Ulpana · Tirgul 1

Hitting Sets

Setting: universe of n elements $V = \{v_1, \dots, v_n\}$

collection of ℓ sets $S_1, \dots, S_\ell \subseteq V$

sets have $\geq s$ elements: $\forall j \ |S_j| \geq s$

Example: V - vertices in the graph

$S_1, \dots, S_\ell$ - each set consists of nbrs of vertices with deg $\geq s$

Goal: find small hitting set $A \subseteq V$:

$$\forall j \quad A \cap S_j \neq \emptyset$$

Remarks: can always find hitting set of size $\min\{n, \ell\}$. not small

finding the smallest hitting set is NP-hard

we'll see how to find "reasonably" small one efficiently

very (very!) useful tool in algorithms

(examples today/later in course)

Thm: There is always a hitting set A of size $O(\frac{n}{s} \log \ell)$

Further, can be found in poly-time

Deterministic construction

natural greedy alg.

- at step i, add $a_i \in V$ that hits the most yet-unhit sets
- repeat until all sets are hit, say after step t
- output $A = \{a_1, \dots, a_t\}$

Need to show: $t \le \frac{n}{s} \ln(\ell)$

Notation: ℓ_i - #unhit sets after step i

$c_i(v)$ - #unhit sets before step i that contain v

"the counter of v before step i"

Obs 1: $\ell_{i-1} - \ell_i = c_i(a_i)$

why? step i decreases #unhit sets by $c_i(a_i)$

Obs 2: $\sum_{v \neq a_1, \dots, a_{i-1}} c_i(v) \ge \ell_{i-1} \cdot s$

why? each unhit set before step i adds 1 to $\ge s$ counters

Obs 3: $c_i(a_i) \ge \frac{1}{n-i+1} \sum_{v \neq a_1, \dots, a_{i-1}} c_i(v)$

why? max $\ge$ avg

combaine obs 1,2,3: $\ell_{i-1} - \ell_i \geq \frac{\ell_{i-1} \cdot s}{n-i+1}$

rearrange: $\ell_i \leq \left(1 - \frac{s}{n-i+1}\right) \ell_{i-1}$

induction: $\ell_i \leq \prod_{k=1}^{i} \left(1 - \frac{s}{n-k+1}\right) \underbrace{\ell_0}_{=\ell}$

$\leq \left(1 - \frac{s}{n}\right)^i \ell$

$1-x < e^{-x} \rightsquigarrow \; < e^{-\frac{is}{n}} \ell$

finalize: if $i \geq \frac{n}{s} \ln(\ell)$ we get $\ell_i < 1 \Rightarrow \ell_i = 0$

because $\ell_{t-1} \geq 1$, it must be that $t-1 \leq \frac{n}{s} \ln(\ell)$

$\blacksquare$

<u>Randomized constructions</u>

Cons: small prob. of error

pros: fast! $O(n)$ time

oblivious: doesn't need to know the sets

(only how many sets are there)

Assume $l \leq n^{100}$ (100 - arbitrary const), want size $O\left(\frac{n}{s} \log n\right)$

Assume $s \gg \ln(n)$ (otherwise can take all elements)

<u>Method 1</u> sample each $v \in V$ into A ind. w.p. $p = \frac{1}{s} \cdot 110 \ln(n)$

prob. to miss specific S_j is $(1-p)^{|S_j|} < e^{-ps} \leq e^{-110 \ln(n)} = \frac{1}{n^{110}}$

by union bound, prob to be hitting set is $\geq 1 - \frac{1}{n^{10}}$

<u>expected</u> size is $np = O\left(\frac{n}{s} \ln n\right)$

Can prove that this is also the size w.h.p (Chernoff)

<u>Method 2</u> sample $A \subseteq V$ uniformly at random from the subsets of V with size $\lceil np \rceil$

prob to miss specific S_j is

$$\frac{n-|S_j|}{n} \cdot \frac{n-|S_j|-1}{n-1} \cdot \frac{n-|S_j|-2}{n-2} \cdots = \prod_{i=1}^{\lceil np \rceil} \frac{n-|S_j|-(i-1)}{n-(i-1)}$$

$$\leq \left(1 - \frac{|S_j|}{n}\right)^{\lceil np \rceil} < e^{-\frac{s}{n} np} = e^{-ps} = \frac{1}{n^{110}}$$

by union bound, get hitting set w.p. $\geq 1 - \frac{1}{n^{10}}$

Baswana-Sen Spanners

Recall: a t-spanner of $G=(V,E)$ is subgraph H s.t.

$$\forall u,v \in V \quad dist_H(u,v) \leq t \cdot dist_G(u,v)$$

Lecture: greedy spanner

Now: different construction using hitting sets

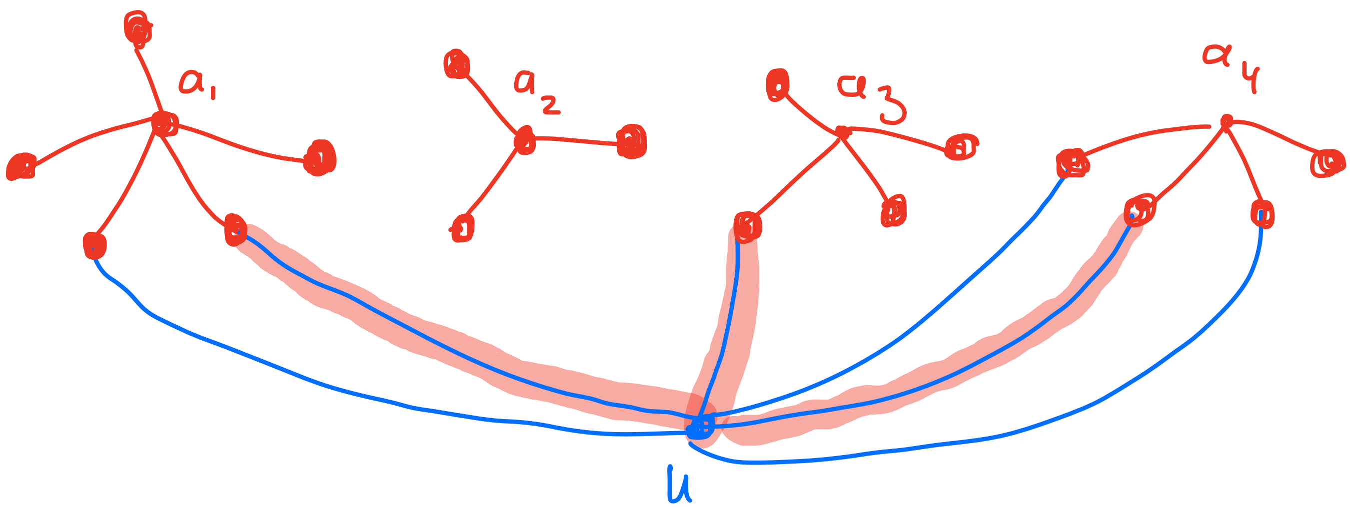
For simplicity: G unweighted, stretch $t=3$

Goal: $|H| = O(n^{1.5} \log n)$ - optimal up to log

Step 1 for every $v \in V$ with deg $\leq \sqrt{n}$, take all edges touching it into H

Step 2
- define $U = \{u \in V \mid deg(u) \geq \sqrt{n}\}$
- build hitting set A for nbrhoods of verts. in U
- for $u \in U$, choose arbitrary nbr $a(u) \in A$ and add the edge $(u, a(u))$ to H
  (we call $a(u)$ the center of u)

Step 3 for each $u \in U$ and $a \in A$:
add <u>one</u> edge (u,w) to $w \in U$ with $a(w) = a$
(if there does not exist such edge - do nothing)



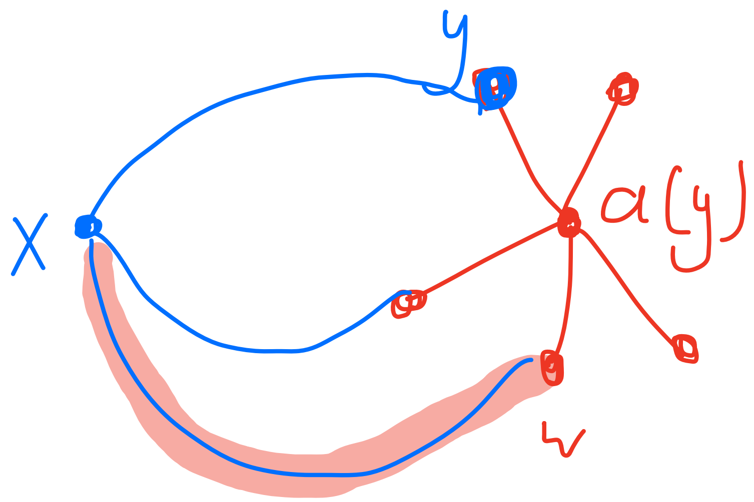
u chooses <u>one</u> edge into each neighboring "star" centered at some $a \in A$

Size analysis: step 1 - $\leq n \cdot \sqrt{n}$
step 2 - $\leq n$
step 3 - $\leq n \cdot |A| = O(n \cdot \sqrt{n} \log n)$

Correctness: as seen in lecture, enough to prove that if $x, y \in V$ are nbrs, then there exist a path of ≤ 3 edges between them in H

- if one of x, y has deg $\leq \sqrt{n}$, then the edge (x,y) was added in Step 1 ✓

- otherwise, $x, y \in \bar{U}$ $\Rightarrow$ the star of $a(y)$ is neighboring to x



So x added an edge to some $w \in \bar{U}$
$a(w) = a(y)$

$$\Rightarrow \quad x \text{ --- } w \text{ --- } a(y) \text{ --- } y \subseteq H$$

added in Step 3

added in step 2

∎