IV. Approximation Algorithms: Covering Problems

Thomas Sauerwald

Easter 2018

UNIVERSITY OF CAMBRIDGE

Outline

Introduction

Vertex Cover

The Set-Covering Problem

Motivation

Many fundamental problems are **NP-complete**, yet they are too important to be abandoned.

Motivation

Many fundamental problems are **NP-complete**, yet they are too important to be abandoned.

Examples: HAMILTON, 3-SAT, VERTEX-COVER, KNAPSACK,...

Motivation

Many fundamental problems are **NP-complete**, yet they are too important to be abandoned.

Examples: HAMILTON, 3-SAT, VERTEX-COVER, KNAPSACK,...

Strategies to cope with NP-complete problems

1. If inputs (or solutions) are small, an algorithm with exponential running time may be satisfactory.
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.

Motivation

Many fundamental problems are **NP-complete**, yet they are too important to be abandoned.

Examples: HAMILTON, 3-SAT, VERTEX-COVER, KNAPSACK,...

Strategies to cope with NP-complete problems

1. If inputs (or solutions) are small, an algorithm with exponential running time may be satisfactory.
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.

Motivation

Many fundamental problems are **NP-complete**, yet they are too important to be abandoned.

Examples: HAMILTON, 3-SAT, VERTEX-COVER, KNAPSACK,...

Strategies to cope with NP-complete problems

1. If inputs (or solutions) are small, an algorithm with exponential running time may be satisfactory.
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.

We will call these **approximation algorithms**.

Performance Ratios for Approximation Algorithms

Approximation Ratio

An algorithm for a problem has approximation ratio $\rho(n)$, if for any input of size n, the cost C of the returned solution and optimal cost C^* satisfy:

$$\max\left(\frac{C}{C^*}, \frac{C^*}{C}\right) \leq \rho(n).$$

Performance Ratios for Approximation Algorithms

Approximation Ratio

An algorithm for a problem has approximation ratio $\rho(n)$, if for any input of size n, the cost C of the returned solution and optimal cost C^* satisfy:

$$\max\left(\frac{C}{C^*}, \frac{C^*}{C}\right) \leq \rho(n).$$

This covers both maximization and minimization problems.

Performance Ratios for Approximation Algorithms

Approximation Ratio

An algorithm for a problem has approximation ratio $\rho(n)$, if for any input of size n, the cost C of the returned solution and optimal cost C^* satisfy:

$$\max\left(\frac{C}{C^*}, \frac{C^*}{C}\right) \le \rho(n).$$

- Maximization problem: $\frac{C^*}{C} \ge 1$

This covers both maximization and minimization problems.

Performance Ratios for Approximation Algorithms

Approximation Ratio

An algorithm for a problem has approximation ratio $\rho(n)$, if for any input of size n, the cost C of the returned solution and optimal cost C^* satisfy:

$$\max\left(\frac{C}{C^*}, \frac{C^*}{C}\right) \leq \rho(n).$$

- Maximization problem: $\frac{C^*}{C} \geq 1$
- Minimization problem: $\frac{C}{C^*} \geq 1$

This covers both maximization and minimization problems.

Performance Ratios for Approximation Algorithms

Approximation Ratio

An algorithm for a problem has approximation ratio $\rho(n)$, if for any input of size n, the cost C of the returned solution and optimal cost C^* satisfy:

$$\max\left(\frac{C}{C^*}, \frac{C^*}{C}\right) \leq \rho(n).$$

- Maximization problem: $\frac{C^*}{C} \geq 1$
- Minimization problem: $\frac{C}{C^*} \geq 1$

This covers both maximization and minimization problems.

For many problems: tradeoff between runtime and approximation ratio.

Performance Ratios for Approximation Algorithms

Approximation Ratio

An algorithm for a problem has approximation ratio $\rho(n)$, if for any input of size n, the cost C of the returned solution and optimal cost C^* satisfy:

$$\max\left(\frac{C}{C^*}, \frac{C^*}{C}\right) \leq \rho(n).$$

- Maximization problem: $\frac{C^*}{C} \geq 1$
- Minimization problem: $\frac{C}{C^*} \geq 1$

This covers both maximization and minimization problems.

For many problems: tradeoff between runtime and approximation ratio.

Approximation Schemes

Performance Ratios for Approximation Algorithms

Approximation Ratio

An algorithm for a problem has approximation ratio $\rho(n)$, if for any input of size n, the cost C of the returned solution and optimal cost C^* satisfy:

$$\max\left(\frac{C}{C^*}, \frac{C^*}{C}\right) \leq \rho(n).$$

- Maximization problem: $\frac{C^*}{C} \geq 1$
- Minimization problem: $\frac{C}{C^*} \geq 1$

This covers both maximization and minimization problems.

For many problems: tradeoff between runtime and approximation ratio.

Approximation Schemes

An approximation scheme is an approximation algorithm, which given any input and $\epsilon > 0$, is a $(1+\epsilon)$-approximation algorithm.

Performance Ratios for Approximation Algorithms

Approximation Ratio

An algorithm for a problem has approximation ratio $\rho(n)$, if for any input of size n, the cost C of the returned solution and optimal cost C^* satisfy:

$$\max\left(\frac{C}{C^*}, \frac{C^*}{C}\right) \leq \rho(n).$$

- Maximization problem: $\frac{C^*}{C} \geq 1$
- Minimization problem: $\frac{C}{C^*} \geq 1$

This covers both maximization and minimization problems.

For many problems: tradeoff between runtime and approximation ratio.

Approximation Schemes

An approximation scheme is an approximation algorithm, which given any input and $\epsilon > 0$, is a $(1+\epsilon)$-approximation algorithm.

- It is a polynomial-time approximation scheme (PTAS) if for any fixed $\epsilon > 0$, the runtime is polynomial in n.

Performance Ratios for Approximation Algorithms

Approximation Ratio

An algorithm for a problem has approximation ratio $\rho(n)$, if for any input of size n, the cost C of the returned solution and optimal cost C^* satisfy:

$$\max\left(\frac{C}{C^*}, \frac{C^*}{C}\right) \leq \rho(n).$$

- Maximization problem: $\frac{C^*}{C} \geq 1$
- Minimization problem: $\frac{C}{C^*} \geq 1$

This covers both maximization and minimization problems.

For many problems: tradeoff between runtime and approximation ratio.

Approximation Schemes

An approximation scheme is an approximation algorithm, which given any input and $\epsilon > 0$, is a $(1+\epsilon)$-approximation algorithm.

- It is a polynomial-time approximation scheme (PTAS) if for any fixed $\epsilon > 0$, the runtime is polynomial in n. For example, $O(n^{2/\epsilon})$.

Performance Ratios for Approximation Algorithms

Approximation Ratio

An algorithm for a problem has approximation ratio $\rho(n)$, if for any input of size n, the cost C of the returned solution and optimal cost C^* satisfy:

$$\max\left(\frac{C}{C^*}, \frac{C^*}{C}\right) \leq \rho(n).$$

- Maximization problem: $\frac{C^*}{C} \geq 1$
- Minimization problem: $\frac{C}{C^*} \geq 1$

This covers both maximization and minimization problems.

For many problems: tradeoff between runtime and approximation ratio.

Approximation Schemes

An approximation scheme is an approximation algorithm, which given any input and $\epsilon > 0$, is a $(1+\epsilon)$-approximation algorithm.

- It is a polynomial-time approximation scheme (PTAS) if for any fixed $\epsilon > 0$, the runtime is polynomial in n. For example, $O(n^{2/\epsilon})$.
- It is a fully polynomial-time approximation scheme (FPTAS) if the runtime is polynomial in both $1/\epsilon$ and n.

Performance Ratios for Approximation Algorithms

Approximation Ratio

An algorithm for a problem has approximation ratio $\rho(n)$, if for any input of size n, the cost C of the returned solution and optimal cost C^* satisfy:

$$\max\left(\frac{C}{C^*}, \frac{C^*}{C}\right) \leq \rho(n).$$

- Maximization problem: $\frac{C^*}{C} \geq 1$
- Minimization problem: $\frac{C}{C^*} \geq 1$

This covers both maximization and minimization problems.

For many problems: tradeoff between runtime and approximation ratio.

Approximation Schemes

An approximation scheme is an approximation algorithm, which given any input and $\epsilon > 0$, is a $(1+\epsilon)$-approximation algorithm.

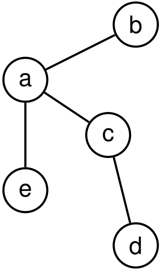
- It is a polynomial-time approximation scheme (PTAS) if for any fixed $\epsilon > 0$, the runtime is polynomial in n. For example, $O(n^{2/\epsilon})$.
- It is a fully polynomial-time approximation scheme (FPTAS) if the runtime is polynomial in both $1/\epsilon$ and n. For example, $O((1/\epsilon)^2 \cdot n^3)$.

Outline

The Vertex-Cover Problem

Vertex Cover Problem

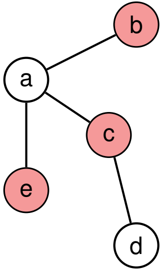
- Given: Undirected graph $G = (V, E)$
- Goal: Find a minimum-cardinality subset $V' \subseteq V$ such that if $(u, v) \in E(G)$, then $u \in V'$ or $v \in V'$.



The Vertex-Cover Problem

Vertex Cover Problem

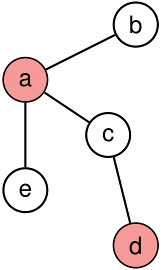
- Given: Undirected graph $G = (V, E)$
- Goal: Find a minimum-cardinality subset $V' \subseteq V$ such that if $(u, v) \in E(G)$, then $u \in V'$ or $v \in V'$.



The Vertex-Cover Problem

Vertex Cover Problem

- Given: Undirected graph $G = (V, E)$
- Goal: Find a minimum-cardinality subset $V' \subseteq V$ such that if $(u, v) \in E(G)$, then $u \in V'$ or $v \in V'$.

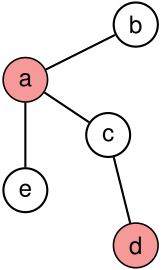


The Vertex-Cover Problem

We are covering edges by picking vertices!

Vertex Cover Problem

- Given: Undirected graph $G = (V, E)$
- Goal: Find a minimum-cardinality subset $V' \subseteq V$ such that if $(u, v) \in E(G)$, then $u \in V'$ or $v \in V'$.



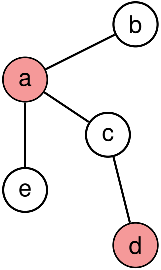
The Vertex-Cover Problem

We are covering edges by picking vertices!

Vertex Cover Problem

- Given: Undirected graph $G = (V, E)$
- Goal: Find a minimum-cardinality subset $V' \subseteq V$ such that if $(u, v) \in E(G)$, then $u \in V'$ or $v \in V'$.

This is an NP-hard problem.



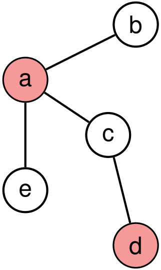
The Vertex-Cover Problem

We are covering edges by picking vertices!

Vertex Cover Problem

- Given: Undirected graph $G = (V, E)$
- Goal: Find a minimum-cardinality subset $V' \subseteq V$ such that if $(u, v) \in E(G)$, then $u \in V'$ or $v \in V'$.

This is an NP-hard problem.



Applications:

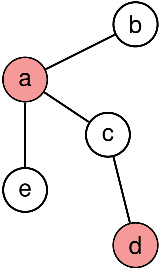
The Vertex-Cover Problem

We are covering edges by picking vertices!

Vertex Cover Problem

- Given: Undirected graph $G = (V, E)$
- Goal: Find a minimum-cardinality subset $V' \subseteq V$ such that if $(u, v) \in E(G)$, then $u \in V'$ or $v \in V'$.

This is an NP-hard problem.



Applications:

- Every edge forms a task, and every vertex represents a person/machine which can execute that task

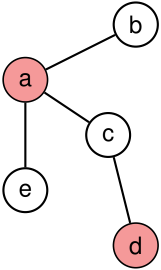
The Vertex-Cover Problem

We are covering edges by picking vertices!

Vertex Cover Problem

- Given: Undirected graph $G = (V, E)$
- Goal: Find a minimum-cardinality subset $V' \subseteq V$ such that if $(u, v) \in E(G)$, then $u \in V'$ or $v \in V'$.

This is an NP-hard problem.



Applications:

- Every edge forms a task, and every vertex represents a person/machine which can execute that task
- Perform all tasks with the minimal amount of resources

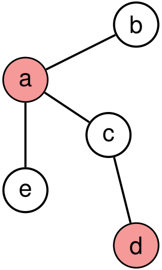
The Vertex-Cover Problem

We are covering edges by picking vertices!

Vertex Cover Problem

- Given: Undirected graph $G = (V, E)$
- Goal: Find a minimum-cardinality subset $V' \subseteq V$ such that if $(u, v) \in E(G)$, then $u \in V'$ or $v \in V'$.

This is an NP-hard problem.



Applications:

- Every edge forms a task, and every vertex represents a person/machine which can execute that task
- Perform all tasks with the minimal amount of resources
- Extensions: weighted vertices or hypergraphs ($\rightsquigarrow$ Set-Covering Problem)

An Approximation Algorithm based on Greedy

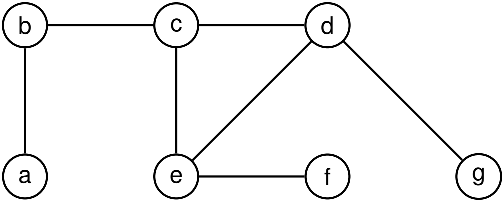
APPROX-VERTEX-COVER(G)

```
C = ∅
E' = G.E
while E' ≠ ∅
    let (u, v) be an arbitrary edge of E'
    C = C ∪ {u, v}
    remove from E' every edge incident on either u or v
return C
```

An Approximation Algorithm based on Greedy

APPROX-VERTEX-COVER(G)

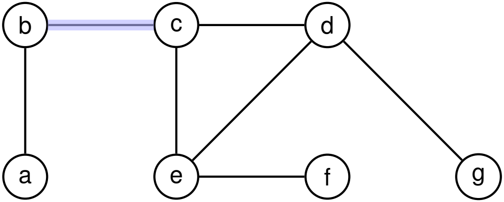
```
C = ∅
E' = G.E
while E' ≠ ∅
    let (u, v) be an arbitrary edge of E'
    C = C ∪ {u, v}
    remove from E' every edge incident on either u or v
return C
```



An Approximation Algorithm based on Greedy

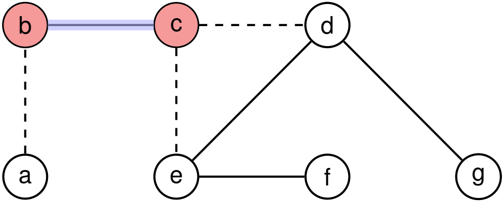
APPROX-VERTEX-COVER(G)

```
C = ∅
E' = G.E
while E' ≠ ∅
    let (u, v) be an arbitrary edge of E'
    C = C ∪ {u, v}
    remove from E' every edge incident on either u or v
return C
```



An Approximation Algorithm based on Greedy

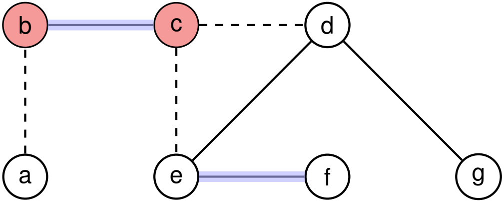
```
APPROX-VERTEX-COVER(G)
C = ∅
E' = G.E
while E' ≠ ∅
    let (u, v) be an arbitrary edge of E'
    C = C ∪ {u, v}
    remove from E' every edge incident on either u or v
return C
```



An Approximation Algorithm based on Greedy

APPROX-VERTEX-COVER(G)

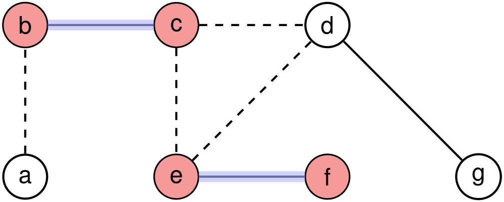
```
C = ∅
E' = G.E
while E' ≠ ∅
    let (u, v) be an arbitrary edge of E'
    C = C ∪ {u, v}
    remove from E' every edge incident on either u or v
return C
```



An Approximation Algorithm based on Greedy

APPROX-VERTEX-COVER(G)

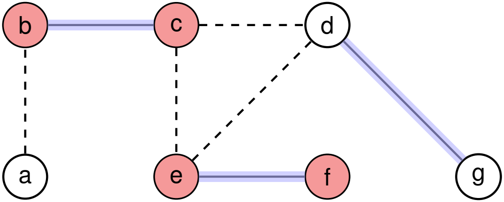
```
C = ∅
E' = G.E
while E' ≠ ∅
    let (u, v) be an arbitrary edge of E'
    C = C ∪ {u, v}
    remove from E' every edge incident on either u or v
return C
```



An Approximation Algorithm based on Greedy

APPROX-VERTEX-COVER(G)

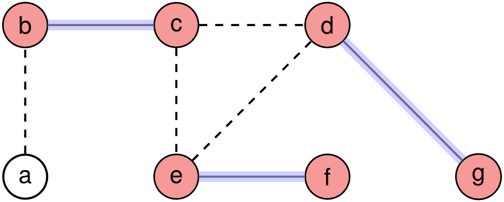
$C = \emptyset$
$E' = G.E$
while $E' \neq \emptyset$
    let (u, v) be an arbitrary edge of E'
    $C = C \cup \{u, v\}$
    remove from E' every edge incident on either u or v
return C



An Approximation Algorithm based on Greedy

APPROX-VERTEX-COVER(G)

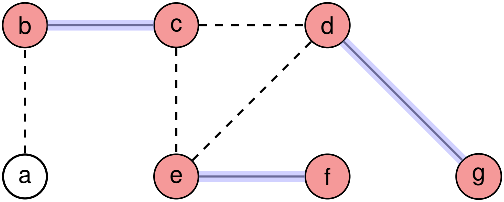
$C = \emptyset$
$E' = G.E$
while $E' \neq \emptyset$
    let (u, v) be an arbitrary edge of E'
    $C = C \cup \{u, v\}$
    remove from E' every edge incident on either u or v
return C



An Approximation Algorithm based on Greedy

APPROX-VERTEX-COVER(G)

$C = \emptyset$
$E' = G.E$
while $E' \neq \emptyset$
    let (u, v) be an arbitrary edge of E'
    $C = C \cup \{u, v\}$
    remove from E' every edge incident on either u or v
return C

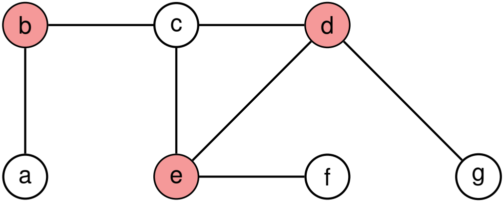


APPROX-VERTEX-COVER produces a set of size 6.

An Approximation Algorithm based on Greedy

APPROX-VERTEX-COVER(G)

```
C = ∅
E' = G.E
while E' ≠ ∅
    let (u, v) be an arbitrary edge of E'
    C = C ∪ {u, v}
    remove from E' every edge incident on either u or v
return C
```



The optimal solution has size 3.

Analysis of Greedy for Vertex Cover

```
APPROX-VERTEX-COVER(G)
C = ∅
E' = G.E
while E' ≠ ∅
    let (u, v) be an arbitrary edge of E'
    C = C ∪ {u, v}
    remove from E' every edge incident on either u or v
return C
```

Analysis of Greedy for Vertex Cover

APPROX-VERTEX-COVER(G)

$C = \emptyset$
$E' = G.E$
while $E' \neq \emptyset$
  let (u, v) be an arbitrary edge of E'
  $C = C \cup \{u, v\}$
  remove from E' every edge incident on either u or v
return C

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Analysis of Greedy for Vertex Cover

APPROX-VERTEX-COVER(G)

```
C = ∅
E' = G.E
while E' ≠ ∅
    let (u, v) be an arbitrary edge of E'
    C = C ∪ {u, v}
    remove from E' every edge incident on either u or v
return C
```

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Proof:

Analysis of Greedy for Vertex Cover

```
APPROX-VERTEX-COVER(G)
C = ∅
E' = G.E
while E' ≠ ∅
    let (u, v) be an arbitrary edge of E'
    C = C ∪ {u, v}
    remove from E' every edge incident on either u or v
return C
```

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Proof:

- Running time is $O(V + E)$ (using adjacency lists to represent E')

Analysis of Greedy for Vertex Cover

APPROX-VERTEX-COVER(G)

```
1  C = ∅
2  E' = G.E
3  while E' ≠ ∅
4      let (u, v) be an arbitrary edge of E'
5      C = C ∪ {u, v}
6      remove from E' every edge incident on either u or v
7  return C
```

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Proof:

- Running time is $O(V + E)$ (using adjacency lists to represent E')
- Let $A \subseteq E$ denote the set of edges picked in line 4

Analysis of Greedy for Vertex Cover

```
APPROX-VERTEX-COVER(G)
1  C = ∅
2  E' = G.E
3  while E' ≠ ∅
4      let (u, v) be an arbitrary edge of E'
5      C = C ∪ {u, v}
6      remove from E' every edge incident on either u or v
7  return C
```

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Proof:

- Running time is $O(V + E)$ (using adjacency lists to represent E')
- Let $A \subseteq E$ denote the set of edges picked in line 4
- Every optimal cover C^* must include at least one endpoint of edges in A,

Analysis of Greedy for Vertex Cover

APPROX-VERTEX-COVER(G)

```
1  C = ∅
2  E' = G.E
3  while E' ≠ ∅
4      let (u, v) be an arbitrary edge of E'
5      C = C ∪ {u, v}
6      remove from E' every edge incident on either u or v
7  return C
```

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Proof:

- Running time is $O(V + E)$ (using adjacency lists to represent E')
- Let $A \subseteq E$ denote the set of edges picked in line 4
- Every optimal cover C^* must include at least one endpoint of edges in A, and edges in A do not share a common endpoint:

Analysis of Greedy for Vertex Cover

```
APPROX-VERTEX-COVER(G)
1  C = ∅
2  E' = G.E
3  while E' ≠ ∅
4      let (u, v) be an arbitrary edge of E'
5      C = C ∪ {u, v}
6      remove from E' every edge incident on either u or v
7  return C
```

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Proof:

- Running time is $O(V + E)$ (using adjacency lists to represent E')
- Let $A \subseteq E$ denote the set of edges picked in line 4
- Every optimal cover C^* must include at least one endpoint of edges in A, and edges in A do not share a common endpoint: $\boxed{|C^*| \geq |A|}$

Analysis of Greedy for Vertex Cover

```
APPROX-VERTEX-COVER(G)
1  C = ∅
2  E′ = G.E
3  while E′ ≠ ∅
4      let (u, v) be an arbitrary edge of E′
5      C = C ∪ {u, v}
6      remove from E′ every edge incident on either u or v
7  return C
```

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Proof:

- Running time is $O(V + E)$ (using adjacency lists to represent E')
- Let $A \subseteq E$ denote the set of edges picked in line 4
- Every optimal cover C^* must include at least one endpoint of edges in A, and edges in A do not share a common endpoint: $\boxed{|C^*| \geq |A|}$
- Every edge in A contributes 2 vertices to $|C|$:

Analysis of Greedy for Vertex Cover

APPROX-VERTEX-COVER(G)

```
1  C = ∅
2  E' = G.E
3  while E' ≠ ∅
4      let (u, v) be an arbitrary edge of E'
5      C = C ∪ {u, v}
6      remove from E' every edge incident on either u or v
7  return C
```

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Proof:

- Running time is $O(V + E)$ (using adjacency lists to represent E')
- Let $A \subseteq E$ denote the set of edges picked in line 4
- Every optimal cover C^* must include at least one endpoint of edges in A, and edges in A do not share a common endpoint: $\boxed{|C^*| \geq |A|}$
- Every edge in A contributes 2 vertices to $|C|$: $\boxed{|C| = 2|A|}$

Analysis of Greedy for Vertex Cover

```
APPROX-VERTEX-COVER(G)
1  C = ∅
2  E′ = G.E
3  while E′ ≠ ∅
4      let (u, v) be an arbitrary edge of E′
5      C = C ∪ {u, v}
6      remove from E′ every edge incident on either u or v
7  return C
```

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Proof:

- Running time is $O(V + E)$ (using adjacency lists to represent E')
- Let $A \subseteq E$ denote the set of edges picked in line 4
- Every optimal cover C^* must include at least one endpoint of edges in A, and edges in A do not share a common endpoint: $\boxed{|C^*| \geq |A|}$
- Every edge in A contributes 2 vertices to $|C|$: $\boxed{|C| = 2|A| \leq 2|C^*|.}$

Analysis of Greedy for Vertex Cover

APPROX-VERTEX-COVER(G)

1 $C = \emptyset$
2 $E' = G.E$
3 **while** $E' \neq \emptyset$
4 let (u, v) be an arbitrary edge of E'
5 $C = C \cup \{u, v\}$
6 remove from E' every edge incident on either u or v
7 **return** C

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Proof:

- Running time is $O(V + E)$ (using adjacency lists to represent E')
- Let $A \subseteq E$ denote the set of edges picked in line 4
- Every optimal cover C^* must include at least one endpoint of edges in A, and edges in A do not share a common endpoint: $\boxed{|C^*| \geq |A|}$
- Every edge in A contributes 2 vertices to $|C|$: $\boxed{|C| = 2|A| \leq 2|C^*|.}$ $\square$

Analysis of Greedy for Vertex Cover

APPROX-VERTEX-COVER(G)

```
1  C = ∅
2  E′ = G.E
3  while E′ ≠ ∅
4      let (u, v) be an arbitrary edge of E′
5      C = C ∪ {u, v}
6      remove from E′
7  return C
```

We can bound the size of the returned solution without knowing the (size of an) optimal solution!

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Proof:

- Running time is $O(V+E)$ (using adjacency lists to represent E')
- Let $A \subseteq E$ denote the set of edges picked in line 4
- Every optimal cover C^* must include at least one endpoint of edges in A, and edges in A do not share a common endpoint: $\boxed{|C^*| \geq |A|}$
- Every edge in A contributes 2 vertices to $|C|$: $\boxed{|C| = 2|A| \leq 2|C^*|.}$ $\square$

Analysis of Greedy for Vertex Cover

APPROX-VERTEX-COVER$(G$

```
1  C = ∅
2  E′ = G.E
3  while E′ ≠ ∅
4      let (u, v) be an arbitrary edge of E′
5      C = C ∪ {u, v}
6      remove from E′
7  return C
```

A "vertex-based" Greedy that adds **one** vertex at each iteration fails to achieve an approximation ratio of 2 (Exercise)!

We can bound the size of the returned solution without knowing the (size of an) optimal solution!

Theorem 35.1

APPROX-VERTEX-COVER is a poly-time 2-approximation algorithm.

Proof:

- Running time is $O(V + E)$ (using adjacency lists to represent E')
- Let $A \subseteq E$ denote the set of edges picked in line 4
- Every optimal cover C^* must include at least one endpoint of edges in A, and edges in A do not share a common endpoint: $\boxed{|C^*| \geq |A|}$
- Every edge in A contributes 2 vertices to $|C|$: $\boxed{|C| = 2|A| \leq 2|C^*|.}$ □

Solving Special Cases

Strategies to cope with NP-complete problems

1. If inputs are small, an algorithm with exponential running time may be satisfactory.
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.

Solving Special Cases

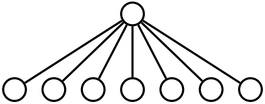
Strategies to cope with NP-complete problems

1. If inputs are small, an algorithm with exponential running time may be satisfactory.
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.

Solving Special Cases

Strategies to cope with NP-complete problems

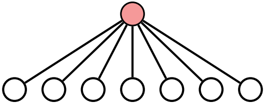
1. If inputs are small, an algorithm with exponential running time may be satisfactory.
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.



Solving Special Cases

Strategies to cope with NP-complete problems

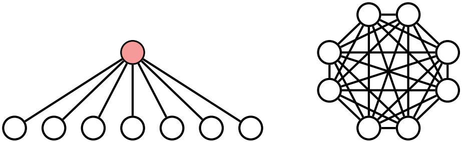
1. If inputs are small, an algorithm with exponential running time may be satisfactory.
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.



Solving Special Cases

Strategies to cope with NP-complete problems

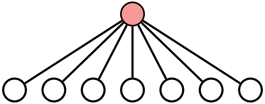
1. If inputs are small, an algorithm with exponential running time may be satisfactory.
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.

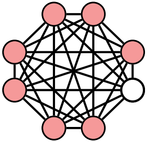


Solving Special Cases

Strategies to cope with NP-complete problems

1. If inputs are small, an algorithm with exponential running time may be satisfactory.
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.

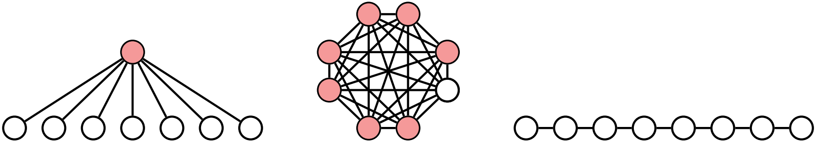




Solving Special Cases

Strategies to cope with NP-complete problems

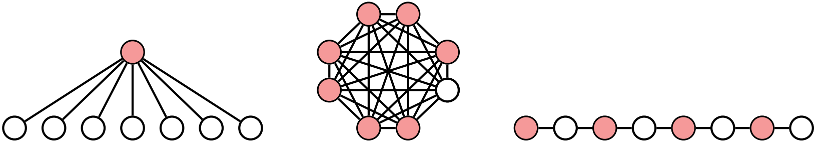
1. If inputs are small, an algorithm with exponential running time may be satisfactory.
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.



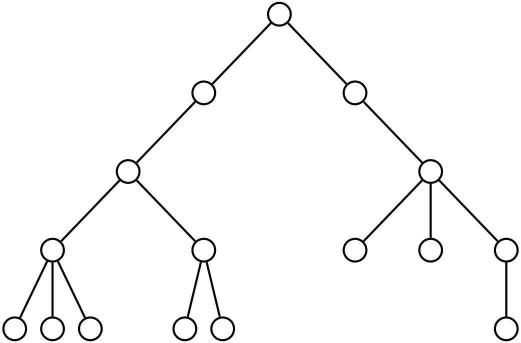
Solving Special Cases

Strategies to cope with NP-complete problems

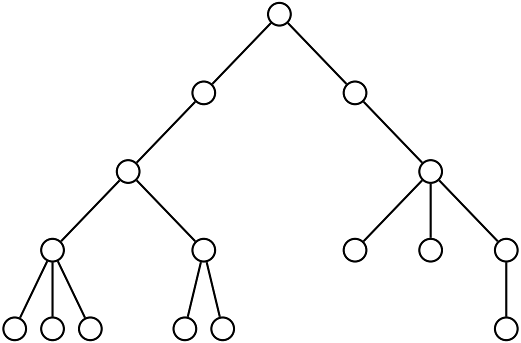
1. If inputs are small, an algorithm with exponential running time may be satisfactory.
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.



Vertex Cover on Trees

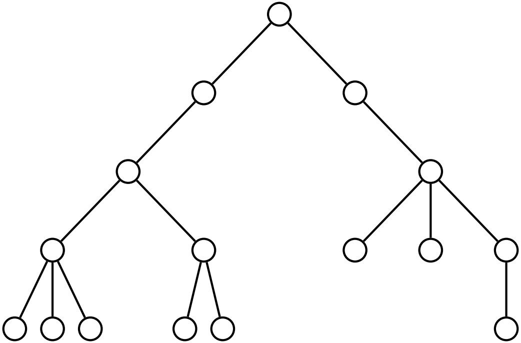


Vertex Cover on Trees



There exists an optimal vertex cover which does not include any leaves.

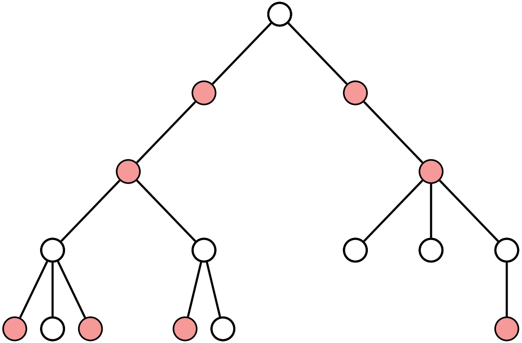
Vertex Cover on Trees



There exists an optimal vertex cover which does not include any leaves.

Exchange-Argument: Replace any leaf in the cover by its parent.

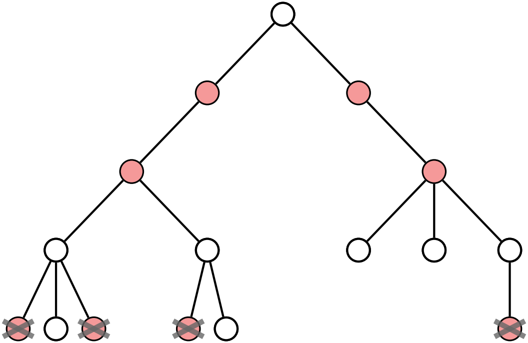
Vertex Cover on Trees



There exists an optimal vertex cover which does not include any leaves.

Exchange-Argument: Replace any leaf in the cover by its parent.

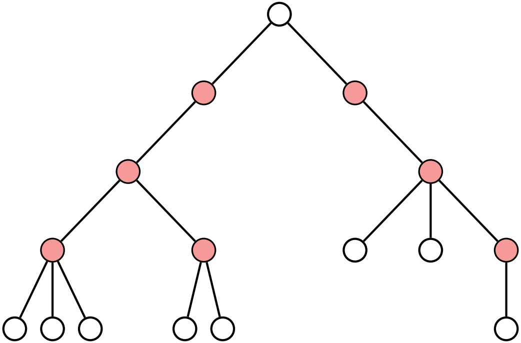
Vertex Cover on Trees



There exists an optimal vertex cover which does not include any leaves.

Exchange-Argument: Replace any leaf in the cover by its parent.

Vertex Cover on Trees



There exists an optimal vertex cover which does not include any leaves.

Exchange-Argument: Replace any leaf in the cover by its parent.

Solving Vertex Cover on Trees

There exists an optimal vertex cover which does not include any leaves.

Solving Vertex Cover on Trees

There exists an optimal vertex cover which does not include any leaves.

VERTEX-COVER-TREES(G)

1: $C = \emptyset$
2: **while** $\exists$ leaves in G
3: Add all parents to C
4: Remove all leaves and their parents from G
5: **return** C

Solving Vertex Cover on Trees

There exists an optimal vertex cover which does not include any leaves.

VERTEX-COVER-TREES(G)

$C = \emptyset$
while $\exists$ leaves in G
    Add all parents to C
    Remove all leaves and their parents from G
return C

Clear: Running time is $O(V)$, and the returned solution is a vertex cover.

Solving Vertex Cover on Trees

There exists an optimal vertex cover which does not include any leaves.

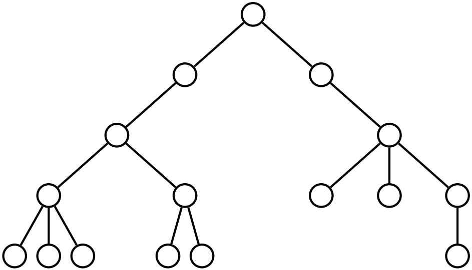
VERTEX-COVER-TREES(G)

1: $C = \emptyset$
2: **while** $\exists$ leaves in G
3: Add all parents to C
4: Remove all leaves and their parents from G
5: **return** C

Clear: Running time is $O(V)$, and the returned solution is a vertex cover.

Solution is also optimal. (Use inductively the existence of an optimal vertex cover without leaves)

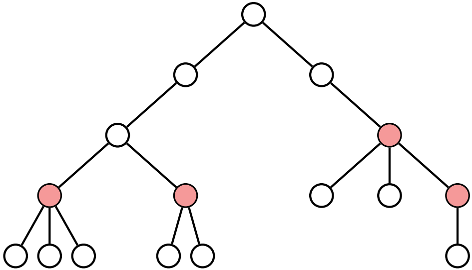
Execution on a Small Example



VERTEX-COVER-TREES(G)

1: $C = \emptyset$
2: **while** $\exists$ leaves in G
3: Add all parents to C
4: Remove all leaves and their parents from G
5: **return** C

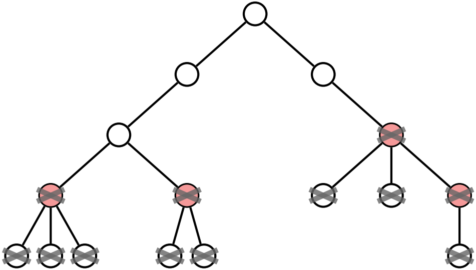
Execution on a Small Example



VERTEX-COVER-TREES(G)

1: $C = \emptyset$
2: **while** $\exists$ leaves in G
3: Add all parents to C
4: Remove all leaves and their parents from G
5: **return** C

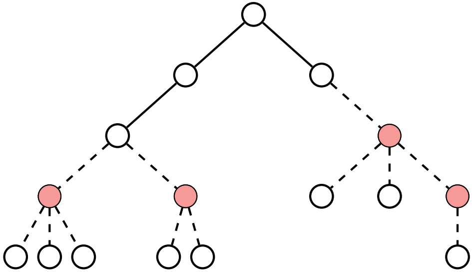
Execution on a Small Example



VERTEX-COVER-TREES(G)

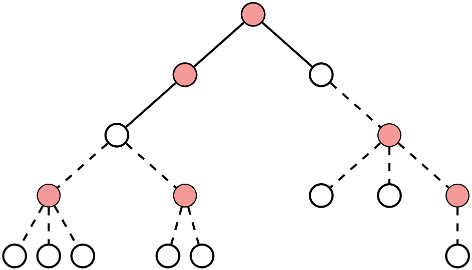
$C = \emptyset$
while $\exists$ leaves in G
    Add all parents to C
    Remove all leaves and their parents from G
return C

Execution on a Small Example



```
VERTEX-COVER-TREES(G)
C = ∅
while ∃ leaves in G
    Add all parents to C
    Remove all leaves and their parents from G
return C
```

Execution on a Small Example



VERTEX-COVER-TREES(G)

1: $C = \emptyset$

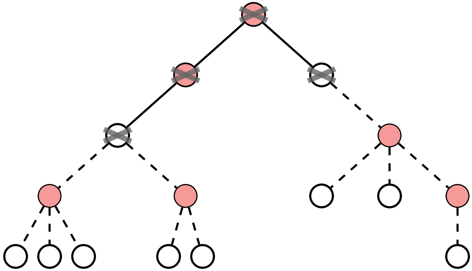
2: **while** $\exists$ leaves in G

3: Add all parents to C

4: Remove all leaves and their parents from G

5: **return** C

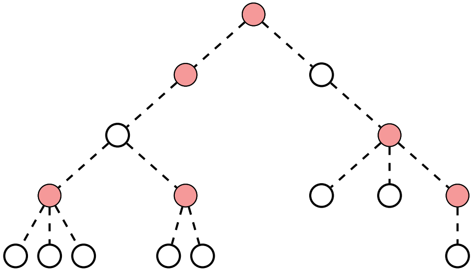
Execution on a Small Example



VERTEX-COVER-TREES(G)

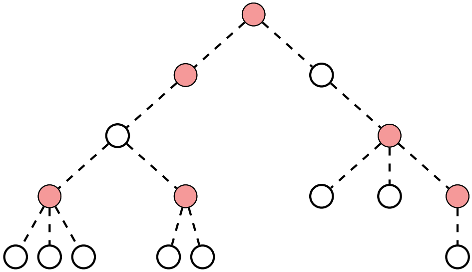
$C = \emptyset$
while $\exists$ leaves in G
    Add all parents to C
    Remove all leaves and their parents from G
return C

Execution on a Small Example



```
VERTEX-COVER-TREES(G)
C = ∅
while ∃ leaves in G
    Add all parents to C
    Remove all leaves and their parents from G
return C
```

Execution on a Small Example



VERTEX-COVER-TREES(G)

```
C = ∅
while ∃ leaves in G
    Add all parents to C
    Remove all leaves and their parents from G
return C
```

Problem can be also solved on bipartite graphs, using Max-Flows and Min-Cuts.

Exact Algorithms

Strategies to cope with NP-complete problems

1. If inputs (or solutions) are small, an algorithm with exponential running time may be satisfactory
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.

Exact Algorithms

Strategies to cope with NP-complete problems

1. If inputs (or solutions) are small, an algorithm with exponential running time may be satisfactory
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.

Exact Algorithms

Such algorithms are called exact algorithms.

Strategies to cope with NP-complete problems

1. If inputs (or solutions) are small, an algorithm with exponential running time may be satisfactory
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.

Exact Algorithms

Such algorithms are called exact algorithms.

Strategies to cope with NP-complete problems

1. If inputs (or solutions) are small, an algorithm with exponential running time may be satisfactory
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.

Focus on instances where the minimum vertex cover is small, that is, **less or equal** than some given integer k.

Exact Algorithms

Such algorithms are called exact algorithms.

Strategies to cope with NP-complete problems

1. If inputs (or solutions) are small, an algorithm with exponential running time may be satisfactory
2. Isolate important special cases which can be solved in polynomial-time.
3. Develop algorithms which find near-optimal solutions in polynomial-time.

Focus on instances where the minimum vertex cover is small, that is, **less or equal** than some given integer k.

Simple Brute-Force Search would take $\approx \binom{n}{k} = \Theta(n^k)$ time.

Towards a more efficient Search

Substructure Lemma

Consider a graph $G = (V, E)$, edge $\{u, v\} \in E(G)$ and integer $k \geq 1$. Let G_u be the graph obtained by deleting u and its incident edges (G_v is defined similarly). Then G has a vertex cover of size k if and only if G_u or G_v (or both) have a vertex cover of size $k-1$.

Towards a more efficient Search

Substructure Lemma

Consider a graph $G = (V, E)$, edge $\{u, v\} \in E(G)$ and integer $k \geq 1$. Let G_u be the graph obtained by deleting u and its incident edges (G_v is defined similarly). Then G has a vertex cover of size k if and only if G_u or G_v (or both) have a vertex cover of size $k - 1$.

Reminiscent of Dynamic Programming.

Towards a more efficient Search

Substructure Lemma

Consider a graph $G = (V, E)$, edge $\{u, v\} \in E(G)$ and integer $k \geq 1$. Let G_u be the graph obtained by deleting u and its incident edges (G_v is defined similarly). Then G has a vertex cover of size k if and only if G_u or G_v (or both) have a vertex cover of size $k - 1$.

Proof:

$\Leftarrow$ Assume G_u has a vertex cover C_u of size $k - 1$.

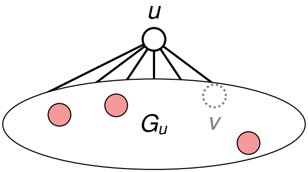
Towards a more efficient Search

Substructure Lemma

Consider a graph $G = (V, E)$, edge $\{u, v\} \in E(G)$ and integer $k \geq 1$. Let G_u be the graph obtained by deleting u and its incident edges (G_v is defined similarly). Then G has a vertex cover of size k if and only if G_u or G_v (or both) have a vertex cover of size $k - 1$.

Proof:

$\Leftarrow$ Assume G_u has a vertex cover C_u of size $k - 1$.



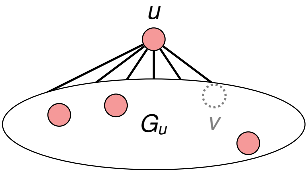
Towards a more efficient Search

Substructure Lemma

Consider a graph $G = (V, E)$, edge $\{u, v\} \in E(G)$ and integer $k \geq 1$. Let G_u be the graph obtained by deleting u and its incident edges (G_v is defined similarly). Then G has a vertex cover of size k if and only if G_u or G_v (or both) have a vertex cover of size $k-1$.

Proof:

$\Leftarrow$ Assume G_u has a vertex cover C_u of size $k-1$.
Adding u yields a vertex cover of G which is of size k



Towards a more efficient Search

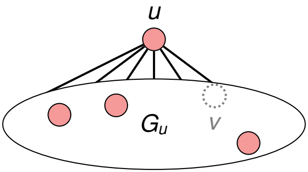
Substructure Lemma

Consider a graph $G = (V, E)$, edge $\{u, v\} \in E(G)$ and integer $k \geq 1$. Let G_u be the graph obtained by deleting u and its incident edges (G_v is defined similarly). Then G has a vertex cover of size k if and only if G_u or G_v (or both) have a vertex cover of size $k - 1$.

Proof:

$\Leftarrow$ Assume G_u has a vertex cover C_u of size $k - 1$.
Adding u yields a vertex cover of G which is of size k

$\Rightarrow$ Assume G has a vertex cover C of size k, which contains, say u.



Towards a more efficient Search

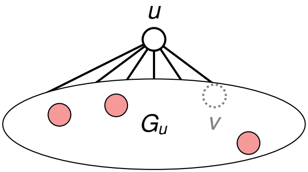
Substructure Lemma

Consider a graph $G = (V, E)$, edge $\{u, v\} \in E(G)$ and integer $k \geq 1$. Let G_u be the graph obtained by deleting u and its incident edges (G_v is defined similarly). Then G has a vertex cover of size k if and only if G_u or G_v (or both) have a vertex cover of size $k - 1$.

Proof:

$\Leftarrow$ Assume G_u has a vertex cover C_u of size $k - 1$.
Adding u yields a vertex cover of G which is of size k

$\Rightarrow$ Assume G has a vertex cover C of size k, which contains, say u.
Removing u from C yields a vertex cover of G_u which is of size $k - 1$. $\square$



A More Efficient Search Algorithm

```
VERTEX-COVER-SEARCH(G, k)
If E = ∅ return ∅
If k = 0 and E ≠ ∅ return ⊥
Pick an arbitrary edge (u, v) ∈ E
S_1 = VERTEX-COVER-SEARCH(G_u, k − 1)
S_2 = VERTEX-COVER-SEARCH(G_v, k − 1)
if S_1 ≠ ⊥ return S_1 ∪ {u}
if S_2 ≠ ⊥ return S_2 ∪ {v}
return ⊥
```

A More Efficient Search Algorithm

VERTEX-COVER-SEARCH(G, k)

If $E = \emptyset$ **return** $\emptyset$
If $k = 0$ and $E \neq \emptyset$ **return** $\perp$
Pick an arbitrary edge $(u, v) \in E$
$S_1 =$ VERTEX-COVER-SEARCH$(G_u, k - 1)$
$S_2 =$ VERTEX-COVER-SEARCH$(G_v, k - 1)$
if $S_1 \neq \perp$ **return** $S_1 \cup \{u\}$
if $S_2 \neq \perp$ **return** $S_2 \cup \{v\}$
return $\perp$

Correctness follows by the Substructure Lemma and induction.

A More Efficient Search Algorithm

$\textsc{Vertex-Cover-Search}(G, k)$

If $E = \emptyset$ **return** $\emptyset$
If $k = 0$ and $E \neq \emptyset$ **return** $\perp$
Pick an arbitrary edge $(u, v) \in E$
$S_1 = \textsc{Vertex-Cover-Search}(G_u, k-1)$
$S_2 = \textsc{Vertex-Cover-Search}(G_v, k-1)$
if $S_1 \neq \perp$ **return** $S_1 \cup \{u\}$
if $S_2 \neq \perp$ **return** $S_2 \cup \{v\}$
return $\perp$

Running time:

A More Efficient Search Algorithm

VERTEX-COVER-SEARCH(G, k)

If $E = \emptyset$ **return** $\emptyset$
If $k = 0$ and $E \neq \emptyset$ **return** $\perp$
Pick an arbitrary edge $(u, v) \in E$
$S_1 =$ VERTEX-COVER-SEARCH$(G_u, k - 1)$
$S_2 =$ VERTEX-COVER-SEARCH$(G_v, k - 1)$
if $S_1 \neq \perp$ **return** $S_1 \cup \{u\}$
if $S_2 \neq \perp$ **return** $S_2 \cup \{v\}$
return $\perp$

Running time:

- Depth k, branching factor 2

A More Efficient Search Algorithm

VERTEX-COVER-SEARCH(G, k)

If $E = \emptyset$ **return** $\emptyset$
If $k = 0$ and $E \neq \emptyset$ **return** $\perp$
Pick an arbitrary edge $(u, v) \in E$
$S_1 =$ VERTEX-COVER-SEARCH$(G_u, k-1)$
$S_2 =$ VERTEX-COVER-SEARCH$(G_v, k-1)$
if $S_1 \neq \perp$ **return** $S_1 \cup \{u\}$
if $S_2 \neq \perp$ **return** $S_2 \cup \{v\}$
return $\perp$

Running time:

- Depth k, branching factor 2 $\Rightarrow$ total number of calls is $O(2^k)$

A More Efficient Search Algorithm

VERTEX-COVER-SEARCH(G, k)

If $E = \emptyset$ **return** $\emptyset$
If $k = 0$ and $E \neq \emptyset$ **return** $\perp$
Pick an arbitrary edge $(u, v) \in E$
$S_1 =$ VERTEX-COVER-SEARCH$(G_u, k-1)$
$S_2 =$ VERTEX-COVER-SEARCH$(G_v, k-1)$
if $S_1 \neq \perp$ **return** $S_1 \cup \{u\}$
if $S_2 \neq \perp$ **return** $S_2 \cup \{v\}$
return $\perp$

Running time:

- Depth k, branching factor 2 $\Rightarrow$ total number of calls is $O(2^k)$
- $O(E)$ work per recursive call

A More Efficient Search Algorithm

$\textsc{Vertex-Cover-Search}(G, k)$

If $E = \emptyset$ **return** $\emptyset$
If $k = 0$ and $E \neq \emptyset$ **return** $\perp$
Pick an arbitrary edge $(u, v) \in E$
$S_1 = \textsc{Vertex-Cover-Search}(G_u, k-1)$
$S_2 = \textsc{Vertex-Cover-Search}(G_v, k-1)$
if $S_1 \neq \perp$ **return** $S_1 \cup \{u\}$
if $S_2 \neq \perp$ **return** $S_2 \cup \{v\}$
return $\perp$

Running time:

- Depth k, branching factor 2 $\Rightarrow$ total number of calls is $O(2^k)$
- $O(E)$ work per recursive call
- Total runtime: $O(2^k \cdot E)$.

A More Efficient Search Algorithm

VERTEX-COVER-SEARCH(G, k)

If $E = \emptyset$ **return** $\emptyset$
If $k = 0$ and $E \neq \emptyset$ **return** $\perp$
Pick an arbitrary edge $(u, v) \in E$
$S_1 =$ VERTEX-COVER-SEARCH$(G_u, k - 1)$
$S_2 =$ VERTEX-COVER-SEARCH$(G_v, k - 1)$
if $S_1 \neq \perp$ **return** $S_1 \cup \{u\}$
if $S_2 \neq \perp$ **return** $S_2 \cup \{v\}$
return $\perp$

Running time:

- Depth k, branching factor 2 $\Rightarrow$ total number of calls is $O(2^k)$
- $O(E)$ work per recursive call
- Total runtime: $O(2^k \cdot E)$.

exponential in k, but much better than $\Theta(n^k)$ (i.e., still polynomial for $k = O(\log n)$)

Outline

Introduction

Vertex Cover

The Set-Covering Problem

The Set-Covering Problem

Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

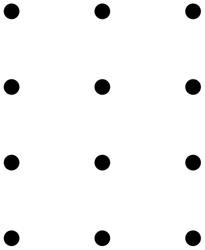
$$\text{s.t.} \qquad X = \bigcup_{S \in \mathcal{C}} S.$$

The Set-Covering Problem

Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

$$\text{s.t.} \qquad X = \bigcup_{S \in \mathcal{C}} S.$$

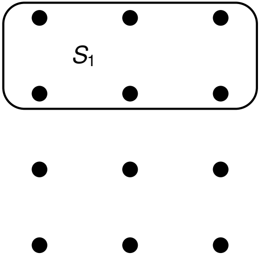


The Set-Covering Problem

Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

$$\text{s.t.} \qquad X = \bigcup_{S \in \mathcal{C}} S.$$

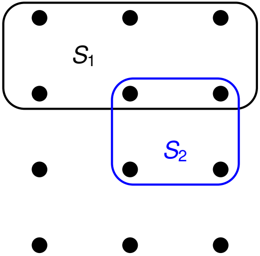


The Set-Covering Problem

Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

$$\text{s.t.} \qquad X = \bigcup_{S \in \mathcal{C}} S.$$

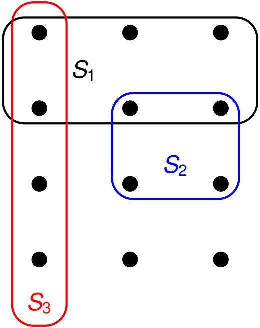


The Set-Covering Problem

Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

$$\text{s.t.} \qquad X = \bigcup_{S \in \mathcal{C}} S.$$

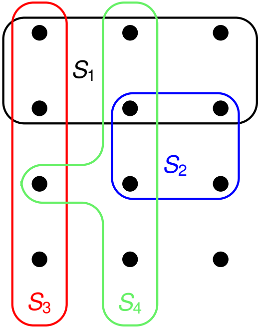


The Set-Covering Problem

Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

$$\text{s.t.} \qquad X = \bigcup_{S \in \mathcal{C}} S.$$

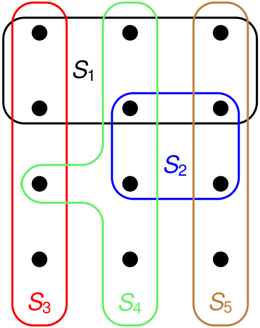


The Set-Covering Problem

Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

$$\text{s.t.} \qquad X = \bigcup_{S \in \mathcal{C}} S.$$

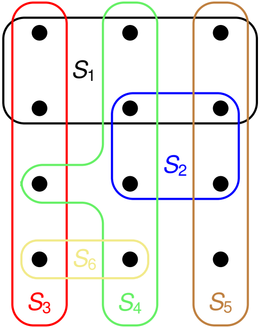


The Set-Covering Problem

Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

$$\text{s.t.} \qquad X = \bigcup_{S \in \mathcal{C}} S.$$



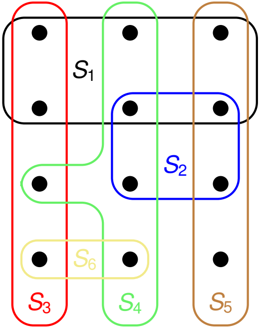
The Set-Covering Problem

Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

$$\text{s.t.} \qquad X = \bigcup_{S \in \mathcal{C}} S.$$

Only solvable if $\bigcup_{S \in \mathcal{F}} S = X$!



The Set-Covering Problem

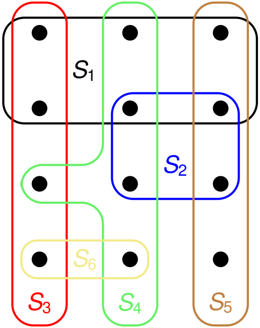
Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

$$\text{s.t.} \qquad X = \bigcup_{S \in \mathcal{C}} S.$$

Number of sets (and not elements)

Only solvable if $\bigcup_{S \in \mathcal{F}} S = X$!



The Set-Covering Problem

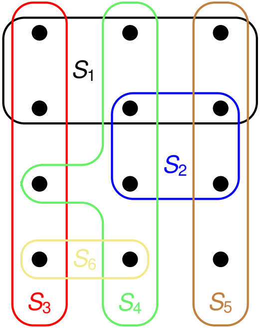
Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

s.t. $$X = \bigcup_{S \in \mathcal{C}} S.$$

Number of sets (and not elements)

Only solvable if $\bigcup_{S \in \mathcal{F}} S = X$!



Remarks:

The Set-Covering Problem

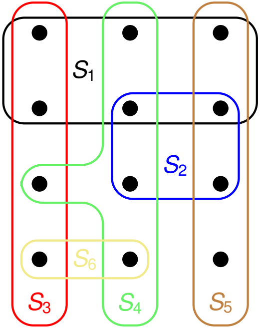
Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

s.t. $$X = \bigcup_{S \in \mathcal{C}} S.$$

Number of sets (and not elements)

Only solvable if $\bigcup_{S \in \mathcal{F}} S = X$!



Remarks:

- generalisation of the vertex-cover problem and hence also NP-hard.

The Set-Covering Problem

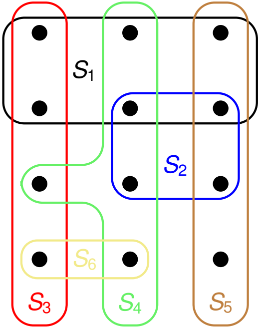
Set Cover Problem

- Given: set X of size n and family of subsets $\mathcal{F}$
- Goal: Find a minimum-size subset $\mathcal{C} \subseteq \mathcal{F}$

$$\text{s.t.} \qquad X = \bigcup_{S \in \mathcal{C}} S.$$

Number of sets (and not elements)

Only solvable if $\bigcup_{S \in \mathcal{F}} S = X$!



Remarks:

- generalisation of the vertex-cover problem and hence also NP-hard.
- models resource allocation problems

Greedy

Strategy: Pick the set S that covers the largest number of uncovered elements.

Greedy

Strategy: Pick the set S that covers the largest number of uncovered elements.

GREEDY-SET-COVER$(X, \mathcal{F})$

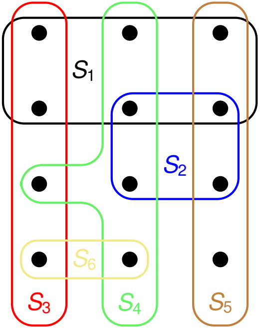
```
U = X
𝒞 = ∅
while U ≠ ∅
      select an S ∈ ℱ that maximizes |S ∩ U|
      U = U − S
      𝒞 = 𝒞 ∪ {S}
return 𝒞
```

Greedy

Strategy: Pick the set S that covers the largest number of uncovered elements.

GREEDY-SET-COVER$(X, \mathcal{F})$

$U = X$
$\mathcal{C} = \emptyset$
while $U \neq \emptyset$
    select an $S \in \mathcal{F}$ that maximizes $|S \cap U|$
    $U = U - S$
    $\mathcal{C} = \mathcal{C} \cup \{S\}$
return $\mathcal{C}$

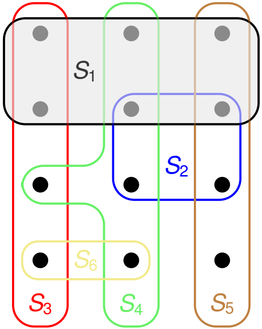


Greedy

Strategy: Pick the set S that covers the largest number of uncovered elements.

GREEDY-SET-COVER$(X, \mathcal{F})$

$U = X$
$\mathcal{C} = \emptyset$
while $U \neq \emptyset$
    select an $S \in \mathcal{F}$ that maximizes $|S \cap U|$
    $U = U - S$
    $\mathcal{C} = \mathcal{C} \cup \{S\}$
return $\mathcal{C}$

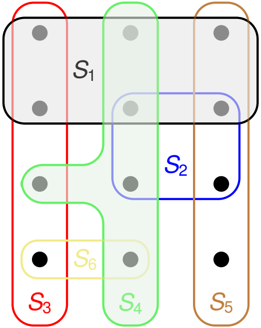


Greedy

Strategy: Pick the set S that covers the largest number of uncovered elements.

GREEDY-SET-COVER$(X, \mathcal{F})$

$U = X$
$\mathcal{C} = \emptyset$
while $U \neq \emptyset$
  select an $S \in \mathcal{F}$ that maximizes $|S \cap U|$
  $U = U - S$
  $\mathcal{C} = \mathcal{C} \cup \{S\}$
return $\mathcal{C}$

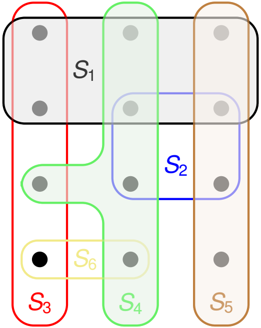


Greedy

Strategy: Pick the set S that covers the largest number of uncovered elements.

GREEDY-SET-COVER$(X, \mathcal{F})$

$U = X$
$\mathcal{C} = \emptyset$
while $U \neq \emptyset$
    select an $S \in \mathcal{F}$ that maximizes $|S \cap U|$
    $U = U - S$
    $\mathcal{C} = \mathcal{C} \cup \{S\}$
return $\mathcal{C}$

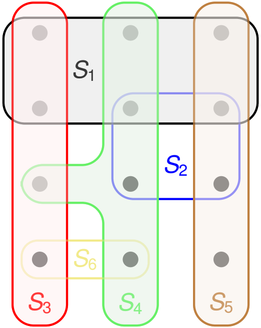


Greedy

Strategy: Pick the set S that covers the largest number of uncovered elements.

GREEDY-SET-COVER$(X, \mathcal{F})$

$U = X$
$\mathcal{C} = \emptyset$
while $U \neq \emptyset$
    select an $S \in \mathcal{F}$ that maximizes $|S \cap U|$
    $U = U - S$
    $\mathcal{C} = \mathcal{C} \cup \{S\}$
return $\mathcal{C}$

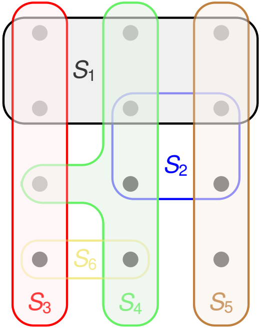


Greedy

Strategy: Pick the set S that covers the largest number of uncovered elements.

GREEDY-SET-COVER$(X, \mathcal{F})$

```
U = X
𝒞 = ∅
while U ≠ ∅
    select an S ∈ 𝓕 that maximizes |S ∩ U|
    U = U − S
    𝒞 = 𝒞 ∪ {S}
return 𝒞
```



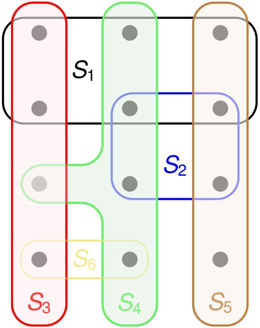
Greedy chooses S_1, S_4, S_5 and S_3 (or S_6), which is a cover of size 4.

Greedy

Strategy: Pick the set S that covers the largest number of uncovered elements.

GREEDY-SET-COVER$(X, \mathcal{F})$

$U = X$
$\mathcal{C} = \emptyset$
while $U \neq \emptyset$
  select an $S \in \mathcal{F}$ that maximizes $|S \cap U|$
  $U = U - S$
  $\mathcal{C} = \mathcal{C} \cup \{S\}$
return $\mathcal{C}$



Greedy chooses S_1, S_4, S_5 and S_3 (or S_6), which is a cover of size 4.

Optimal cover is $\mathcal{C} = \{S_3, S_4, S_5\}$

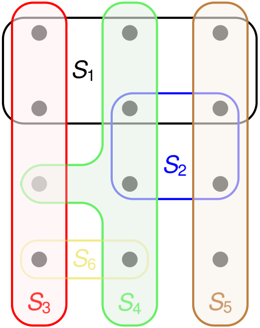
Greedy

Strategy: Pick the set S that covers the largest number of uncovered elements.

GREEDY-SET-COVER$(X, \mathcal{F})$

$U = X$
$\mathcal{C} = \emptyset$
while $U \neq \emptyset$
    select an $S \in \mathcal{F}$ that maximizes $|S \cap U|$
    $U = U - S$
    $\mathcal{C} = \mathcal{C} \cup \{S\}$
return $\mathcal{C}$

Can be easily implemented to run in time polynomial in $|X|$ and $|\mathcal{F}|$



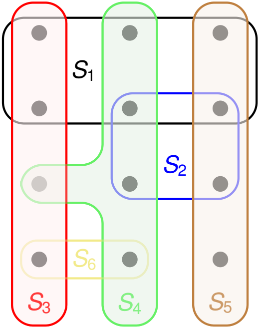
Greedy

Strategy: Pick the set S that covers the largest number of uncovered elements.

GREEDY-SET-COVER$(X, \mathcal{F})$

```
U = X
𝒞 = ∅
while U ≠ ∅
    select an S ∈ 𝓕 that maximizes |S ∩ U|
    U = U − S
    𝒞 = 𝒞 ∪ {S}
return 𝒞
```

Can be easily implemented to run in time polynomial in $|X|$ and $|\mathcal{F}|$



How good is the approximation ratio?

Approximation Ratio of Greedy

Theorem 35.4

GREEDY-SET-COVER is a polynomial-time $\rho(n)$-algorithm, where

$$\rho(n) = H(\max\{|S| \colon |S| \in \mathcal{F}\})$$

Approximation Ratio of Greedy

Theorem 35.4

GREEDY-SET-COVER is a polynomial-time $\rho(n)$-algorithm, where

$$\rho(n) = H(\max\{|S| \colon |S| \in \mathcal{F}\})$$

$H(k) := \sum_{i=1}^{k} \frac{1}{i} \leq \ln(k) + 1$

Approximation Ratio of Greedy

Theorem 35.4

GREEDY-SET-COVER is a polynomial-time $\rho(n)$-algorithm, where

$$\rho(n) = H(\max\{|S| : |S| \in \mathcal{F}\}) \leq \ln(n) + 1.$$

$H(k) := \sum_{i=1}^{k} \frac{1}{i} \leq \ln(k) + 1$

Approximation Ratio of Greedy

Theorem 35.4

GREEDY-SET-COVER is a polynomial-time $\rho(n)$-algorithm, where

$$\rho(n) = H(\max\{|S| : |S| \in \mathcal{F}\}) \leq \ln(n) + 1.$$

$H(k) := \sum_{i=1}^{k} \frac{1}{i} \leq \ln(k) + 1$

Idea: Distribute cost of 1 for each added set over newly covered elements.

Approximation Ratio of Greedy

Theorem 35.4

GREEDY-SET-COVER is a polynomial-time $\rho(n)$-algorithm, where

$$\rho(n) = H(\max\{|S| : |S| \in \mathcal{F}\}) \leq \ln(n) + 1.$$

$H(k) := \sum_{i=1}^{k} \frac{1}{i} \leq \ln(k) + 1$

Idea: Distribute cost of 1 for each added set over newly covered elements.

Definition of cost

If an element x is covered for the first time by set S_i in iteration i, then

$$c_x := \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}.$$

Approximation Ratio of Greedy

Theorem 35.4

GREEDY-SET-COVER is a polynomial-time $\rho(n)$-algorithm, where

$$\rho(n) = H(\max\{|S| : |S| \in \mathcal{F}\}) \leq \ln(n) + 1.$$

$H(k) := \sum_{i=1}^{k} \frac{1}{i} \leq \ln(k) + 1$

Idea: Distribute cost of 1 for each added set over newly covered elements.

Definition of cost

If an element x is covered for the first time by set S_i in iteration i, then

$$c_x := \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}.$$

Notice that in the mathematical analysis, S_i is the set chosen in iteration i - not to be confused with the sets $S_1, S_2, \ldots, S_6$ in the example.

Illustration of Costs for Greedy picking S_1, S_4, S_5 and S_3

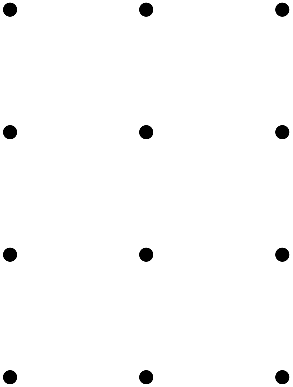


Illustration of Costs for Greedy picking S_1, S_4, S_5 and S_3

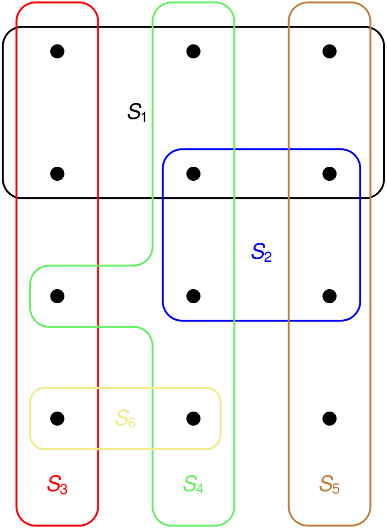


Illustration of Costs for Greedy picking S_1, S_4, S_5 and S_3

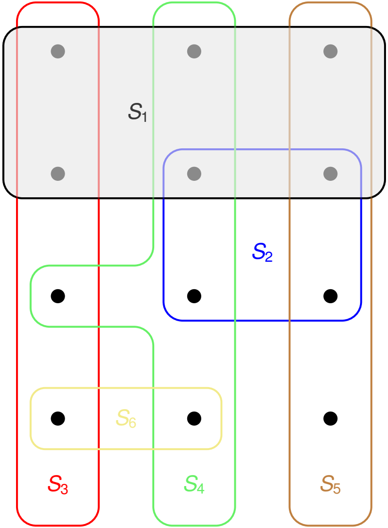


Illustration of Costs for Greedy picking S_1, S_4, S_5 and S_3

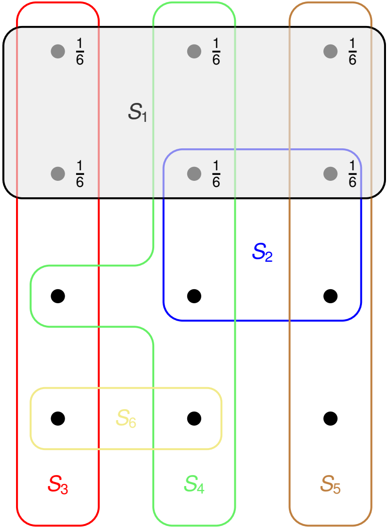


Illustration of Costs for Greedy picking S_1, S_4, S_5 and S_3

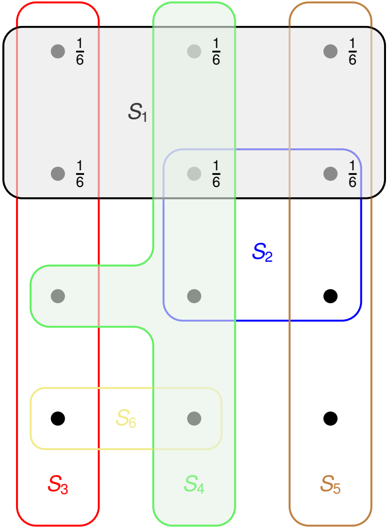


Illustration of Costs for Greedy picking S_1, S_4, S_5 and S_3

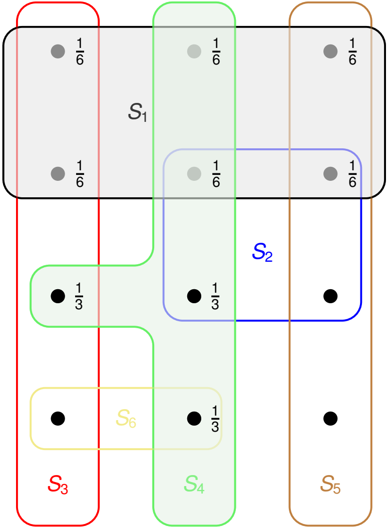


Illustration of Costs for Greedy picking S_1, S_4, S_5 and S_3

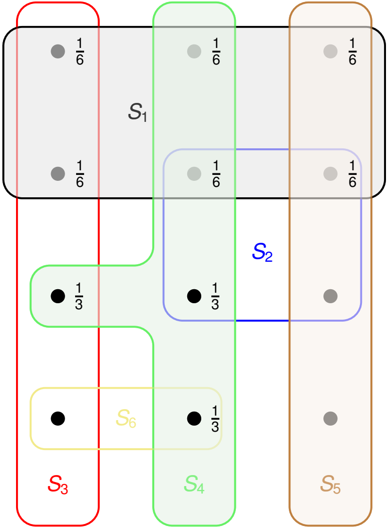


Illustration of Costs for Greedy picking S_1, S_4, S_5 and S_3

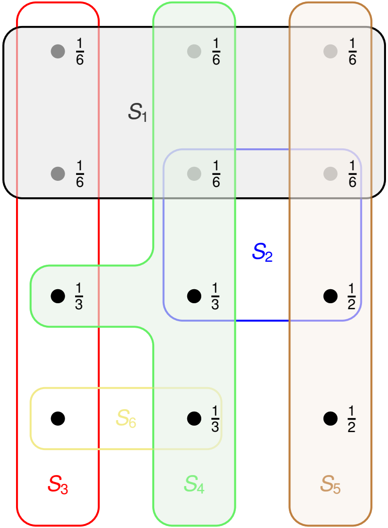


Illustration of Costs for Greedy picking S_1, S_4, S_5 and S_3

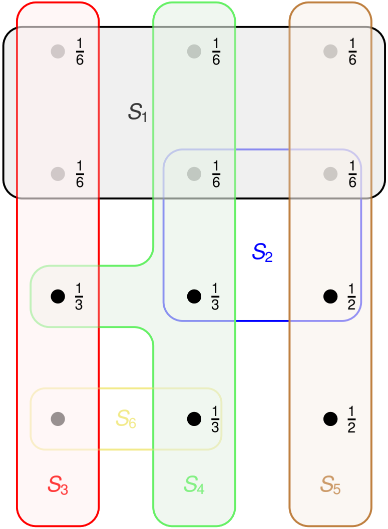


Illustration of Costs for Greedy picking S_1, S_4, S_5 and S_3

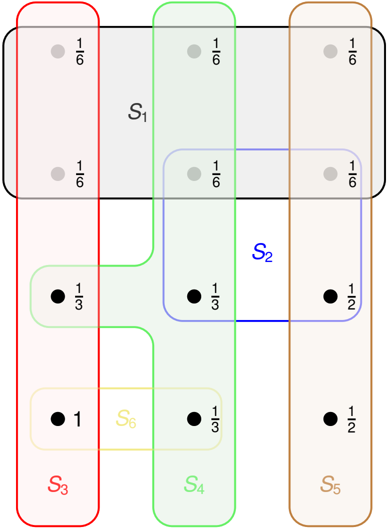
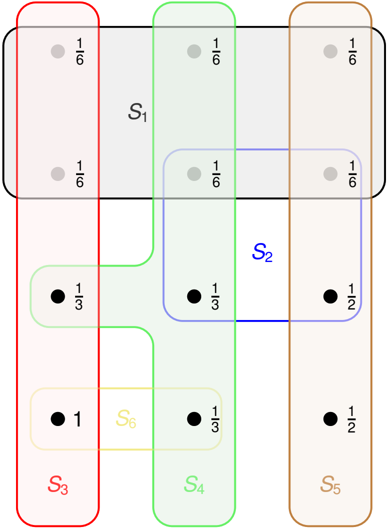
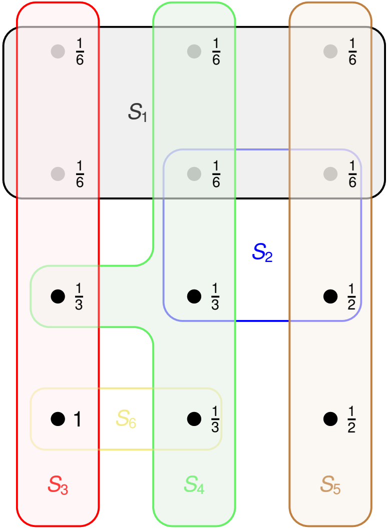


Illustration of Costs for Greedy picking S_1, S_4, S_5 and S_3



$$\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{3}+\frac{1}{3}+\frac{1}{3}+\frac{1}{2}+\frac{1}{2}+1 = ??$$

Illustration of Costs for Greedy picking S_1, S_4, S_5 and S_3



$$\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{3}+\frac{1}{3}+\frac{1}{3}+\frac{1}{2}+\frac{1}{2}+1 = 4$$

Proof of Theorem 35.4 (1/2)

Definition of cost

If x is covered for the first time by a set S_i, then $c_x := \frac{1}{\left|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})\right|}$.

Proof of Theorem 35.4 (1/2)

Definition of cost

If x is covered for the first time by a set S_i, then $c_x := \frac{1}{\left|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})\right|}$.

Proof.

- Each step of the algorithm assigns one unit of cost, so

(1)

Proof of Theorem 35.4 (1/2)

Definition of cost

If x is covered for the first time by a set S_i, then $c_x := \frac{1}{\left|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})\right|}$.

Proof.

- Each step of the algorithm assigns one unit of cost, so

$$|\mathcal{C}| = \sum_{x \in X} c_x \tag{1}$$

Proof of Theorem 35.4 (1/2)

Definition of cost

If x is covered for the first time by a set S_i, then $c_x := \frac{1}{\left|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})\right|}$.

Proof.

- Each step of the algorithm assigns one unit of cost, so

$$|\mathcal{C}| = \sum_{x \in X} c_x \tag{1}$$

- Each element $x \in X$ is in at least one set in the optimal cover $\mathcal{C}^*$, so

Proof of Theorem 35.4 (1/2)

Definition of cost

If x is covered for the first time by a set S_i, then $c_x := \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$.

Proof.

- Each step of the algorithm assigns one unit of cost, so

$$|\mathcal{C}| = \sum_{x \in X} c_x \tag{1}$$

- Each element $x \in X$ is in at least one set in the optimal cover $\mathcal{C}^*$, so

$$\sum_{S \in \mathcal{C}^*} \sum_{x \in S} c_x \geq \sum_{x \in X} c_x \tag{2}$$

Proof of Theorem 35.4 (1/2)

Definition of cost

If x is covered for the first time by a set S_i, then $c_x := \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$.

Proof.

- Each step of the algorithm assigns one unit of cost, so

$$|\mathcal{C}| = \sum_{x \in X} c_x \tag{1}$$

- Each element $x \in X$ is in at least one set in the optimal cover $\mathcal{C}^*$, so

$$\sum_{S \in \mathcal{C}^*} \sum_{x \in S} c_x \geq \sum_{x \in X} c_x \tag{2}$$

- Combining 1 and 2 gives

Proof of Theorem 35.4 (1/2)

Definition of cost

If x is covered for the first time by a set S_i, then $c_x := \frac{1}{\left|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})\right|}$.

Proof.

- Each step of the algorithm assigns one unit of cost, so

$$|\mathcal{C}| = \sum_{x \in X} c_x \tag{1}$$

- Each element $x \in X$ is in at least one set in the optimal cover $\mathcal{C}^*$, so

$$\sum_{S \in \mathcal{C}^*} \sum_{x \in S} c_x \geq \sum_{x \in X} c_x \tag{2}$$

- Combining 1 and 2 gives

$$|\mathcal{C}| \leq \sum_{S \in \mathcal{C}^*} \sum_{x \in S} c_x$$

Proof of Theorem 35.4 (1/2)

Definition of cost

If x is covered for the first time by a set S_i, then $c_x := \frac{1}{\left|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})\right|}$.

Proof.

- Each step of the algorithm assigns one unit of cost, so

$$|\mathcal{C}| = \sum_{x \in X} c_x \tag{1}$$

- Each element $x \in X$ is in at least one set in the optimal cover $\mathcal{C}^*$, so

$$\sum_{S \in \mathcal{C}^*} \sum_{x \in S} c_x \geq \sum_{x \in X} c_x \tag{2}$$

- Combining 1 and 2 gives

$$|\mathcal{C}| \leq \sum_{S \in \mathcal{C}^*} \sum_{x \in S} c_x$$

Key Inequality: $\sum_{x \in S} c_x \leq H(|S|)$.

Proof of Theorem 35.4 (1/2)

Definition of cost

If x is covered for the first time by a set S_i, then $c_x := \frac{1}{\left|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})\right|}$.

Proof.

- Each step of the algorithm assigns one unit of cost, so

$$|\mathcal{C}| = \sum_{x \in X} c_x \tag{1}$$

- Each element $x \in X$ is in at least one set in the optimal cover $\mathcal{C}^*$, so

$$\sum_{S \in \mathcal{C}^*} \sum_{x \in S} c_x \geq \sum_{x \in X} c_x \tag{2}$$

- Combining 1 and 2 gives

$$|\mathcal{C}| \leq \sum_{S \in \mathcal{C}^*} \sum_{x \in S} c_x \leq \sum_{S \in \mathcal{C}^*} H(|S|)$$

Key Inequality: $\sum_{x \in S} c_x \leq H(|S|)$.

Proof of Theorem 35.4 (1/2)

Definition of cost

If x is covered for the first time by a set S_i, then $c_x := \frac{1}{\left|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})\right|}$.

Proof.

- Each step of the algorithm assigns one unit of cost, so

$$|\mathcal{C}| = \sum_{x \in X} c_x \tag{1}$$

- Each element $x \in X$ is in at least one set in the optimal cover $\mathcal{C}^*$, so

$$\sum_{S \in \mathcal{C}^*} \sum_{x \in S} c_x \geq \sum_{x \in X} c_x \tag{2}$$

- Combining 1 and 2 gives

$$|\mathcal{C}| \leq \sum_{S \in \mathcal{C}^*} \sum_{x \in S} c_x \leq \sum_{S \in \mathcal{C}^*} H(|S|) \leq |\mathcal{C}^*| \cdot H(\max\{|S| : S \in \mathcal{F}\}) \qquad \square$$

Key Inequality: $\sum_{x \in S} c_x \leq H(|S|)$.

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

Remaining uncovered elements in S

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

Sets chosen by the algorithm

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x \in S} c_x$$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x \in S} c_x = \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x \in S} c_x = \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x \in S} c_x = \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$$

- Further, by definition of the Greedy-Set-Cover:

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x \in S} c_x = \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$$

- Further, by definition of the Greedy-Set-Cover:

$$|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| \geq |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|$$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x \in S} c_x = \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$$

- Further, by definition of the Greedy-Set-Cover:

$$|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| \geq |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| = u_{i-1}.$$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x\in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x\in S} c_x = \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$$

- Further, by definition of the GREEDY-SET-COVER:

$$|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| \geq |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| = u_{i-1}.$$

- Combining the last inequalities gives:

$$\sum_{x\in S} c_x$$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x \in S} c_x = \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$$

- Further, by definition of the GREEDY-SET-COVER:

$$|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| \geq |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| = u_{i-1}.$$

- Combining the last inequalities gives:

$$\sum_{x \in S} c_x \leq \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{u_{i-1}}$$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x \in S} c_x = \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$$

- Further, by definition of the GREEDY-SET-COVER:

$$|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| \geq |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| = u_{i-1}.$$

- Combining the last inequalities gives:

$$\sum_{x \in S} c_x \leq \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{u_{i-1}} = \sum_{i=1}^{k} \sum_{j=u_i+1}^{u_{i-1}} \frac{1}{u_{i-1}}$$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x \in S} c_x = \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$$

- Further, by definition of the GREEDY-SET-COVER:

$$|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| \geq |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| = u_{i-1}.$$

- Combining the last inequalities gives:

$$\sum_{x \in S} c_x \leq \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{u_{i-1}} = \sum_{i=1}^{k} \sum_{j=u_i+1}^{u_{i-1}} \frac{1}{u_{i-1}}$$

$$\leq \sum_{i=1}^{k} \sum_{j=u_i+1}^{u_{i-1}} \frac{1}{j}$$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x \in S} c_x = \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$$

- Further, by definition of the Greedy-Set-Cover:

$$|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| \geq |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| = u_{i-1}.$$

- Combining the last inequalities gives:

$$\begin{aligned}
\sum_{x \in S} c_x \leq \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{u_{i-1}} &= \sum_{i=1}^{k} \sum_{j=u_i+1}^{u_{i-1}} \frac{1}{u_{i-1}} \\
&\leq \sum_{i=1}^{k} \sum_{j=u_i+1}^{u_{i-1}} \frac{1}{j} \\
&= \sum_{i=1}^{k} (H(u_{i-1}) - H(u_i))
\end{aligned}$$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x \in S} c_x = \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$$

- Further, by definition of the Greedy-Set-Cover:

$$|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| \geq |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| = u_{i-1}.$$

- Combining the last inequalities gives:

$$\begin{aligned} \sum_{x \in S} c_x \leq \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{u_{i-1}} &= \sum_{i=1}^{k} \sum_{j=u_i+1}^{u_{i-1}} \frac{1}{u_{i-1}} \\ &\leq \sum_{i=1}^{k} \sum_{j=u_i+1}^{u_{i-1}} \frac{1}{j} \\ &= \sum_{i=1}^{k} (H(u_{i-1}) - H(u_i)) = H(u_0) - H(u_k) \end{aligned}$$

Proof of Theorem 35.4 (2/2)

Proof of the Key Inequality $\boxed{\sum_{x \in S} c_x \leq H(|S|)}$

- For any $S \in \mathcal{F}$ and $i = 1, 2, \ldots, |\mathcal{C}| = k$ let $u_i := |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)|$

$\Rightarrow$ $|X| = u_0 \geq u_1 \geq \cdots \geq u_{|\mathcal{C}|} = 0$ and $u_{i-1} - u_i$ counts the items in S covered first time by S_i.

$\Rightarrow$

$$\sum_{x \in S} c_x = \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})|}$$

- Further, by definition of the Greedy-Set-Cover:

$$|S_i \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| \geq |S \setminus (S_1 \cup S_2 \cup \cdots \cup S_{i-1})| = u_{i-1}.$$

- Combining the last inequalities gives:

$$\begin{aligned}
\sum_{x \in S} c_x \leq \sum_{i=1}^{k} (u_{i-1} - u_i) \cdot \frac{1}{u_{i-1}} &= \sum_{i=1}^{k} \sum_{j=u_i+1}^{u_{i-1}} \frac{1}{u_{i-1}} \\
&\leq \sum_{i=1}^{k} \sum_{j=u_i+1}^{u_{i-1}} \frac{1}{j} \\
&= \sum_{i=1}^{k} (H(u_{i-1}) - H(u_i)) = H(u_0) - H(u_k) = H(|S|). \qquad \square
\end{aligned}$$

Set-Covering Problem (Summary)

Theorem 35.4

GREEDY-SET-COVER is a polynomial-time $\rho(n)$-algorithm, where

$$\rho(n) = H(\max\{|S|: |S| \in \mathcal{F}\}) \leq \ln(n) + 1.$$

Set-Covering Problem (Summary)

Theorem 35.4

GREEDY-SET-COVER is a polynomial-time $\rho(n)$-algorithm, where

$$\rho(n) = H(\max\{|S| \colon |S| \in \mathcal{F}\}) \leq \ln(n) + 1.$$

- Is the bound on the approximation ratio in Theorem 35.4 tight?
- Is there a better algorithm?

Set-Covering Problem (Summary)

Theorem 35.4

GREEDY-SET-COVER is a polynomial-time $\rho(n)$-algorithm, where

$$\rho(n) = H(\max\{|S| \colon |S| \in \mathcal{F}\}) \leq \ln(n) + 1.$$

- Is the bound on the approximation ratio in Theorem 35.4 tight?
- Is there a better algorithm?

Lower Bound

Unless P=NP, there is no $c \cdot \ln(n)$ polynomial-time approximation algorithm for some constant $0 < c < 1$.

Set-Covering Problem (Summary)

Theorem 35.4

GREEDY-SET-COVER is a polynomial-time $\rho(n)$-algorithm, where

$$\rho(n) = H(\max\{|S| : |S| \in \mathcal{F}\}) \leq \ln(n) + 1.$$

Can be applied to the Vertex Cover Problem for Graphs with maximum degree 3 to obtain approximation ratio of $1 + \frac{1}{2} + \frac{1}{3} < 2$.

- Is the bound on the approximation ratio in Theorem 35.4 tight?
- Is there a better algorithm?

Lower Bound

Unless P=NP, there is no $c \cdot \ln(n)$ polynomial-time approximation algorithm for some constant $0 < c < 1$.

Set-Covering Problem (Summary)

The same approach also gives an approximation ratio of $O(\ln(n))$ if there exists a cost function $c : S \to \mathbb{Z}^+$

Theorem 35.4

GREEDY-SET-COVER is a polynomial-time $\rho(n)$-algorithm, where

$$\rho(n) = H(\max\{|S| : |S| \in \mathcal{F}\}) \leq \ln(n) + 1.$$

Can be applied to the Vertex Cover Problem for Graphs with maximum degree 3 to obtain approximation ratio of $1 + \frac{1}{2} + \frac{1}{3} < 2$.

- Is the bound on the approximation ratio in Theorem 35.4 tight?
- Is there a better algorithm?

Lower Bound

Unless P=NP, there is no $c \cdot \ln(n)$ polynomial-time approximation algorithm for some constant $0 < c < 1$.

Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$

Example where the solution of Greedy is bad

Instance

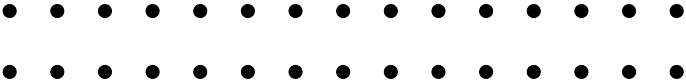
- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)

Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)

$k = 4, n = 30$:

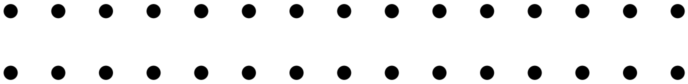


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements

$k = 4, n = 30$:

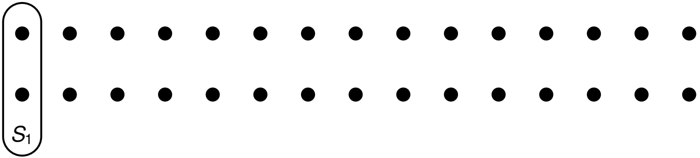


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements

$k = 4, n = 30$:

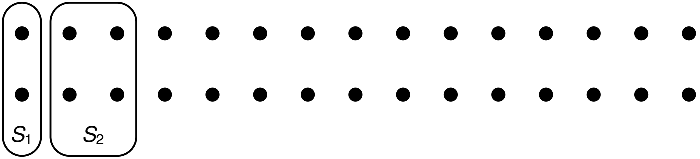


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements

$k = 4, n = 30$:

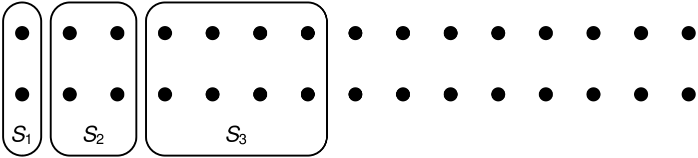


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements

$k = 4, n = 30$:

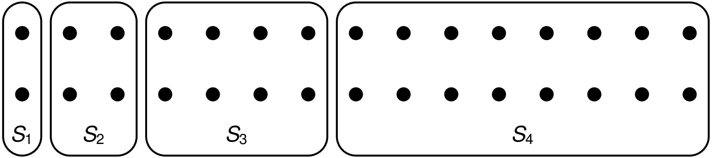


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements

$k = 4, n = 30$:

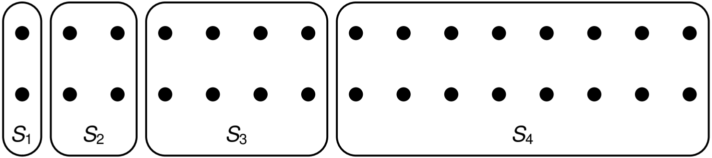


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements
- Sets T_1, T_2 are disjoint and each set contains half of the elements of each set $S_1, S_2, \ldots, S_k$

$k = 4, n = 30$:

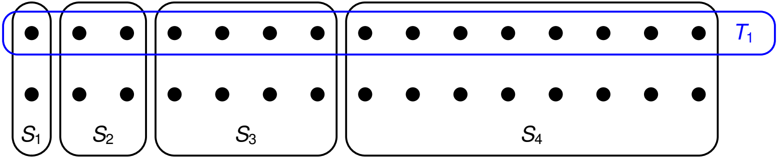


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements
- Sets T_1, T_2 are disjoint and each set contains half of the elements of each set $S_1, S_2, \ldots, S_k$

$k = 4, n = 30$:

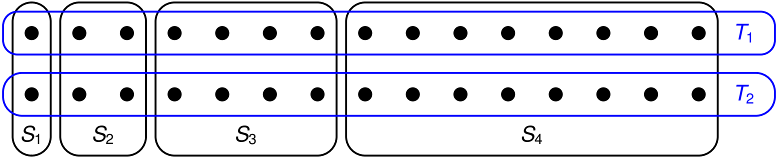


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements
- Sets T_1, T_2 are disjoint and each set contains half of the elements of each set $S_1, S_2, \ldots, S_k$

$k = 4, n = 30$:

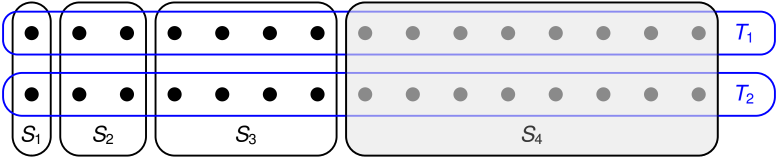


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements
- Sets T_1, T_2 are disjoint and each set contains half of the elements of each set $S_1, S_2, \ldots, S_k$

$k = 4, n = 30$:

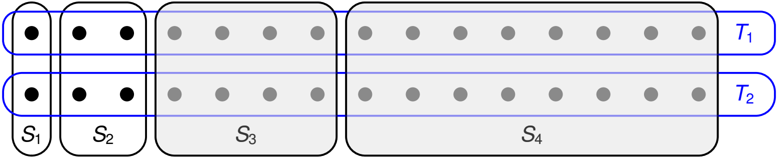


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements
- Sets T_1, T_2 are disjoint and each set contains half of the elements of each set $S_1, S_2, \ldots, S_k$

$k = 4, n = 30$:

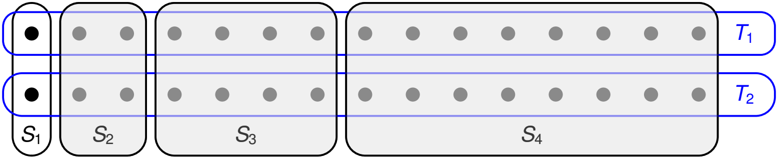


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements
- Sets T_1, T_2 are disjoint and each set contains half of the elements of each set $S_1, S_2, \ldots, S_k$

$k = 4, n = 30$:

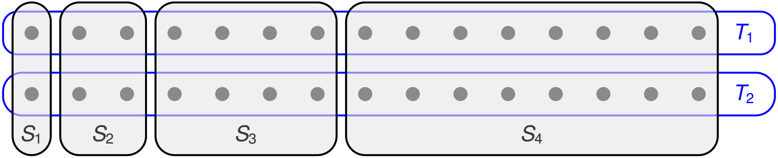


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements
- Sets T_1, T_2 are disjoint and each set contains half of the elements of each set $S_1, S_2, \ldots, S_k$

$k = 4, n = 30$:

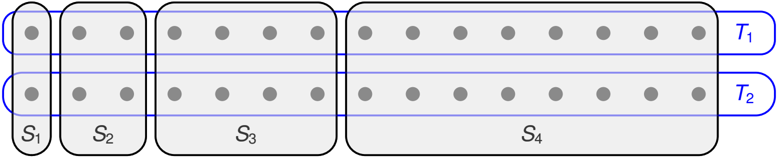


Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements
- Sets T_1, T_2 are disjoint and each set contains half of the elements of each set $S_1, S_2, \ldots, S_k$

$k = 4, n = 30$:



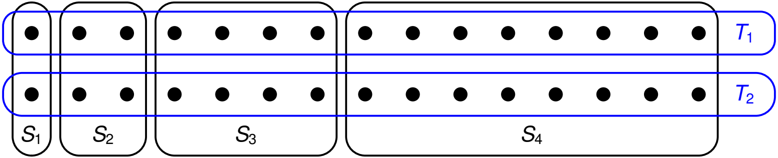
Solution of Greedy consists of k sets.

Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements
- Sets T_1, T_2 are disjoint and each set contains half of the elements of each set $S_1, S_2, \ldots, S_k$

$k = 4, n = 30$:



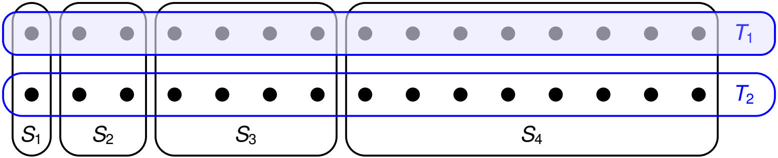
Solution of Greedy consists of k sets.

Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements
- Sets T_1, T_2 are disjoint and each set contains half of the elements of each set $S_1, S_2, \ldots, S_k$

$k = 4, n = 30$:



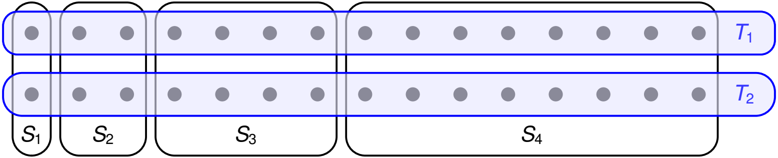
Solution of Greedy consists of k sets.

Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements
- Sets T_1, T_2 are disjoint and each set contains half of the elements of each set $S_1, S_2, \ldots, S_k$

$k = 4, n = 30$:



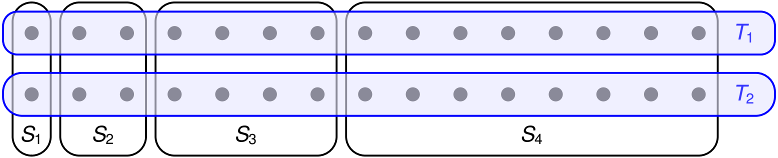
Solution of Greedy consists of k sets.

Example where the solution of Greedy is bad

Instance

- Given any integer $k \geq 3$
- There are $n = 2^{k+1} - 2$ elements overall (so $k \approx \log_2 n$)
- Sets $S_1, S_2, \ldots, S_k$ are pairwise disjoint and each set contains $2, 4, \ldots, 2^k$ elements
- Sets T_1, T_2 are disjoint and each set contains half of the elements of each set $S_1, S_2, \ldots, S_k$

$k = 4, n = 30$:



Solution of Greedy consists of k sets.

Optimum consists of 2 sets.