Techniques in proving the hardness of approximation: Reductions from label cover problem

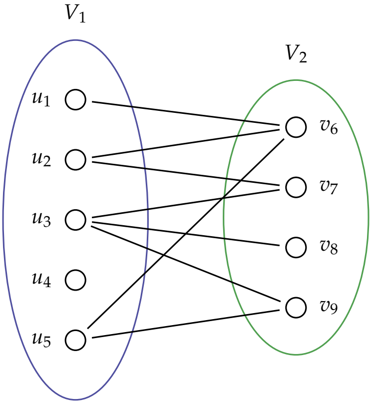
Theja Tulabandhula
Massachusetts Institute of Technology

[Shmoys and Williamson 2011: Sec 16.4]

November 18, 2011

Given

- Bipartite graph (V_1, V_2, E) and label sets L_1 and L_2.
- Every edge (u, v) has set of acceptable label-pairs $\subset L_1 \times L_2$.
- Edge (u, v) is satisfied if it is assigned an acceptable label-pair.



Edge	label-pairs
(u_1, v_6)	{(blue,red), (green,blue), (black,black)}
(u_2, v_6)	{(green,blue)}
⋮	⋮
(u_5, v_9)	...

PROBLEMS

Maximization label cover (MaxLabelCover)

- Assign exactly one label to each vertex $w \in V_1 \cup V_2$.
- Assign labels so that many edges are satisfied.

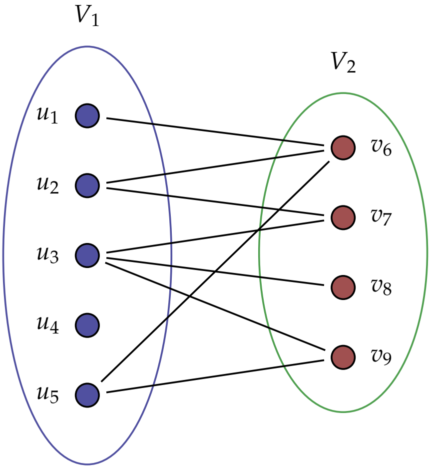
Minimization label cover (MinLabelCover)

- Assign a set $\mathcal{L}_w$ of labels for each vertex $w \in V_1 \cup V_2$.
- For each edge (u, v), $\mathcal{L}_u \times \mathcal{L}_v$ should contain an acceptable label pair.
- Minimize the total number of labels used: $\sum_{u \in V_1} |\mathcal{L}_u| + \sum_{v \in V_2} |\mathcal{L}_v|$.

What we will cover in this talk

1. MaxLabelCover is hard to approximate.
2. Unweighted Set Cover problem is hard to approximate.
3. (Very) strong hardness of MaxLabelCover.

EXAMPLE OF MAXLABELCOVER

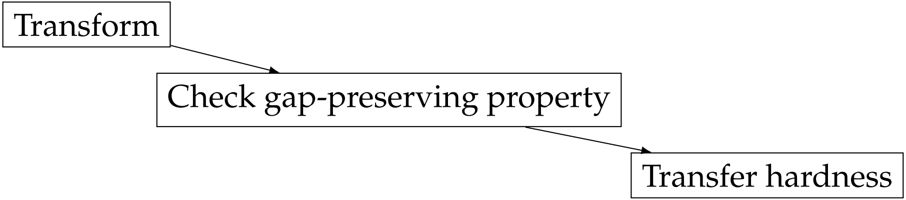


- If edge (u_1, v_6) has only (blue,green) and all edges including (u_1, v_6) and (u_5, v_6) have only (blue,red) as acceptable label-pairs then value of solution above is $|E| - 1$.

GAP-PRESERVING REDUCTION FROM MAX E3SAT

We will show this in three steps:

- Come up with a transformation.
- Show that it is gap preserving.
  - Roughly, preserves “gap” in fraction of constraints/clauses satisfied to distinguish between “YES”/“NO” instances.
- Transfer the hardness of E3SAT to MaxLabelCover.



TRANSFORMATION

- Given an instance I of MAX E3SAT with m clauses.
- Create I' with:
  - vertex $u_i \in V_1$ for each variable x_i and
  - vertex $v_j \in V_2$ for each clause C_j.
  - edge (u_i, v_j) if literal x_i appears in clause C_j.
- $L_1 = \{T, F\}$
  $L_2 = L_1 \times L_1 \times L_1$.

Intuition:

- $u_i \in V_1$ with label b_i relates to variable x_i.
- $v_j \in V_2$ with label (b_p, b_q, b_r)relates to clause C_j.

If acceptable label-pairs are all $(b_i, (b_p, b_q, b_r)) \in L_1 \times L_2$ such that C_j is satisfied, then b_i is same as b_p or b_q or b_r.

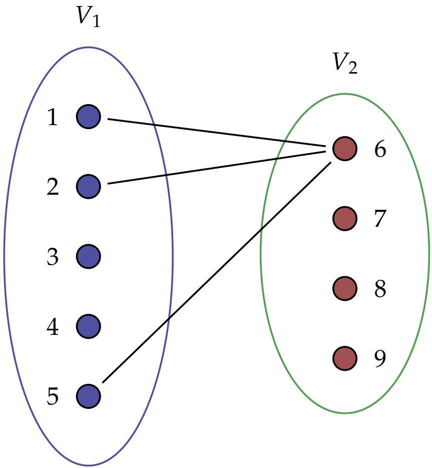
EXAMPLE

Let $x_1 \vee \bar{x}_2 \vee \bar{x}_5$ be the clause. Let vertex $v_6 \in V_2$ represent it. Then, $(u_1, v_6), (u_2, v_6), (u_5, v_6) \in E$.

(u_1, v_6) label-pairs:

$\{x_1, (x_1, x_2, x_5)\}$

$\{\ \{T, (T, T, T)\},$
$\{T, (T, T, F)\},$
$\{T, (T, F, T)\},$
$\{T, (T, F, F)\},$
$\{F, (F, T, F)\},$
$\{F, (F, F, T)\},$
$\{F, (F, F, F)\}\}$



Given label for v_j and (u_i, v_j), there is at most one acceptable label for u_i.

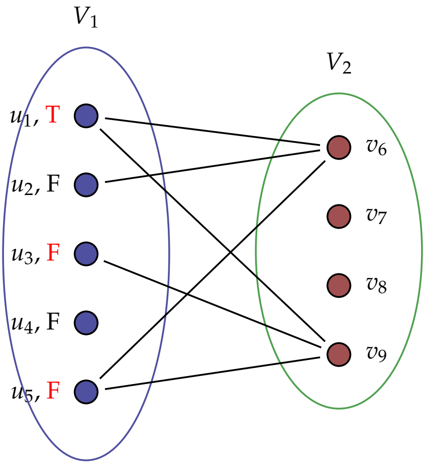
CHECK GAP-PRESERVING PROPERTY

Claim: Given an assignment of $\{x_i\}$ satisfying k clauses for I, can construct a solution to I' satisfying $3k + 2(m-k)$ edges.

Proof:

1. label $u_i \in V_1$ with that of x_i.
2. label $v_j \in V_2$ with that corresponding to (x_p, x_q, x_r).
3. For each satisfied clause, all 3 edges are satisfied.
4. For each unsatisfied clause, flip values to satisfy 2 of the 3 edges.

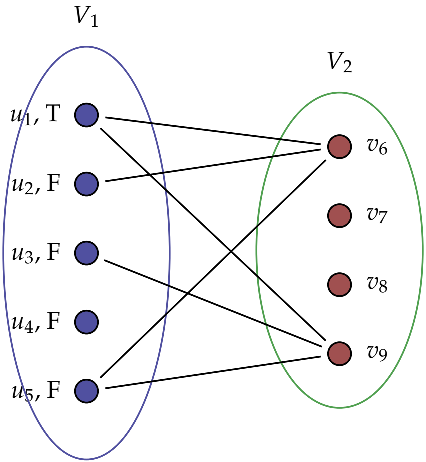
Example: If $x_1 \vee \bar{x}_2 \vee \bar{x}_5$ corresponds to v_6, then 3 satisfied edges.
If $\bar{x}_1 \vee x_3 \vee x_5$ corresponds to v_9: then only 2 satisfied edges.



Claim: Given a solution to I' with $3k + 2(m - k)$ edges satisfied, can obtain an assignment to E3SAT with k clauses satisfied.

Proof:

- Assign x_i in E3SAT based on label of $u_i \in V_1$.
- Each $v_j \in V_2$ with 3 satisfied edges corresponds to a satisfied clause.



TRANSFER HARDNESS OF APPROXIMATION FOR E3SAT

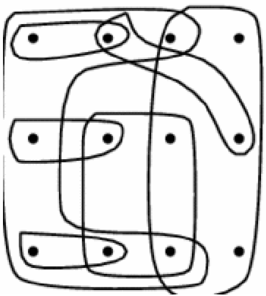
- Know: NP-hard to distinguish E3SAT where all m clauses are satisfiable and E3SAT where at most $m(\frac{7}{8}+\delta)$ clauses are satisfiable.
- If m clauses satisfiable
  - then all $3m$ edges are satisfiable.
- If $m(\frac{7}{8}+\delta)$ clauses satisfiable
  - then all $3m(\frac{7}{8}+\delta)+2m(\frac{1}{8}-\delta)=(\frac{23}{8}+\delta)m$ edges are satisfiable.
- Distinguishing all vs $\frac{23}{24}+\frac{\delta}{3}$ fraction of the $3m$ edges is NP-hard.

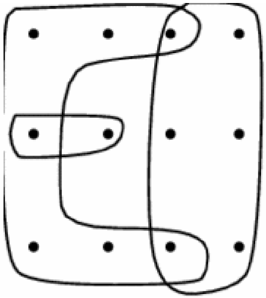
Theorem

There is no $\alpha > \frac{23}{24}$ approximation to MaxLabelCover unless P = NP.

UNWEIGHTED SET COVER

- Given ground set $A \times B$ with $|A \times B| = N$ and $S_{11}, ..., S_{KJ} \subseteq A \times B$.
- Cover C is a collection of some S_{kj} such that $\cup_{S_{kj} \in C} S_{kj} = A \times B$.
- Minimize $|C|$.





Start with two things

1 Hardness of MaxLabelCover[1]:

There exist d-regular I of MaxLabelCover such that distinguishing
- I whose all edges are satisfiable
- I whose at most $1/\log^2(|L_1||E|)$ fraction edges are satisfiable

is not possible unless NP algorithms run in time $O(n^{O(\log\log n)})$.

2 Property:

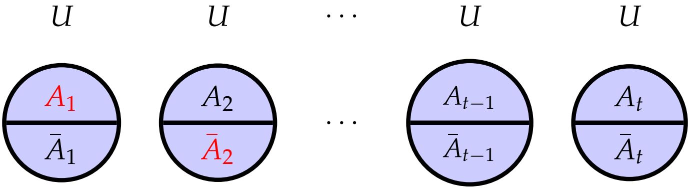
For these hard instances, given a label for $v \in V_2$ and edge (u, v), there is at most one acceptable label for $u \in V_1$.

$\boxed{\text{Transform}} \rightarrow \boxed{\text{Check gap-preserving property}} \rightarrow \boxed{\text{Transfer hardness}}$

[1] Assume this version for now. n is the input size of instance

NEED NOTION OF A PARTITION SYSTEM

- Consider tuple $\{U, t, h\}$ with set U and scalar t, h.
  - Pairs of sets $\{A_i, \bar{A}_i\}_{i=1}^{t}$ with $\bar{A}_i = U - A_i$.
  - Pick any h pairs, and one set from each pair. Call it B_{i_j}.
  - Property: $\cup_{j=1,\ldots,h} B_{i_j} \subsetneq U$.



- Example:
  - Let $h = 2$.
  - Pick any 2 set pairs out of t above.
  - Say, $A_1, \bar{A}_2$. Then, $A_1 \cup \bar{A}_2 \subsetneq U$

TRANSFORMATION

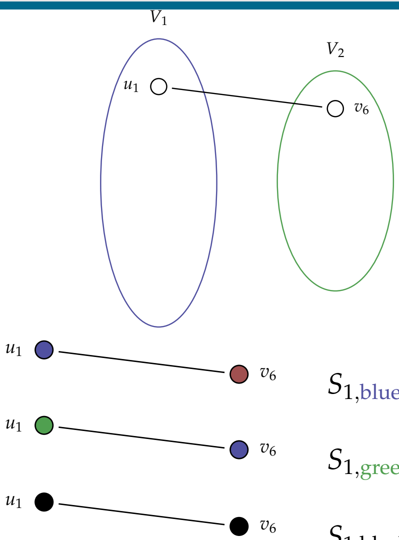
Given a label cover instance I:

- Create set $S_{u,i}$ for every $u \in V_1, i \in L_1$.
- Create set $S_{v,j}$ for every $v \in V_2, j \in L_2$.
- Ground set $E \times U$ for some U with $N = |E \times U|$.

Exactly what we need for set cover instance: a ground set of size N and a collection of sets.

Intuition: Create sets such that

- if there is a satisfiable label-pair (i, j) for edge (u, v) in MaxLabelCover, then for some k
  - $\{(u, v)\} \times A_k \subset S_{u,i}$ and
  - $\{(u, v)\} \times \bar{A}_k \subset S_{v,j}$
  - $\Rightarrow \{(u, v)\} \times U \subset S_{u,i} \cup S_{v,j}$.
  - two sets contain $\{(u, v)\} \times U$
- else take atleast $h = \Omega(\log N)$ sets to contain $\{(u, v)\} \times U$.



$S_{1,\text{blue}}, S_{6,\text{red}}$

$S_{1,\text{green}}, S_{6,\text{blue}}$

$S_{1,\text{black}}, S_{6,\text{black}}$

- Let $L_1 = L_2 = \{$blue,red,green,black$\}$.
- Acceptable label-pairs for (u_1, v_6): $\{$(blue,red),(green,blue),(black,black)$\}$

TRANSFORMATION

Construct sets using a partition system

- Let $\{U, t, h\}$ be the partition system with $t = |L_1|$ and $h = \log |L_1||E|$.
- $S_{u,i} := \{((u, v), a) : (u, v) \in E, a \in A_i\}$
- $S_{v,j} := \{((u, v), b) : (u, v) \in E, b \in \bar{A}_i$ where i such that (i, j) acceptable label-pair[2]$\}$.

$S_{1,\text{blue}}, S_{6,\text{red}}$

- u_1 has 3 label-pairs.
- We have $\{A_1, \bar{A}_1, \ldots, A_4, \bar{A}_4\}$.
- $S_{1,\text{blue}} = \{(u_1, v_6)\} \times A_1$.
- $S_{6,\text{red}} = \{(u_1, v_6)\} \times \bar{A}_1 \cup \{(u_3, v_6)\} \times \bar{A}_\alpha \cup \{(u_5, v_6)\} \times \bar{A}_\beta$

[2]Property: given label j and edge (u, v), there is at most one acceptable label i.

CHECK GAP-PRESERVING PROPERTY

- Claim: Given a solution to label cover with all edges satisfied, there is a solution to set cover using only $|V_1| + |V_2|$ sets.
  Proof:
  - If $u \in V_1$ labeled with label i, choose set $S_{u,i}$. Similarly $S_{v,j}$.
  - Since, label-pair (i, j) is acceptable, $\{u, v\} \times U = \{(u, v)\} \times (A_i \cup \bar{A}_i) \subseteq S_{u,i} \cup S_{v,j}$.
  - All edges satisfied implies ground set is covered.

- Claim: Given a solution to set cover with at most $\frac{h}{8}(|V_1| + |V_2|)$ sets, we can find solution to a label cover instance satisfying at least $\frac{2}{h^2}|E|$ edges.
  Proof:
  - For $u \in V_1, n_u$ is the number of sets $S_{u,i}$ in set cover. Similarly n_v.
  - Let E_1 be set of edges with $n_u \geq \frac{h}{2}$. Similarly E_2.
  - There are at most $1/4$ vertices of $V_1 \cup V_2$ with n_u or $n_v \geq h/2$.
  - d-regular label cover implies $|E_1 \cup E_2| \leq \frac{1}{2}|E|$.
  - Let $E_0 = E - E_1 - E_2$.
  - Any edge in E_0 has $n_u + n_v < h$, so $\#\{S_{u,i}, S_{v,j}\}$ in set cover $< h$.

Proof (continued):

- By property of partition system, for $\{(u, v)\} \times U$ to be covered, for every (u, v) $\exists i, j$ such that $S_{v,j} \supseteq \{(u, v)\} \times \bar{A}_i$.
- Suppose, pick i randomly for each u and j for each v.
- For each $(u, v) \in E_0$, there is at least one label-pair out of $(h/2)^2$ that satisfies the edge. $\Pr[\mathit{satisfied}] \geq (2/h)^2$.
- Expected number of edges satisfied is at least

$$\sum_{(u,v)\in E_0} \frac{4}{h^2} = \frac{4}{h^2}|E_0| \geq \frac{2}{h^2}|E|.$$

TRANSFER HARDNESS OF APPROXIMATION FOR MAXLABELCOVER

- Claim: There is no $\frac{1}{32}\log N$ approximation for unweighted set cover[3] unless NP algorithms run in $O(n^{O(\log\log n)})$ time.
  Proof:
  - Given $(V_1, V_2, E), L_1, L_2$, acceptable label-pairs.
  - $h = \log |L_1||E|$ and $t = |L_1|$. $|U| = 2^{2h+2}t^2 = 4|E|^2|L_1|^4$.
  - $N = |E||U| = 4|E|^3|L_1|^4 \leq 2^{4h}$
  - If all edges satisfiable, set cover has a solution with $|V_1| + |V_2|$.

[3] N is the size of ground set

Proof (continued):

- If we obtain a cover of size $\leq \frac{h}{8}(|V_1| + |V_2|)$, can get solution to MaxLabelCover with at least $2/h^2$ fraction of edges satisfied.
- Allows us to distinguish all vs at most $1/h^2$ fraction.
- If we have $\alpha = h/8$ approx. alg for set cover, then distinguish such instances in P time.
- $h/8 \geq \frac{1}{32} \log N$.
- Thus there is no approx. alg with $\alpha = \frac{1}{32} \log N$ for unweighted set cover unless NP algorithms run in time $O(n^{O(\log \log n)})$.

Q.E.D

CONSTRUCTION OF THE PARTITION SYSTEM

- Claim: Given h, t there is a randomized algorithm for constructing partition of size $s = 2^{2h+2}t^2$ with high probability.
  Proof:
  - Pick A_i uniformly at random from 2^U.
  - Suppose, we selected $i_1 < \ldots < i_h$ and B_{i_j} at random.
  - Number of ways: $\binom{t}{h}2^h$.
  - $\Pr[\cup_{j=1}^h B_{i_j} = U | \{B_{i_j}\}] = \left(1 - \frac{1}{2^h}\right)^s$. Because,
    - $\Pr[u \notin \cup_{j=1}^h B_{i_j}] = \Pi_{j=1}^h \Pr[u \notin \text{the set we choose}] = \frac{1}{2^h}$.
  - $\Pr[\{A_i\} \text{ do not form a partition}] \leq \binom{t}{h}2^h\left(1 - \frac{1}{2^h}\right)^s \leq \frac{1}{2^h}$
  - Substitute $h = \log |L_1||E_1|$ and we get: the sets A_i form the desired partition with probability $1 - \frac{1}{|L_1||E|}$.

STRONGER HARDNESS: REDUCING THE PROBLEM TO ITSELF.

- Transformation
  - Given $I = \{(V_1, V_2, E), L_1, L_2, \text{accept. label-pairs}\}$
  - Create I': $V_1' = V_1 \times \cdots V_1$ (k times). Similarly V_2', E', L_1', L_2'.
  - Edge $(u, v) \in E'$ with component pairs being edges of I is satisfiable iff component-wise satisfiable.
- Gap-preserving reduction
  - Claim: if in I, given an edge and label for its vertex in V_2, there is at most one label for the other end in V_1, then similar holds for I'.
  - Claim: d-regular implies d^k-regular and size is poly.
  - Claim: If $|E|(1-\delta)$ satisfied in I, then $\leq |E'|(1-\delta)^{ck}$ satisfied in I'.
- Hardness
  - There is a c such that for any k, approximation guarantee better than $(1-\delta)^{ck}$ for some δ is impossible unless NP algorithms run in $O(n^k)$.
  - No constant factor approximation exists.

WEAKENING THE HYPOTHESIS $P \neq NP$

- Quasipolynomial time: Running time $O(n^{O(\log^c n)})$ for constant c.
- Another Hardness result for MaxLabelCover
  - For any $\epsilon > 0$, there is no $2^{-log^{1-\epsilon} m}$ approximation[4] with d-regular instances unless NP has quasipolynomial time algorithms.
- Leads to the version we used for Unweighted Set Cover:
  - There are d-regular instances such that in P time, we cannot distinguish between instances with all edges satisfiable and those with at most $1/\log^2(|L_1||E|)$ fraction edges satisfiable unless NP has each problem running in $O(n^{O(\log \log n)})$

[4] m = $|E|$, number of edges in the instance.

SUMMARY

- Thus we have seen: reductions from NP-C problems, reductions that preserve approximations, PCP based reductions for constraint satisfaction problems, and in

  this talk: reductions from maximization label cover problem.
- In particular,
  1. MaxLabelCover is hard to approximate (reduced from MAX E3SAT).
  2. Unweighted Set Cover is hard to approximate (reduced from hard d-regular MaxLabelCover).
  3. (Very) strong hardness of MaxLabelCover (similar to Max. Independent Set problem).
- We did not see: MinlabelCover and its uses in hardness of approximation.

HARDNESS OF APPROXIMATING $(5,3)$-REGULAR MAXIMIZATION LABEL COVER

- Consider a restricted version of E3SAT with every variable appearing in exactly 5 clauses.
  - Claim: Still NP-hard to distinguish full satisfiability from fractional.
- Transformation
  - Each vertex $u_i \in V_1$ corresponding to x_i has 5 edges.
  - Each $v_j \in V_2$ has degree 3.
- Gap-preserving reduction
  - Same as original MAX E3SAT $\rightarrow$ MaxLabelCover.
  - Property still holds: for label b for $v \in V_2$, there is at most one acceptable label a for $u \in V_1$.
- Hardness
  - NP-hard to distinguish full satisfiability vs α satisfiability for some $\alpha < 1$.

HARDNESS OF APPROXIMATING 15-REGULAR MAXIMIZATION LABEL COVER

- Transformation
  - Given $(5,3)$-regular $I = \{(V_1, V_2, E), L_1, L_2, \text{accept. label-pairs}\}$, create $I' = \{(V_1', V_2', E'), L_1', L_2', \text{accept. label-pairs}\}$.
  - $V_1' = V_1 \times V_2; V_2' = V_2 \times V_1; L_1' = L_1; L_2' = L_2$
  - Have edge between $(u, v) \in V_1'$ and $(v', u') \in V_2'$ iff (u, v') and (u', v) are edges in I.
  - This edge satisfied if same labels satisfy edge (u, v') of I.
- Gap-preserving reduction
  - For any (u', v): each $(u, v') \in E$ corresponds to an edge in I′.
  - Claim: if edges of I are satisfied, then related edges of I' are also satisfied.
  - Claim: I' is 15-regular.
  - Claim: Given an edge and a label for vertex in V_2', there is at most one label for the other end of the edge in I' if this is true for I
- Hardness of α-approximation follows.

APPROXIMABILITY OF MINIMIZATION LABEL COVER

Theorem

- A $2^{\log^{1-\epsilon} m}$-approx. for d-regular MinLabelCover implies a $2^{-\log^{1-\epsilon'} m}$-approx. for d-regular instances of MaxLabelCover.