$

$$\geq LP - n^2 \cdot \sqrt{2/T}$$

Remark. choose $c = 200$, and Expected integral soln. $\geq OPT(1-\varepsilon)$ (since $OPT \geq \frac{n^2}{200}$ on dense graphs). Need $O(1/\varepsilon^2)$-level SA. The algorithm runs in time $n^{O(1/\varepsilon^2)}$.