

Linear-Work

Parallel Set Cover

and Variants Using MaNIS

Kanat Tangwongsan

Carnegie Mellon University

(Joint work with Guy Blelloch and Richard Peng)

Min Set Cover

Min Set Cover Max Cover

Min Set Cover

Max Cover

Min-Sum Set Cover

Min Set Cover

Max Cover

Min-Sum Set Cover

Facility Location

Min Set Cover

Max Cover

Min-Sum Set Cover

Facility Location

Asymmetric k-Center

Min Set Cover

Max Cover

Min-Sum Set Cover

Facility Location

Asymmetric k-Center

Min Set Cover

Max Cover

Min-Sum Set Cover

Facility Location

Asymmetric k-Center

while (not done)
 pick highest utility option

simple greedy solution

Min Set Cover

Max Cover

Min-Sum Set Cover

Facility Location

Asymmetric k-Center

while (not done)
 pick highest utility option

simple greedy solution

inherently sequential

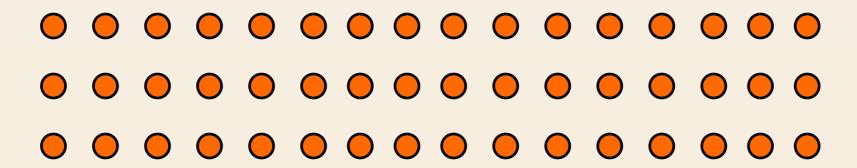


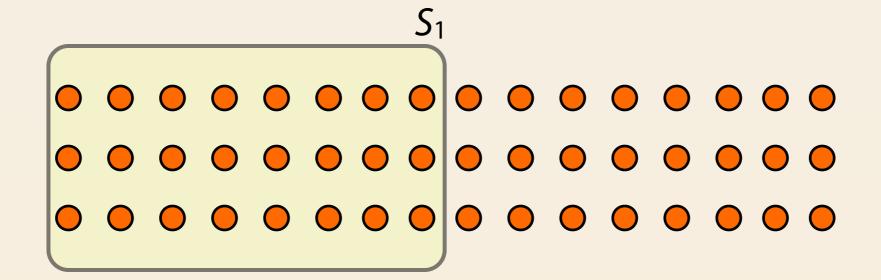
Min Set Cover

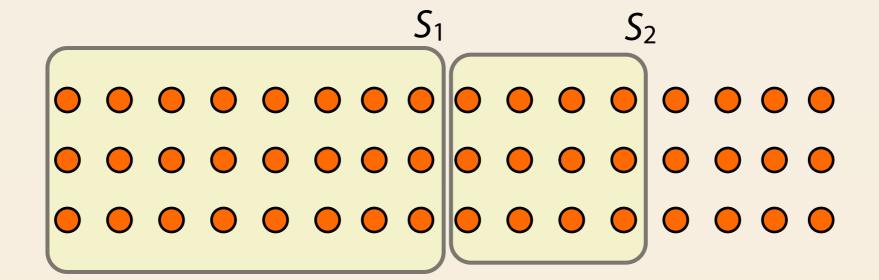
This Talk: MaNIS: Maximal Nearly Independent Set techniques for parallel approximate greedy algorithms

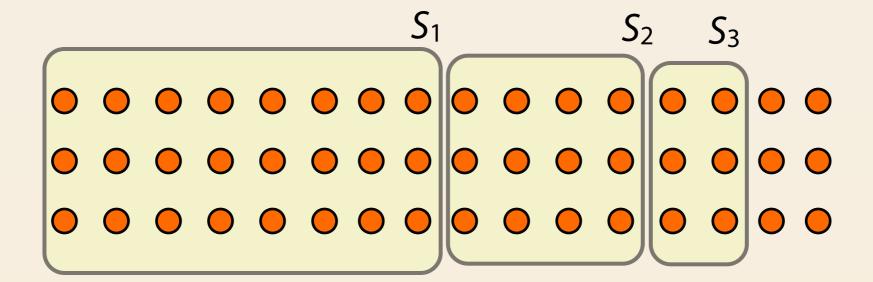
How to select a *maximal* collection of *nearly* non-overlapping sets in linear work and polylog depth?

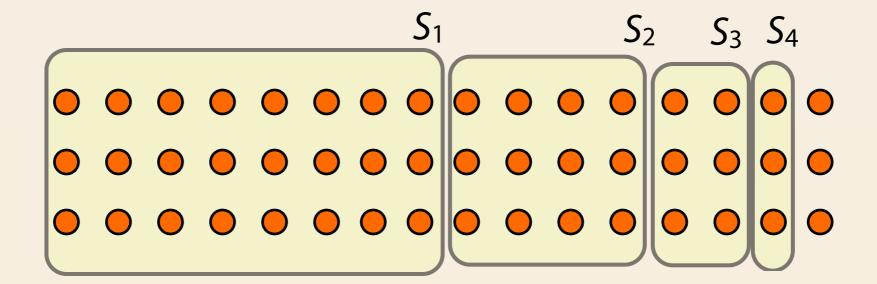
Asymmetric k-Center

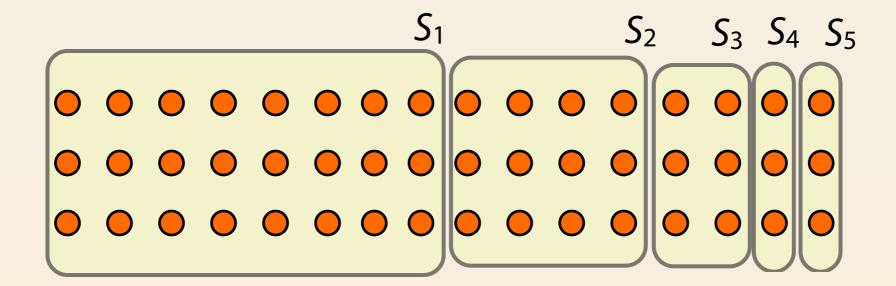


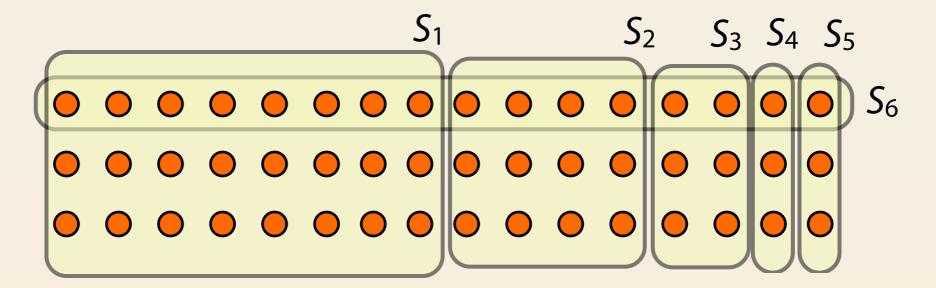


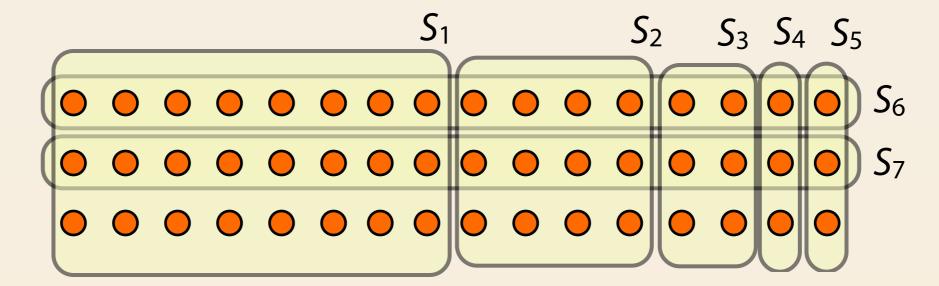


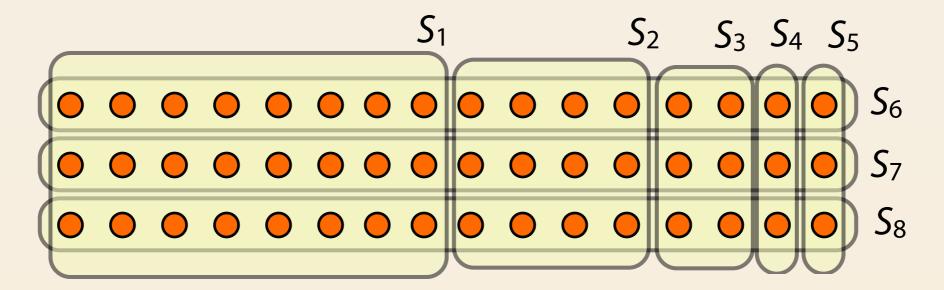




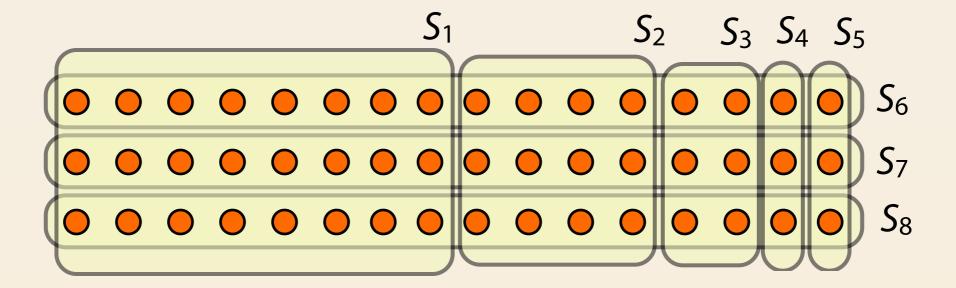






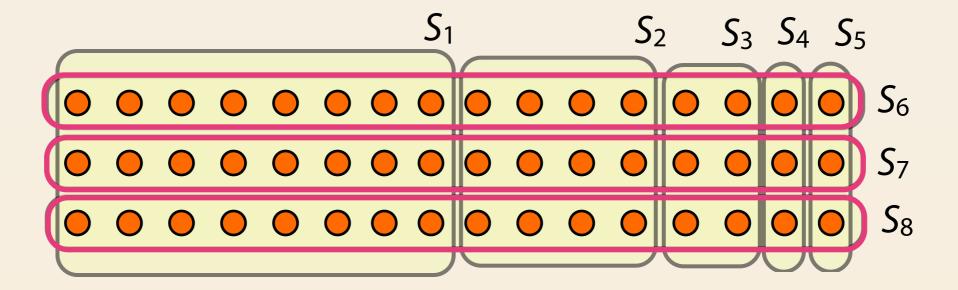


Instance: elements and sets covering them



Task: cover all elements using fewest sets

Instance: elements and sets covering them

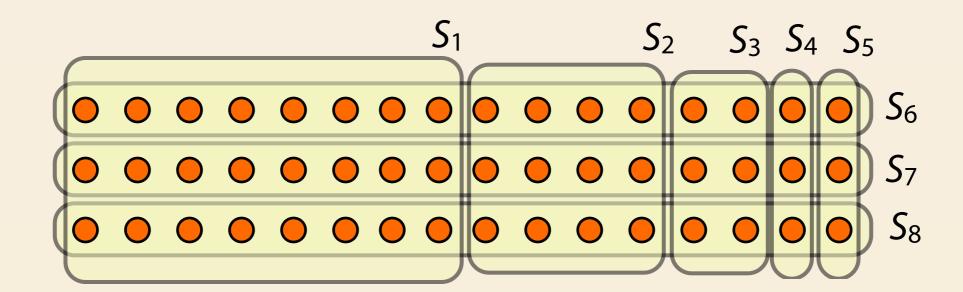


Task: cover all elements using fewest sets

[Johnson'74, Chvatal'79]

For t = 0, 1, ... until elts all covered

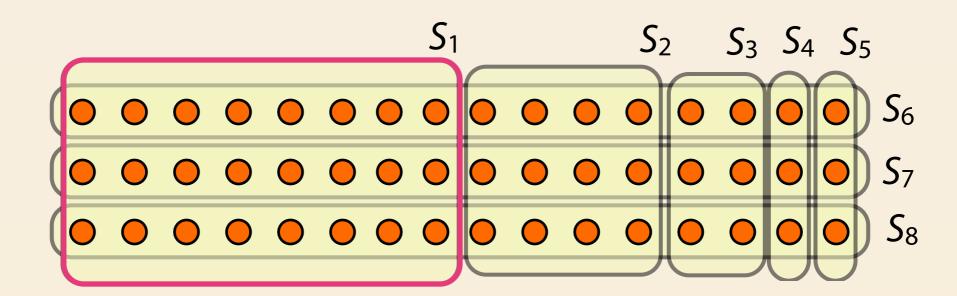
Pick the set that covers the most **new** elements (say X_t)



[Johnson'74, Chvatal'79]

For t = 0, 1, ... until elts all covered

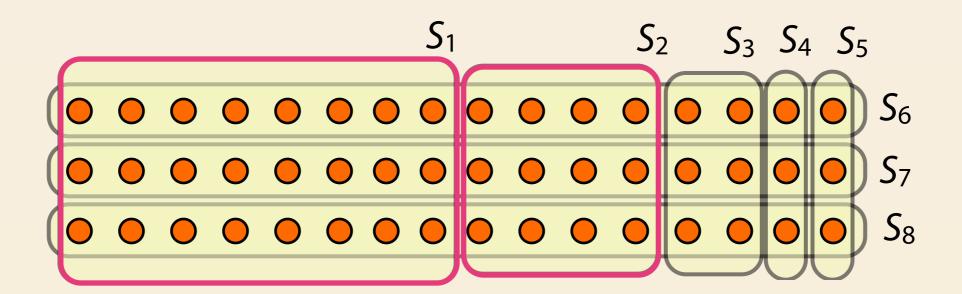
Pick the set that covers the most **new** elements (say X_t)



[Johnson'74, Chvatal'79]

For t = 0, 1, ... until elts all covered

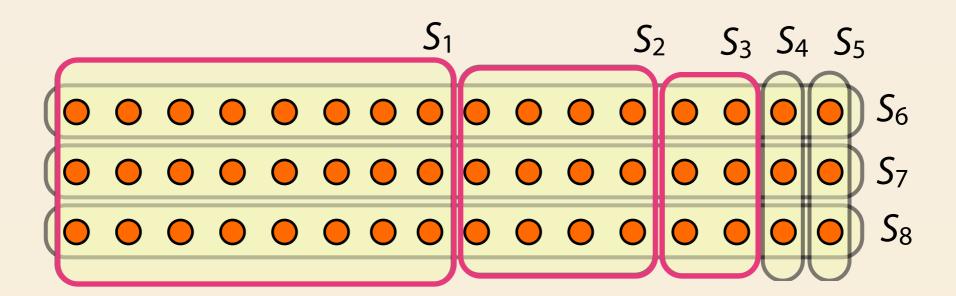
Pick the set that covers the most **new** elements (say X_t)



[Johnson'74, Chvatal'79]

For t = 0, 1, ... until elts all covered

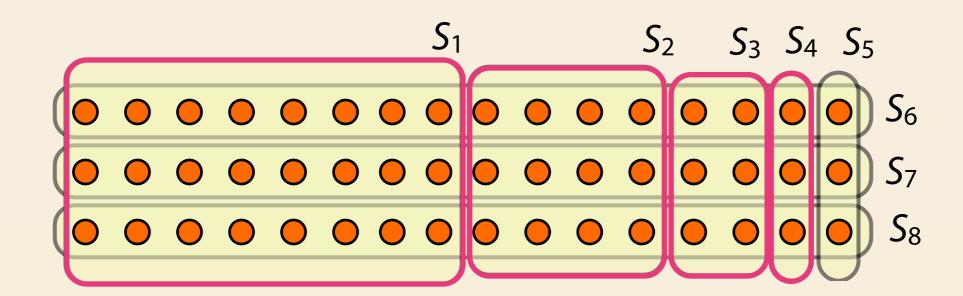
Pick the set that covers the most **new** elements (say X_t)



[Johnson'74, Chvatal'79]

For t = 0, 1, ... until elts all covered

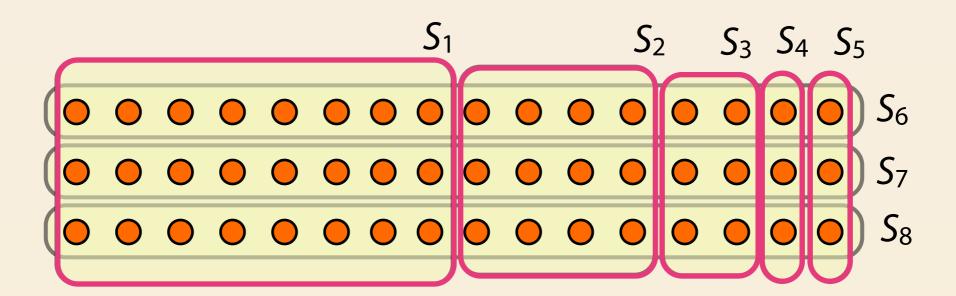
Pick the set that covers the most **new** elements (say X_t)



[Johnson'74, Chvatal'79]

For t = 0, 1, ... until elts all covered

Pick the set that covers the most **new** elements (say X_t)



[Johnson'74, Chvatal'79]

For t = 0, 1, ... until elts all covered

Pick the set that covers the most **new** elements (say X_t)

 X_t = new elements covered

Thm: Greedy is a (ln *n* + 1)-approximation.

[Johnson'74, Chvatal'79]

[Solution of the content of the conte

[Johnson'74, Chvatal'79]

For t = 0, 1, ... until elts all covered

Pick the set that covers the most **new** elements (say X_t)

 X_t = new elements covered

Thm: Greedy is a $(\ln n + 1)$ -approximation.

[Johnson'74, Chvatal'79]

Thm: Can't beat greedy unless P = NP.

[Raz-Safra'97, Feige'98, Alon et al.'06]

Thm: Faithfully greedy set cover is P-complete.

[Bongiovanni et al.'95, Blelloch et al.'11]

Thm: Faithfully greedy set cover is P-complete.

[Bongiovanni et al.'95, Blelloch et al.'11]

m = sum of set sizes, n = # of elements

Thm: Faithfully greedy set cover is P-complete.

[Bongiovanni et al.'95, Blelloch et al.'11]

m = sum of set sizes, n = # of elements

Berger, Rompel, and Shor'94

- $(1+\epsilon)(1+\ln n)$ -approx, RNC O($m\log^4 m$)-work

Rajagopalan and Vazirani'98

- $(2+\epsilon)(1+\ln n)$ -approx, RNC O($m\log^2 m$)-work

Our Result for Set Cover:

 $(1+\epsilon)(1+\ln n)$ -approximation, O(m)-work, $O(\log^3 m)$ -depth

m = sum of set sizes, n = # of elements

Berger, Rompel, and Shor'94

- $(1+\epsilon)(1+\ln n)$ -approx, RNC O($m\log^4 m$)-work

Rajagopalan and Vazirani'98

- $(2+\epsilon)(1+\ln n)$ -approx, RNC O($m\log^2 m$)-work

Our Result for Set Cover:

 $(1+\epsilon)(1+\ln n)$ -approximation, O(m)-work, $O(\log^3 m)$ -depth

m = sum of set sizes, n = # of elements

Our Main Result:

- Formulation of MaNIS
- O(|E|)-work, $O(\log^2 |E|)$ -depth algorithm for MaNIS

handling multiple sets simultaneously

For t = 0, 1, ... until elts all covered

Pick the set that covers the most **new** elements (say X_t)

 X_t = new elements covered

handling multiple sets simultaneously

For t = 0, 1, ... until elts all covered

Pick the set that covers the most new elements (say X_t)

 X_t = new elements covered

handling multiple sets simultaneously

For t = 0, 1, ... until elts all covered

Pick the set that covers the most **new** elements (say X_t)

 X_t = new elements covered

Identify **all** sets that cover at least $(1-\varepsilon)|X_t|$ and bulk-process them a way that mimics greedy

handling multiple sets simultaneously

For t = 0, 1, ... until elts all covered

Pick the set that covers the most **new** elements (say X_t)

 X_t = new elements covered

Identify **all** sets that cover at least $(1-\varepsilon)|X_t|$ and bulk-process them a way that mimics greedy

Observation:

If $|X_{t+1}| \leq (1 - \varepsilon)|X_t|$, then # of rounds is $O(\log_{1+\varepsilon} n)$.

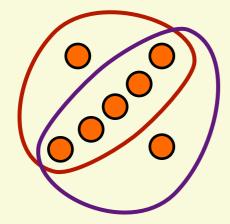
Given sets covering roughly $|X_t|$ (i.e., between $(1-\varepsilon)|X_t|$ and $|X_t|$)

Want: each set chosen to cover roughly $|X_t|$ new elts.

Given sets covering roughly $|X_t|$ (i.e., between $(1-\varepsilon)|X_t|$ and $|X_t|$)

Want: each set chosen to cover roughly $|X_t|$ new elts.

Example 1: lots of sharing

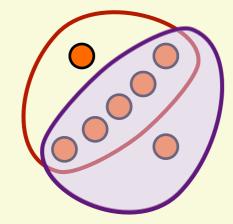


intended behavior: choose one

Given sets covering roughly $|X_t|$ (i.e., between $(1-\varepsilon)|X_t|$ and $|X_t|$)

Want: each set chosen to cover roughly $|X_t|$ new elts.

Example 1: lots of sharing

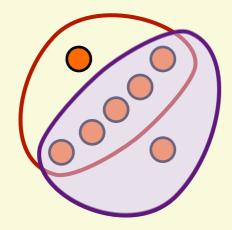


intended behavior: choose one

Given sets covering roughly $|X_t|$ (i.e., between $(1-\varepsilon)|X_t|$ and $|X_t|$)

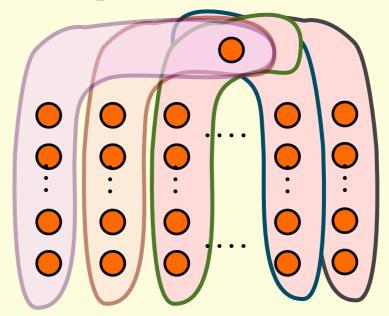
Want: each set chosen to cover roughly $|X_t|$ new elts.

Example 1: lots of sharing



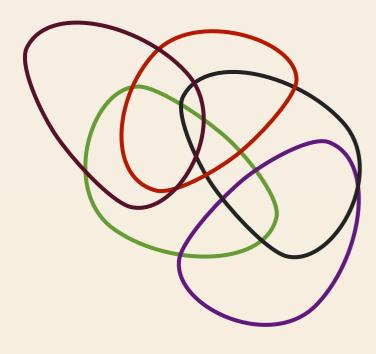
intended behavior: choose one

Example 2: little sharing

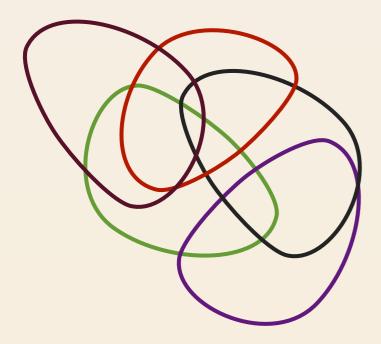


intended behavior: choose all

(cont'd)



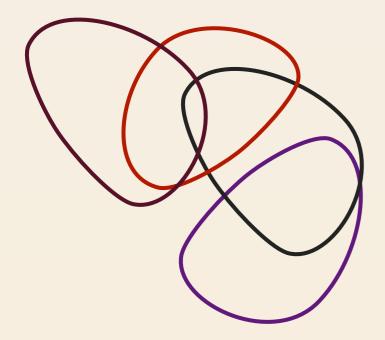
(cont'd)



1

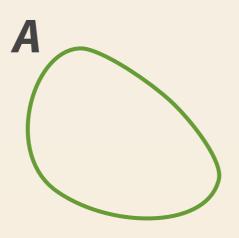
PICK:

(cont'd)

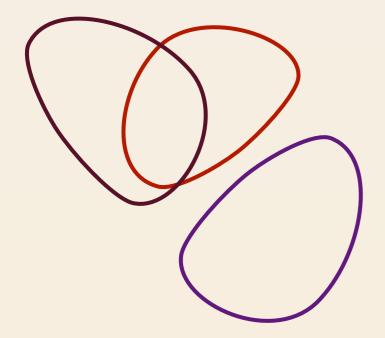




PICK:

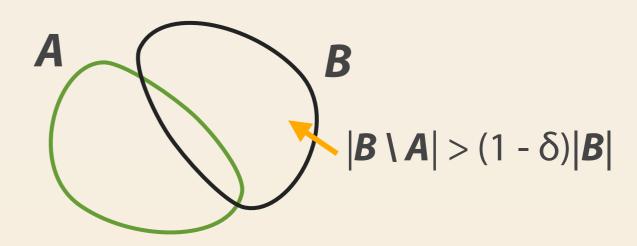


(cont'd)

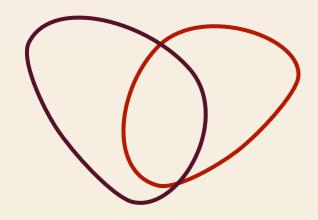


1

PICK:

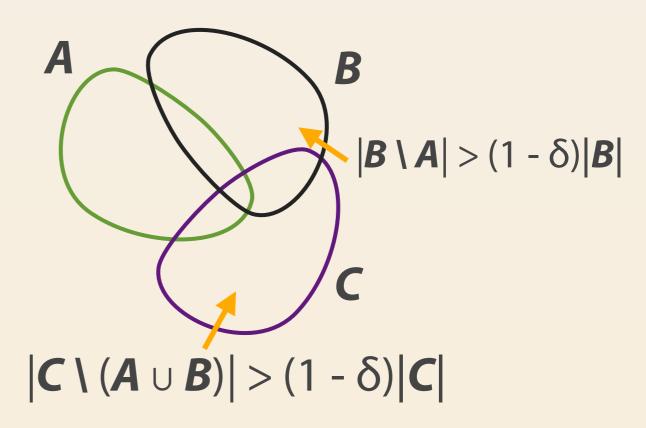


(cont'd)



1

PICK:

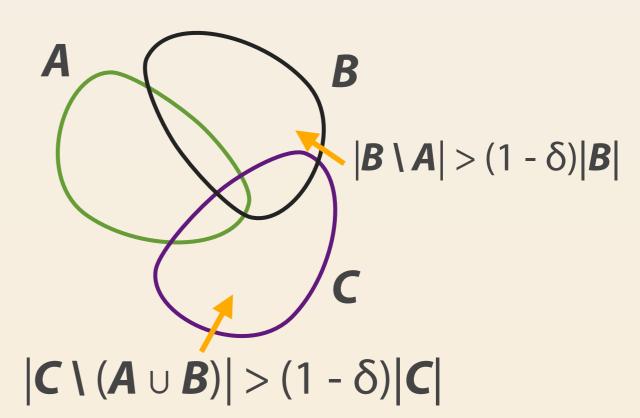


(cont'd)

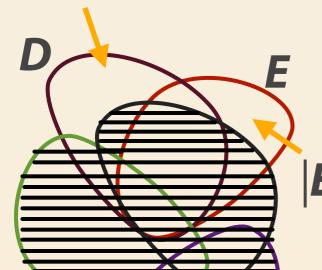
1

PICK:

if has a small-overlapping sequence



|**D**\ (**A** ∪ **B** ∪ **C**) | < (1 -
$$\epsilon$$
)|**D**|



2

REJECT: if retains too few elts

$$|E \setminus (A \cup B \cup C)| < (1 - \varepsilon)|E|$$

formalizing our intuitions

Input: SETS = collection of sets

For $\delta \geq \varepsilon$, (ε, δ) -MaNIS is $J = \langle S_1, ..., S_k \rangle \subseteq$ SETS such that

formalizing our intuitions

Input: SETS = collection of sets

For $\delta \geq \varepsilon$, (ε, δ) -MaNIS is $J = \langle S_1, ..., S_k \rangle \subseteq$ SETS such that

1 Nearly Independent.

$$|S_i \setminus (\cup_{j < i} S_j)| \ge (1 - \delta)|S_i|$$

formalizing our intuitions

Input: SETS = collection of sets

For $\delta \geq \varepsilon$, (ε, δ) -MaNIS is $J = \langle S_1, ..., S_k \rangle \subseteq$ SETS such that

1 Nearly Independent.

Only "slightly" worse than greedy

$$|S_i \setminus (\cup_{j < i} S_j)| \ge (1 - \delta)|S_i|$$

formalizing our intuitions

Input: SETS = collection of sets

For $\delta \geq \varepsilon$, (ε, δ) -MaNIS is $J = \langle S_1, ..., S_k \rangle \subseteq$ SETS such that

1 Nearly Independent.

Only "slightly" worse than greedy

$$|S_i \setminus (\cup_{j < i} S_j)| \ge (1 - \delta)|S_i|$$

Maximal. Each $R_i \notin J$ satisfies $covered = S_1 \cup S_2 \cup \cdots \cup S_k$ $|R_i \setminus covered| < (1 - \varepsilon)|R_i|$

For $\delta \geq \varepsilon$, (ε, δ) -MaNIS is $J = \langle S_1, ..., S_k \rangle \subseteq$ SETS such that

1 Nearly Independent.

$$|S_i \setminus (\cup_{j < i} S_j)| \ge (1 - \delta)|S_i|$$

Maximal. Each $R_i \notin J$ satisfies covered = $S_1 \cup S_2 \cup \cdots \cup S_k$

$$|R_i \setminus \mathsf{covered}| < (1 - \varepsilon)|R_i|$$

For $\delta \geq \varepsilon$, (ε, δ) -MaNIS is $J = \langle S_1, ..., S_k \rangle \subseteq$ SETS such that

1 Nearly Independent.

$$|S_i \setminus (\cup_{j < i} S_j)| \ge (1 - \delta)|S_i|$$

Maximal. Each $R_i \notin J$ satisfies $covered = S_1 \cup S_2 \cup \cdots \cup S_k$ $|R_i \setminus covered| < (1 - \varepsilon)|R_i|$

MaNIS needs not be unique

For $\delta \geq \varepsilon$, (ε, δ) -MaNIS is $J = \langle S_1, ..., S_k \rangle \subseteq$ SETS such that

1 Nearly Independent.

$$|S_i \setminus (\cup_{j < i} S_j)| \ge (1 - \delta)|S_i|$$

2 Maximal. Each $R_i \notin J$ satisfies covered = $S_1 \cup S_2 \cup \cdots \cup S_k$ $|R_i \setminus \text{covered}| < (1 - \varepsilon)|R_i|$

MaNIS needs not be unique

small overlaps on average

$$|\bigcup_i S_i| \ge (1 - \delta) \sum_i |S_i|$$

For $\delta \geq \varepsilon$, (ε, δ) -MaNIS is $J = \langle S_1, ..., S_k \rangle \subseteq$ SETS such that

1 Nearly Independent.

$$|S_i \setminus (\cup_{j < i} S_j)| \ge (1 - \delta)|S_i|$$

Maximal. Each $R_i \notin J$ satisfies $covered = S_1 \cup S_2 \cup \cdots \cup S_k$ $|R_i \setminus covered| < (1 - \varepsilon)|R_i|$

MaNIS needs not be unique

 $\delta = \varepsilon = 0$ is maximal set packing

small overlaps on average

small overlaps
$$|S_i| \ge (1 - \delta) \sum_i |S_i|$$

For $\delta \geq \varepsilon$, (ε, δ) -MaNIS is $J = \langle S_1, ..., S_k \rangle \subseteq$ SETS such that

1 Nearly Independent.

$$|S_i \setminus (\cup_{j < i} S_j)| \ge (1 - \delta)|S_i|$$

Maximal. Each $R_i \notin J$ satisfies $covered = S_1 \cup S_2 \cup \cdots \cup S_k$ $|R_i \setminus covered| < (1 - \varepsilon)|R_i|$

MaNIS needs not be unique

$$\delta = \varepsilon = 0$$
 is maximal set packing

small overlaps on average

small overlaps
$$|S_i| \leq (1-\delta) \sum_i |S_i|$$

simple O(|E|) sequential algo

Implicit in algorithms from previous work

For $\delta \geq \epsilon$:

Implicit in algorithms from previous work

For $\delta \geq \epsilon$:

Berger, Rompel, and Shor'94 (also Chierichetti, Kumar, and Tomkins'10)

 (ε, δ) -MaNIS on average

RNC $O(|E| \log^3 |E|)$ -work

Implicit in algorithms from previous work

For $\delta \geq \epsilon$:

Berger, Rompel, and Shor'94 (also Chierichetti, Kumar, and Tomkins'10)

RNC $O(|E| \log^3 |E|)$ -work

 (ϵ, δ) -MaNIS on average super-linear work

Implicit in algorithms from previous work

For $\delta \geq \epsilon$:

Berger, Rompel, and Shor'94 (also Chierichetti, Kumar, and Tomkins'10)

RNC $O(|E| \log^3 |E|)$ -work

Rajagopalan and Vazirani'98

RNC $O(|E| \log |E|)$ -work

 (ϵ, δ) -MaNIS on average super-linear work

 (ε, δ) -MaNIS

Implicit in algorithms from previous work

For $\delta \geq \epsilon$:

Berger, Rompel, and Shor'94 (also Chierichetti, Kumar, and Tomkins'10)

RNC $O(|E| \log^3 |E|)$ -work

Rajagopalan and Vazirani'98

RNC $O(|E| \log |E|)$ -work

 (ϵ, δ) -MaNIS on average super-linear work

(ε, δ)-MaNIS δ > 1/2

Implicit in algorithms from previous work

For $\delta \geq \epsilon$:

Berger, Rompel, and Shor'94 (also Chierichetti, Kumar, and Tomkins'10)

RNC $O(|E| \log^3 |E|)$ -work

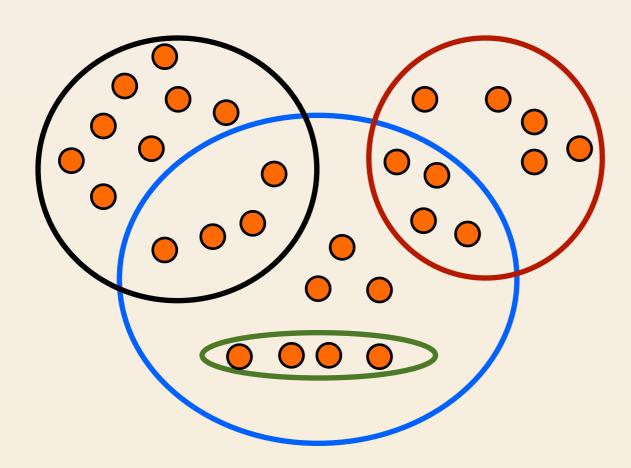
Rajagopalan and Vazirani'98

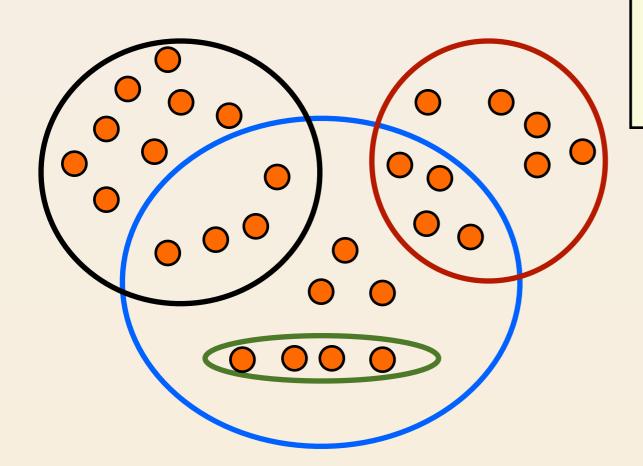
RNC $O(|E| \log |E|)$ -work

 (ϵ, δ) -MaNIS on average super-linear work

(ε, δ)-MaNIS δ > 1/2

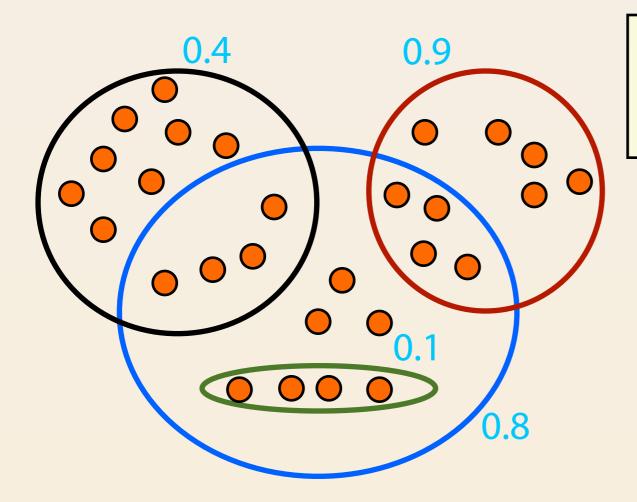
This Work: simple RNC linear work (ϵ , δ)-MaNIS





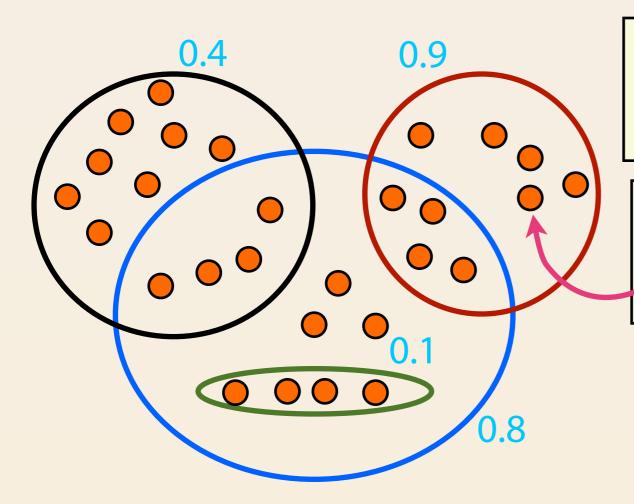
Step 1:

Each set picks a random val \in [0, 1]



Step 1:

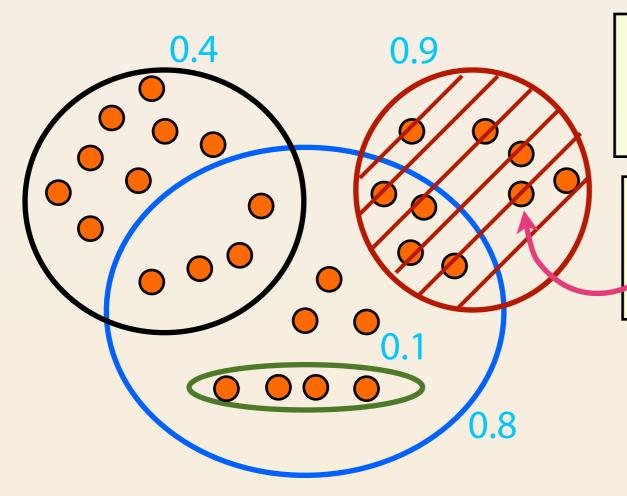
Each set picks a random val \in [0, 1]



Step 1:

Each set picks a random val \in [0, 1]

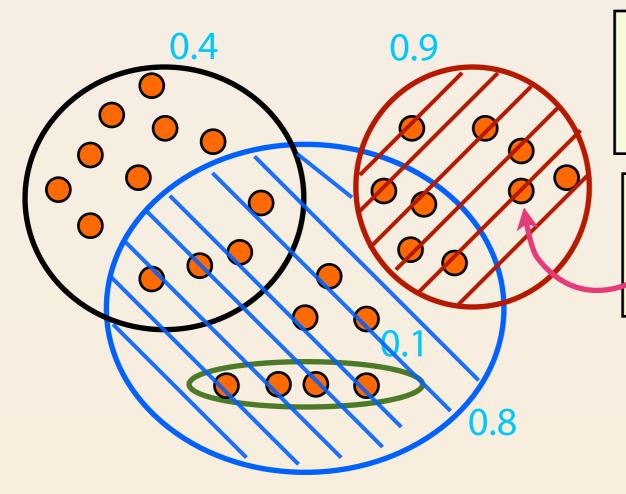
Step 2:



Step 1:

Each set picks a random val \in [0, 1]

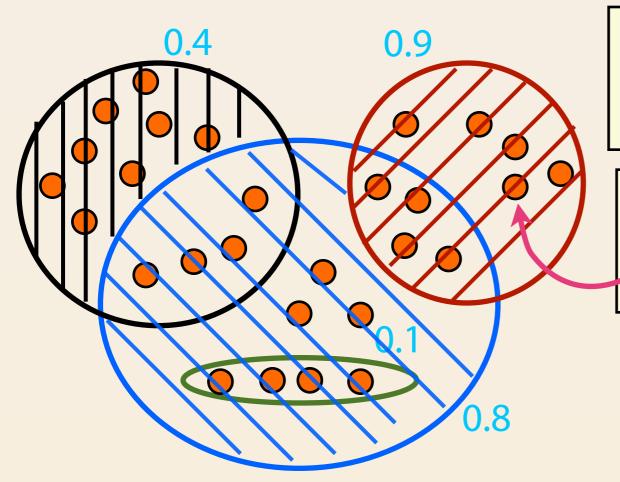
Step 2:



Step 1:

Each set picks a random val \in [0, 1]

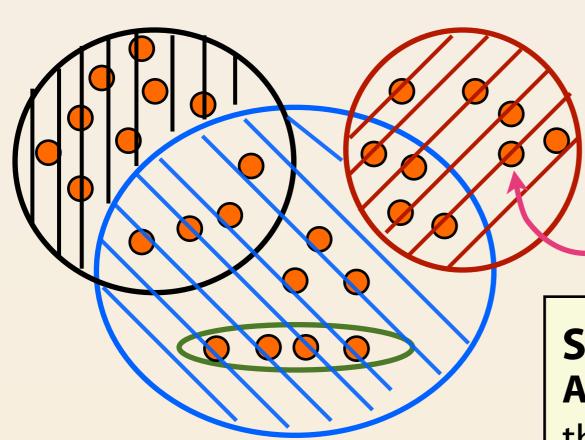
Step 2:



Step 1:

Each set picks a random val \in [0, 1]

Step 2:



Step 1:

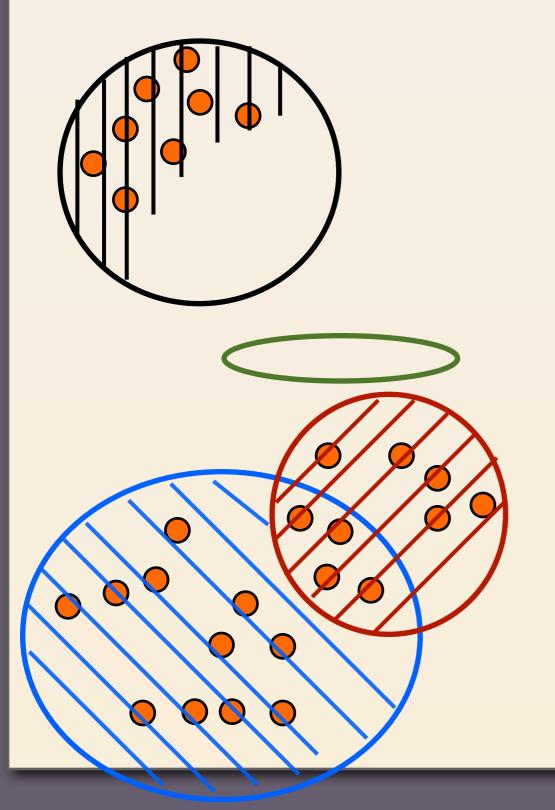
Each set picks a random val \in [0, 1]

Step 2:

 $val(elt) = max\{ val(S) : elt \in S \}$

Step 3:

ACCEPT *S* if $> (1-\delta)$ fraction of *S* has the same val as *S*.



Step 1:

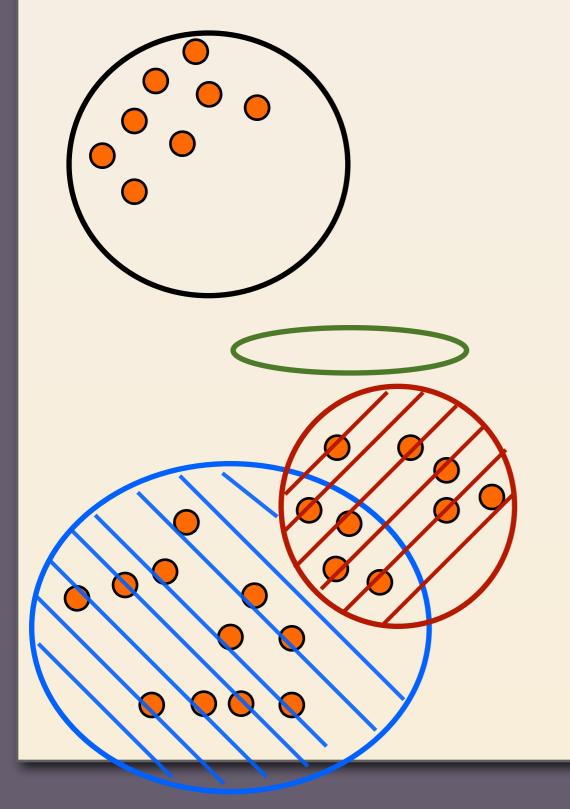
Each set picks a random val \in [0, 1]

Step 2:

 $val(elt) = max\{ val(S) : elt \in S \}$

Step 3:

ACCEPT *S* if $> (1-\delta)$ fraction of *S* has the same val as *S*.



Step 1:

Each set picks a random val \in [0, 1]

Step 2:

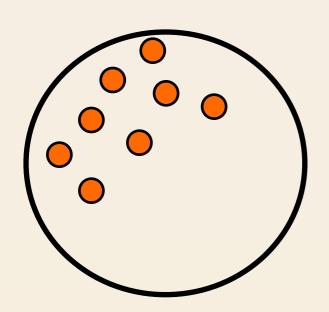
 $val(elt) = max\{ val(S) : elt \in S \}$

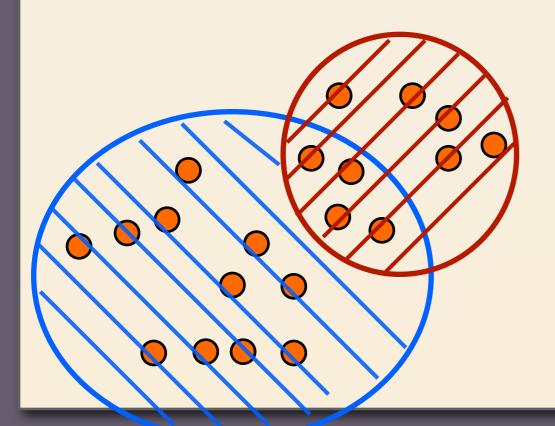
Step 3:

ACCEPT *S* if > $(1-\delta)$ fraction of *S* has the same val as *S*.

Step 4:

REJECT *S* if < (1- ϵ) frac of *S* remains





Step 1:

Each set picks a random val \in [0, 1]

Step 2:

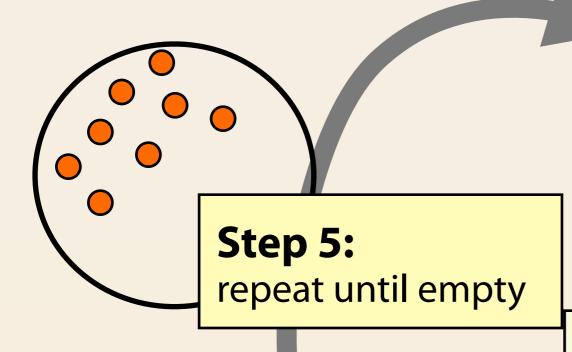
 $val(elt) = max\{ val(S) : elt \in S \}$

Step 3:

ACCEPT *S* if $> (1-\delta)$ fraction of *S* has the same val as *S*.

Step 4:

REJECT *S* if < (1- ϵ) frac of *S* remains



Step 1:

Each set picks a random val \in [0, 1]

Step 2:

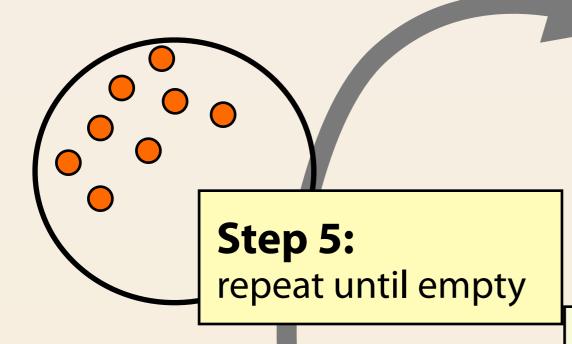
 $val(elt) = max\{ val(S) : elt \in S \}$

Step 3:

ACCEPT *S* if > $(1-\delta)$ fraction of *S* has the same val as *S*.

Step 4:

REJECT *S* if < (1- ϵ) frac of *S* remains



Step 1:

Each set picks a random val \in [0, 1]

Step 2:

 $val(elt) = max\{ val(S) : elt \in S \}$

Step 3:

ACCEPT S if $> (1-\delta)$ fraction of S has the same val as S.

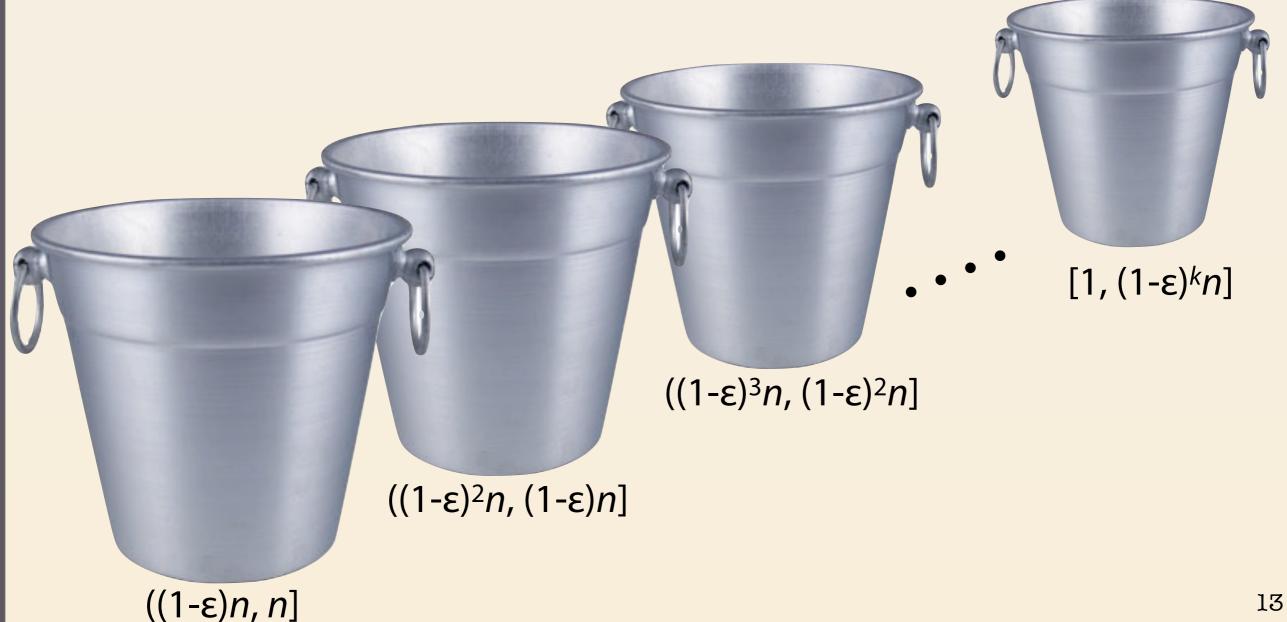
Step 4:

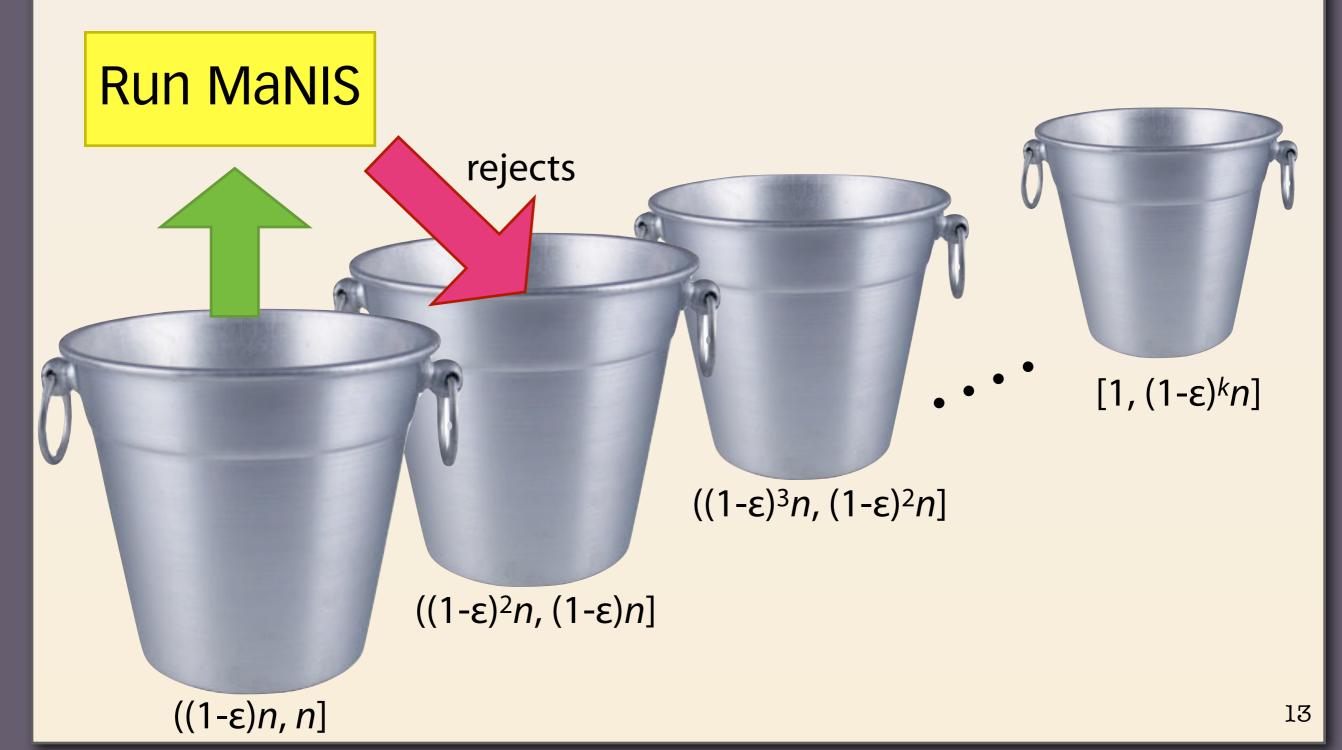
REJECT S if < (1- ε) frac of S remains

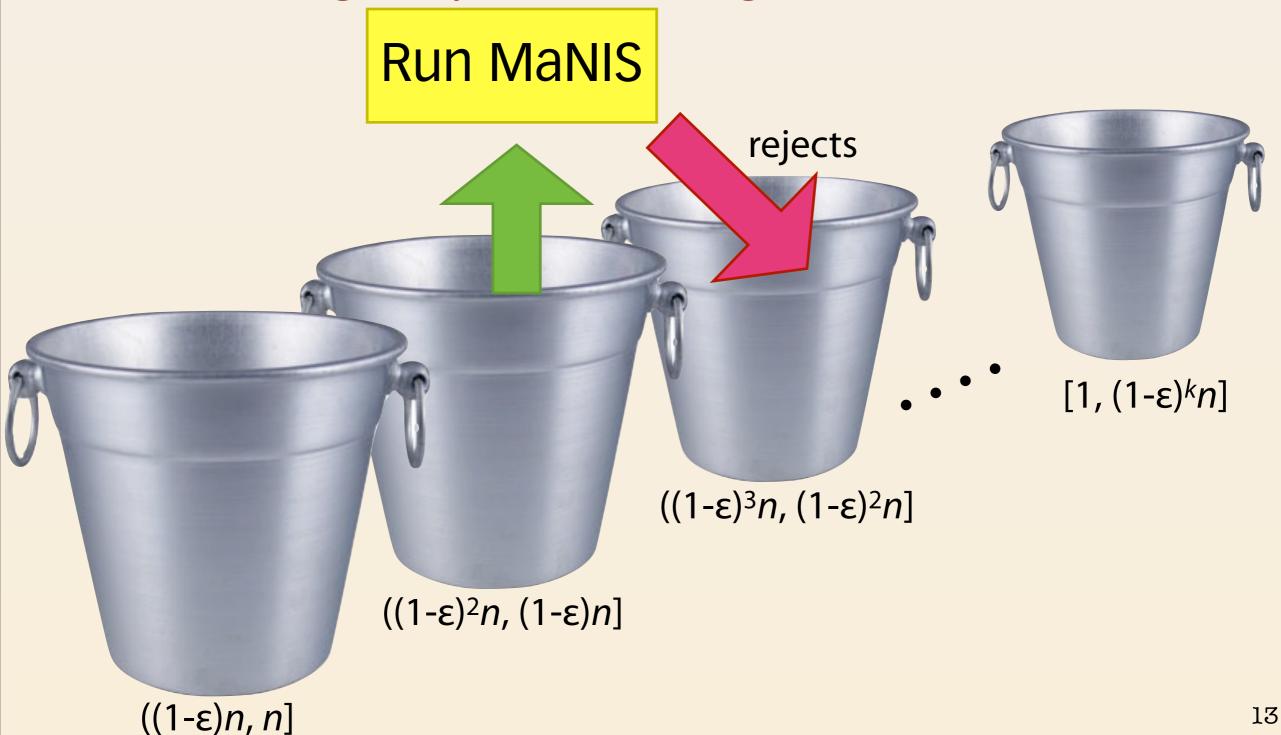
Theorem:

O(|E|)-work, $O(\log^2 |E|)$ -depth algorithm for (ε, δ) -MaNIS

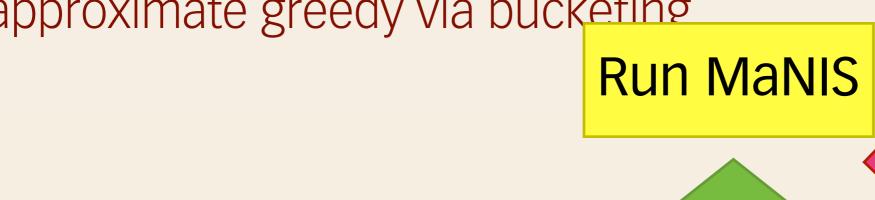


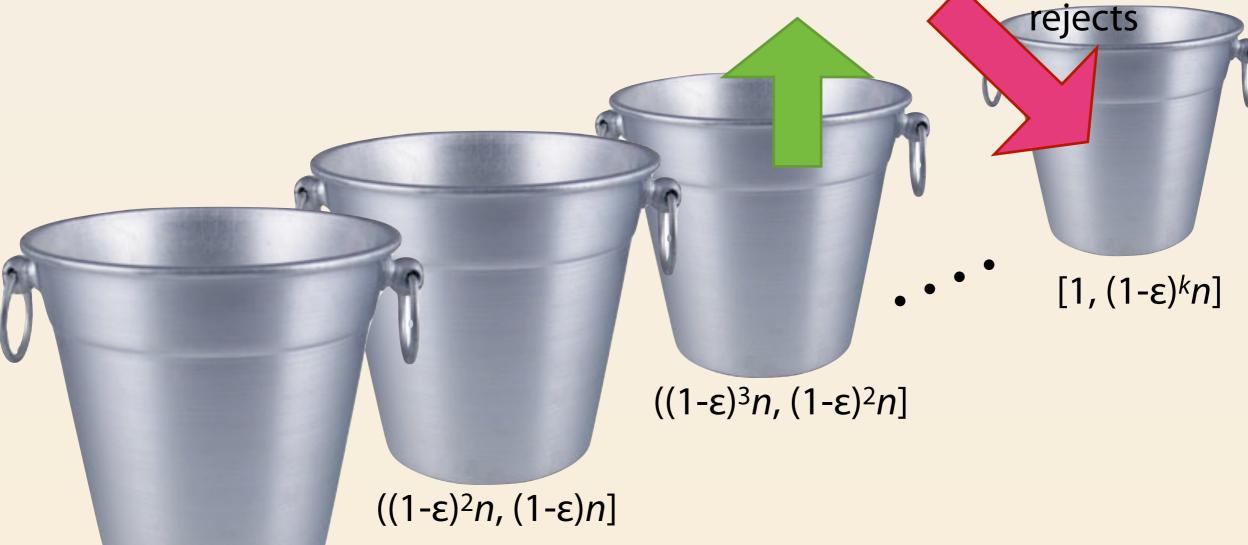






approximate greedy via bucketing





 $((1-\varepsilon)n, n]$

(cont'd)

Initial bucketing: linear work

Iterating through buckets:

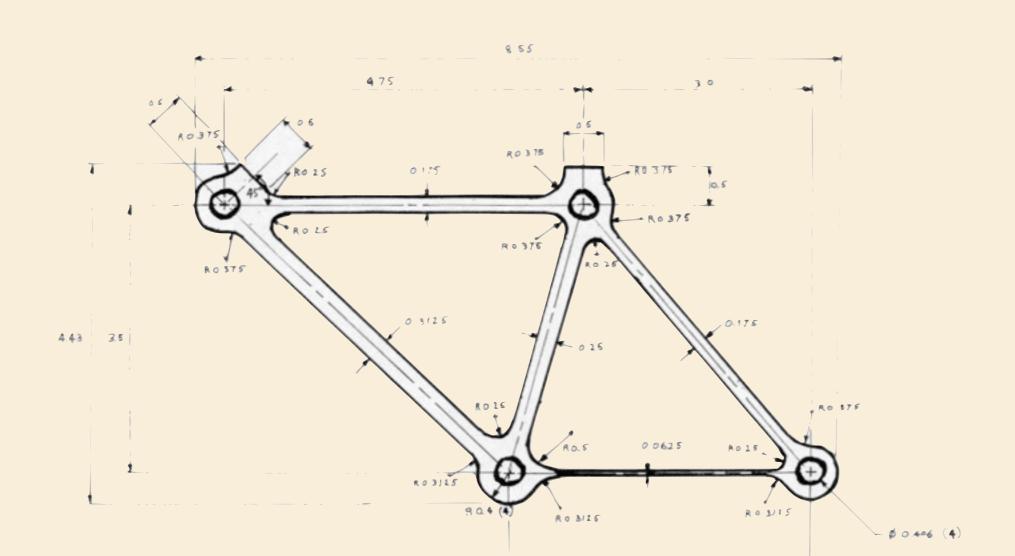
MaNIS is O(m) = O(sum of set sizes in that bucket)

When a set changes buckets, size shrinks by ϵ

Total work: Linear in sum of set sizes

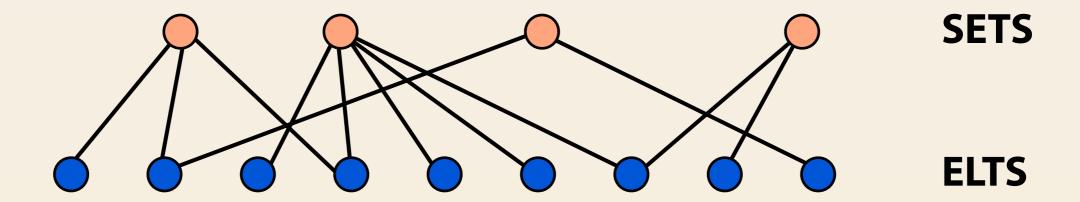
Now for a proof sketch....

O(|E|)-work, $O(\log^2 |E|)$ -depth



Proof's Overview

a bipartite graph view:



Key Lemma:

Each iteration of MaNIS removes a constant fraction of the edges.

Each iteration of MaNIS removes a constant fraction of the edges.

For each set S, $W_S = \text{"# of edges S responsible for deleting"}$

Each iteration of MaNIS removes a constant fraction of the edges.

For each set S, $W_S = \text{"# of edges S responsible for deleting"}$ if S not accepted: $W_S = 0$

Each iteration of MaNIS removes a constant fraction of the edges.

For each set S, $W_S = \text{"# of edges S responsible for deleting"}$

if S not accepted: $W_S = 0$

if S accepted:

 W_S = sum of degrees of S's elts with same val as S

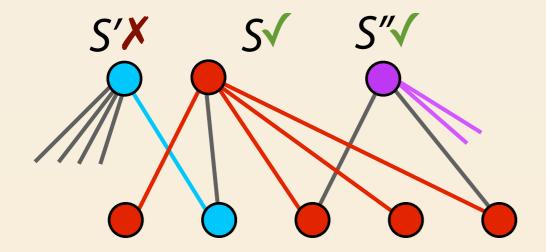
Each iteration of MaNIS removes a constant fraction of the edges.

For each set S, $W_S = "# of edges S responsible for deleting"$

if S not accepted: $W_S = 0$

if S accepted:

 W_S = sum of degrees of S's elts with same val as S



$$W_{S'}=0$$

$$W_S = 1 + 0 + 2 + 1 + 2$$

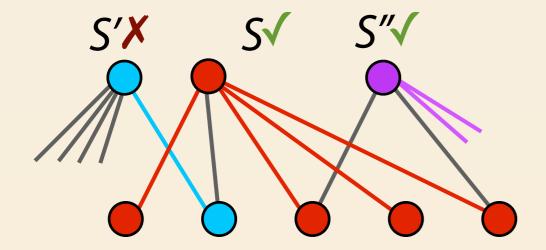
Each iteration of MaNIS removes a constant fraction of the edges.

Observation:

edges removed
$$\geq \sum_{S \in SETS} W_S$$

if S accepted:

 W_S = sum of degrees of S's elts with same val as S



$$W_{S'} = 0$$

 $W_S = 1 + 0 + 2 + 1 + 2$

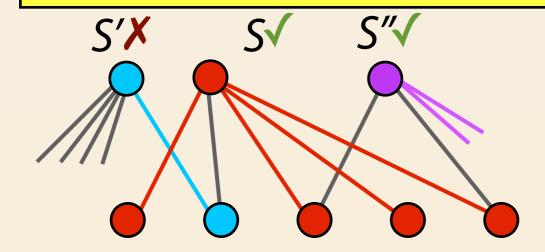
Each iteration of MaNIS removes a constant fraction of the edges.

Observation:

edges removed
$$\geq \sum_{S \in SETS} W_S$$

Technical Claim:

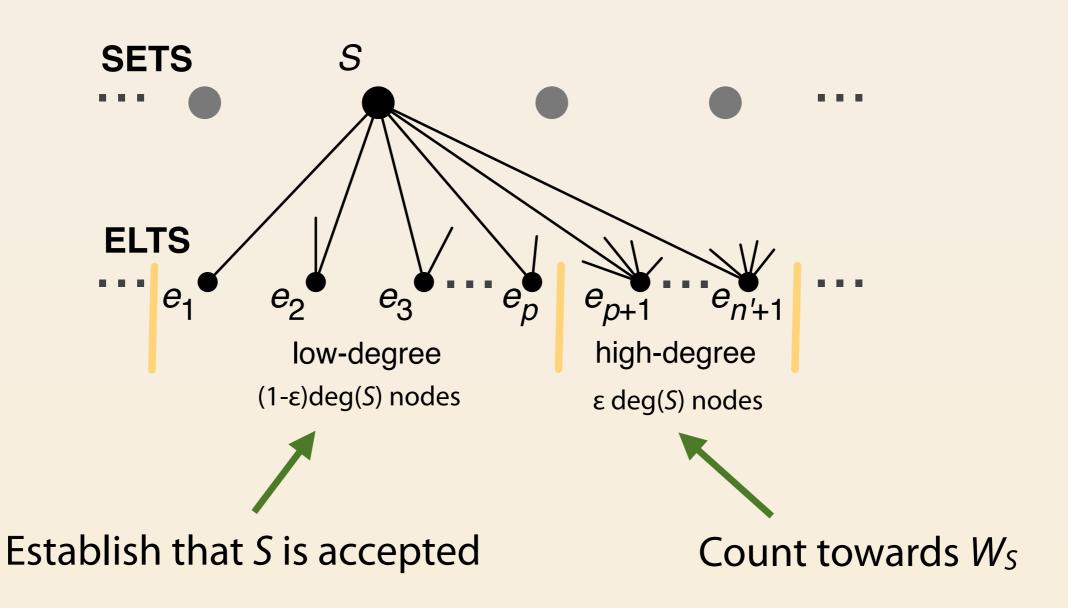
$$\mathbb{E}[W_S] \ge c \cdot \deg(S)$$



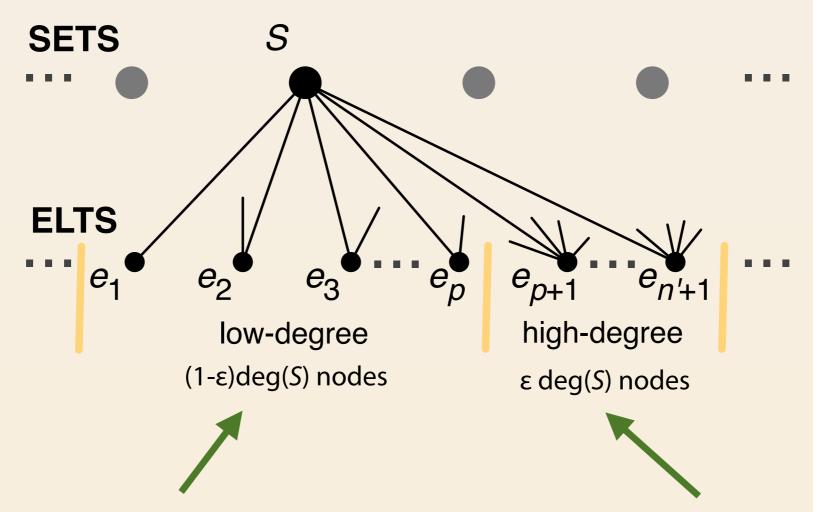
$$W_{S'} = 0$$

$$W_S = 1 + 0 + 2 + 1 + 2$$

$$\mathbb{E}[W_S] \ge c \cdot \deg(S)$$



$$\mathbb{E}[W_S] \ge c \cdot \deg(S)$$



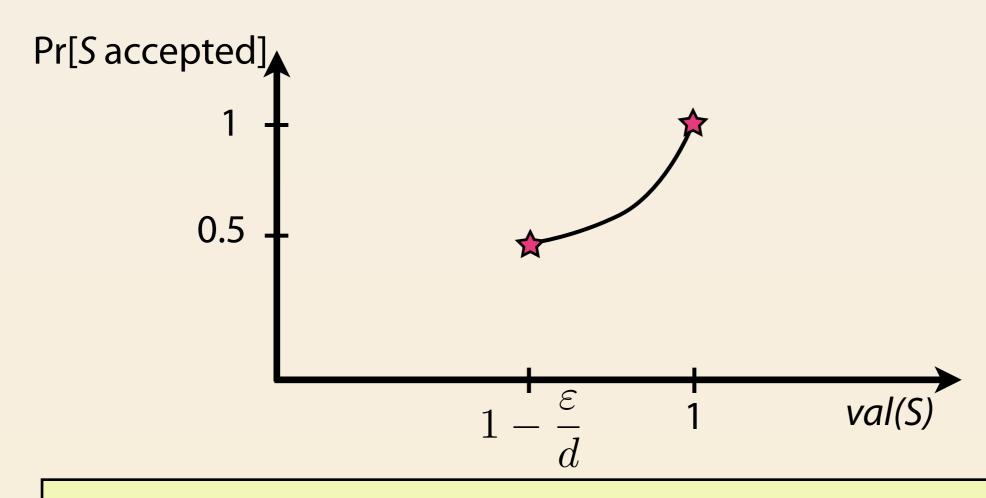
Establish that S is accepted

Count towards W_S

Suppose every node in the low degree group has degree $\leq d$, and every node in the high degree group has degree > d.

$$\mathbb{E}[W_S] \ge c \cdot \deg(S)$$

Use low-degree neighbors of S to establish that S is picked



Claim: If $val(S) \ge 1 - \varepsilon/d$, then $Pr[S \text{ accepted}] \ge 1/2$

$$\mathbb{E}[W_S] \ge c \cdot \deg(S)$$

Claim: If $val(S) \ge 1 - \varepsilon/d$, then $Pr[S \text{ accepted}] \ge 1/2$

Use high-degree neighbors to account for Ws

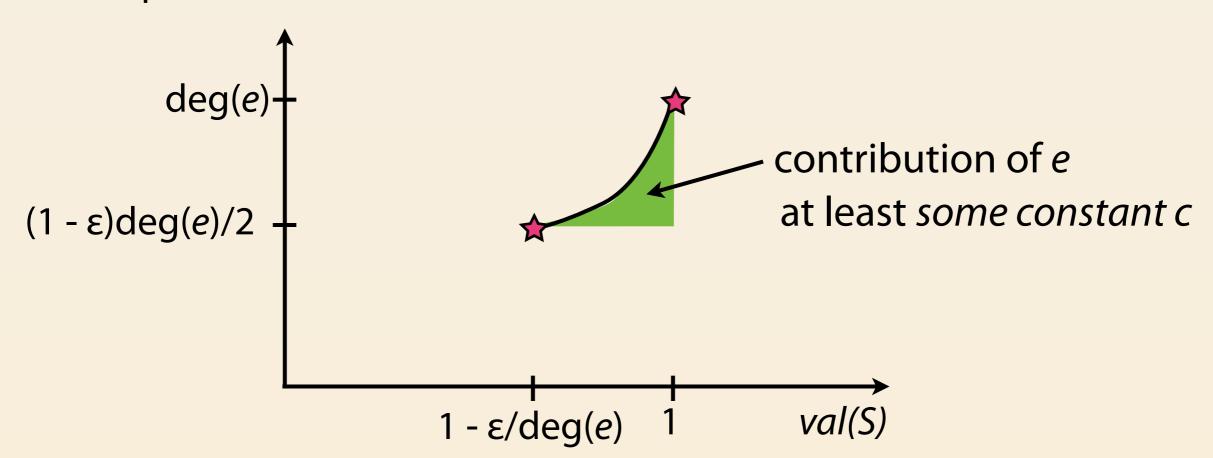
Claim: For a high-degree node e, if $val(S) \ge 1 - \varepsilon/\deg(e)$ and S is accepted, then $\Pr[e \text{ has the same value as } S] \ge 1 - \varepsilon$

$$\mathbb{E}[W_S] \ge c \cdot \deg(S)$$

Claim: If $val(S) \ge 1 - \varepsilon/d$, then $Pr[S \text{ accepted}] \ge 1/2$

Use high-degree neighbors to account for Ws

Claim: For a high-degree node e, if $val(S) \ge 1 - \varepsilon/\deg(e)$ and S is accepted, then $\Pr[e \text{ has the same value as } S] \ge 1 - \varepsilon$



Min Set Cover

Max Cover

Min-Sum Set Cover

Facility Location

Asymmetric k-Center



Min Set Cover

Max Cover

Min-Sum Set Cover

Facility Location

Asymmetric k-Center





Min Set Cover

 $(1+\epsilon)(1+\ln n)$ -approx, linear work

Max Cover

 $(1 - 1/e - \varepsilon)$ -approx, linear work

Min-Sum Set Cover

 $(4+\epsilon)$ -approx, linear work

Facility Location (1.861+ε)-approx, O(nlog n)-work

Asymmetric k-Center

O(log* n)-approx, work-efficient



Min Set Cover

 $(1+\epsilon)(1+\ln n)$ -approx, linear work

Max Cover

 $(1 - 1/e - \varepsilon)$ -approx, linear work

Min-Sum Set Cover

 $(4+\epsilon)$ -approx, linear work

Facility Location (1.861+ε)-approx, O(nlog n)-work

Asymmetric k-Center

O(log* n)-approx, work-efficient



Common Tool: (ε,δ) -MaNIS in O(|E|)-work, $O(\log^2 |E|)$ -depth

Min Set Cover

 $(1+\epsilon)(1+\ln n)$ -approx, linear work

Max Cover

 $(1 - 1/e - \varepsilon)$ -approx, linear work

Min-Sum Set Cover

 $(4+\epsilon)$ -approx, linear work

Facility Location (1.861+ε)-approx, O(nlog n)-work

Asymmetric k-Center

O(log* n)-approx, work-efficient



Thank you!

Common Tool: (ε,δ) -MaNIS in O(|E|)-work, $O(\log^2 |E|)$ -depth