Proceedings of Machine Learning Research vol :1–14, 2026 39th Annual Conference on Learning Theory

# Active Learning and Covering Problems with Precedence

**author names withheld**

**Editor:** Under Review for COLT 2026

## Abstract

In the Bayesian Active Learning a hidden hypothesis is required to be uncovered. To do so, the learner is allowed to perform tests, each of which reveals partial information about the hidden hypothesis. Upon receiving this information, the learner adaptively selects the next test to be performed. The goal is to uncover the hidden hypothesis while performing as few tests as possible in the worst or average case.

In the covering problems, we are given a set of items and a collection of subsets that cover these items. The objective is to select a sequence of subsets that covers all items, which minimizing the worst or average covering cost.

For both types of problems, a natural constraint may arise that some tests can only be performed only after certain other tests (or some subsets can only be selected after selecting certain other subsets). We model such constraints using directed acyclic graphs (DAGs) that impose precedence on the tests or subsets. This paper explores the connection of active learning and covering problems under such constraints.

We show that given any bicriteria $(O(1)\,, \alpha)$-approximation ratio for the Precedence Constrained Set Cover, we can obtain an $O(\alpha \cdot \log n)$-approximation ratio for the Worst Case Active Learning with precedence constraints, where $n$ is the number of hypothesis. Similarly, we prove that given any $O(\beta)$-approximation ratio for the Precedence Constrained Min-Sum Set Cover, we can obtain an $O(\beta \cdot \log n)$-approximation ratio for the Average Case Active Learning with Precedence Constraints. Finally, we provide several approximation algorithms for the Set Cover and Min-Sum Set Cover problems with various types of precedence constraints.

**Keywords:** Bayesian active learning, Set cover, Precedence constraints, Approximation Algorithms, Decision Trees

## 1. Introduction

Consider following problems:

- The *Precedence Constrained Bayesian Active Learning Problem* consists a set of $\mathcal{H}$ of $n$ hypothesis, a set $\mathcal{T}$ of $m$ tests and a DAG (directed acyclic graph) $\mathcal{F} = \{\mathcal{T}, \preceq\}$ encoding the precedence constraints between available tests. Among $\mathcal{H}$ a hidden hypothesis is required to be encovered. To do so, the learner is allowed to perform tests, each of which reveals partial information about the hidden hypothesis. Upon receiving this information, the learner adaptively selects the next test to be performed. Importantly, in order to perform such test the learner needs to perform all of its predecesors in $\mathcal{F}$ first. The goal is to uncover the hidden hypothesis while performing as few tests as possible. Depending on the chosen criterion we distinguish between the *Precedence Constrained Worst Case Active Learning* (PCWCAL) and *Precedence Constrained Average Case Active Learning* (PCACAL) problems.

- The *Precedence Constrained Covering Problem* consists of a set of $n$ items $\mathcal{U}$, a collection $\mathcal{S}$ of $m$ subsets of $\mathcal{U}$ that cover these items, and a DAG $\mathcal{F} = \{\mathcal{S}, \preceq\}$ encoding the precedence constraints between available subsets. The goal is to select a sequence of tests that covers at least $K$ items. Depending on the chosen criterion we distinguish between the *Precedence Constrained Set Cover* (PCSC) and *Precedence Constrained Min-Sum Set Cover* (PCMSSC) problems. In the first we are only interested in minimizing the number of selected subsets, while in the second we want to minimize the average time it takes to cover an item.

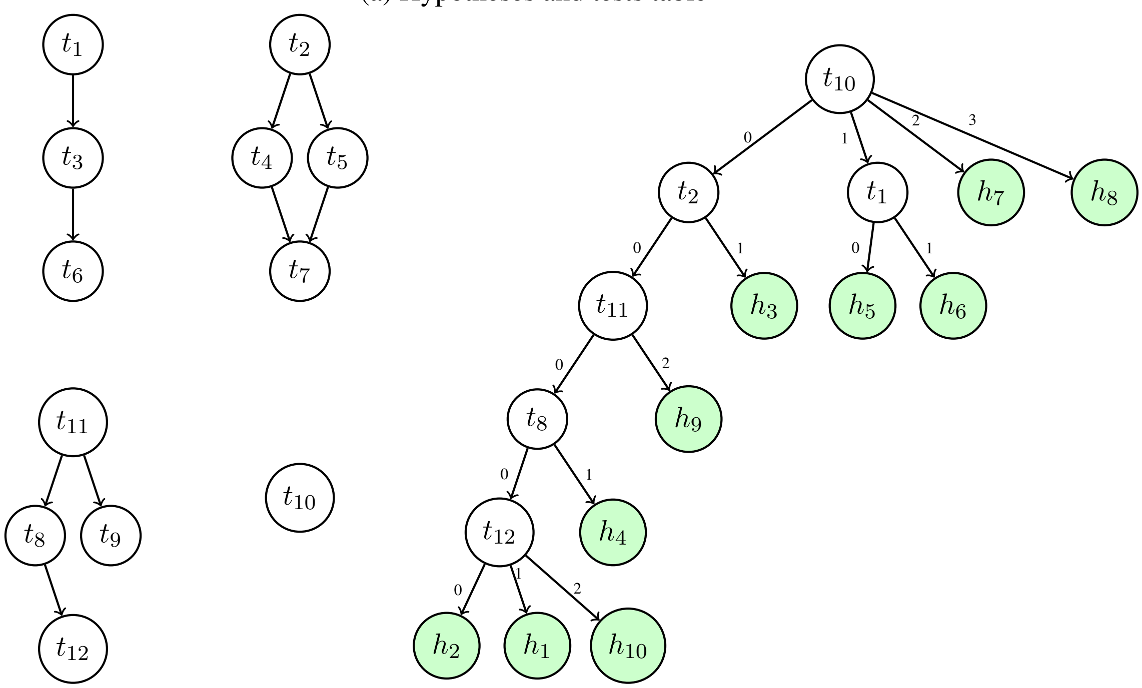
### 1.1. Our results and techniques

| **precedence/problem** | **PCSC** | **PCMSSC** | **PCWCAL** | **PCACAL** |
|---|---|---|---|---|
| none | $O(\log n)$ | 4 | $O(\log n)$ | $O(\log n)$ |
| inforest | $O(\log n)$* | 4 | $O(\log n)$* | $O(\log n)$* |
| outforest | $O(\log^2 n)$** | $O(\log n)$** | $O(\log^2 n)$* | $O(\log^2 n)$* |
| general | $O(\sqrt{n} \log n)$* | $O(\sqrt{n})$ | $O(\sqrt{n} \log n)$* | $O(\sqrt{n} \log n)$* |

Table 1: Approximation algorithms for various covering and active learning problems under different precedence constraints. (* denotes new results, ** denotes previously unmentioned corollaries of known results)

| | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ | $t_7$ | $t_8$ | $t_9$ | $t_{10}$ | $t_{11}$ | $t_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $h_1$ | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 |
| $h_2$ | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $h_3$ | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $h_4$ | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| $h_5$ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| $h_6$ | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| $h_7$ | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 2 | 0 | 0 |
| $h_8$ | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 |
| $h_9$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 2 | 0 |
| $h_{10}$ | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |

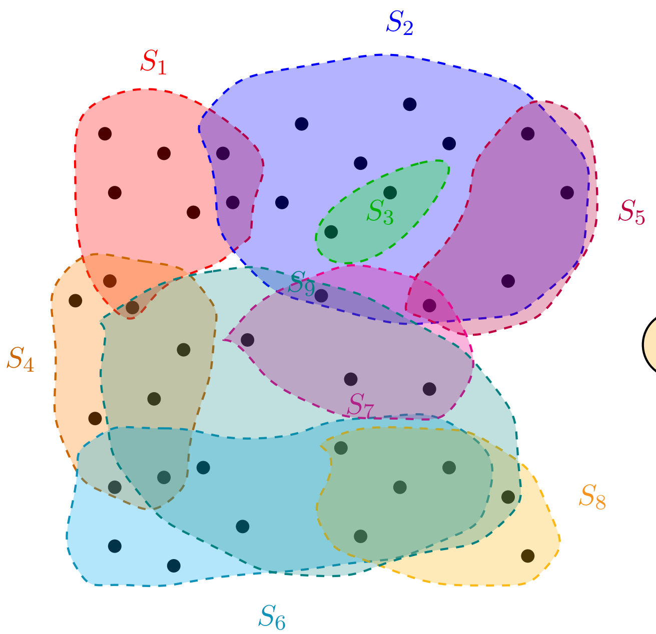
(a) Hypotheses and tests table



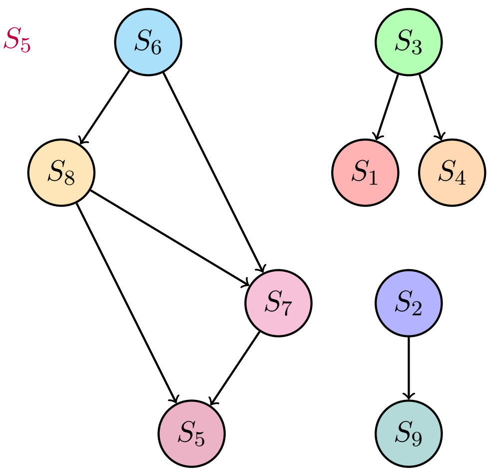

(b) Precedence DAG

(c) Valid decision tree

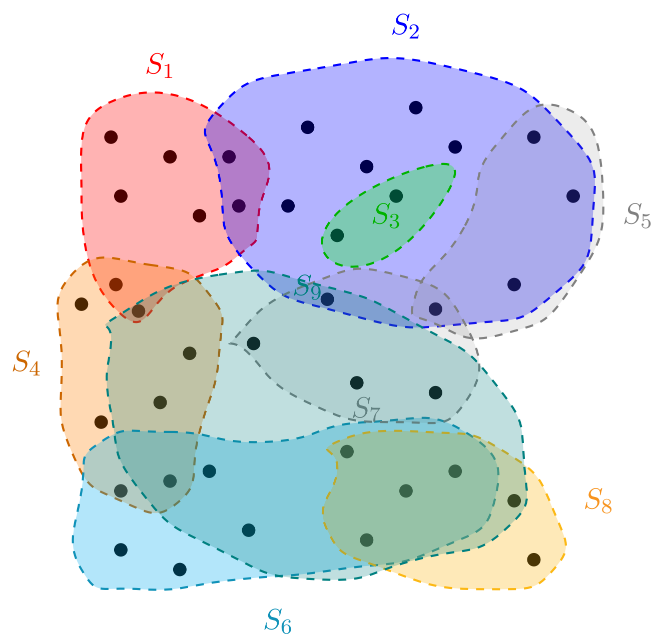
Figure 1: Example of a PCAL instance with 10 hypotheses and 12 tests. (a) Hypotheses-tests table. (b) Precedence DAG with four components. (c) A valid decision tree solution respecting precedence constraints.




(a) Universe and covering sets




(b) Precedence DAG




(c) Solution: selected sets $\{S_1, S_2, S_6, S_8, S_9\}$ (in color)

Figure 2: Example of a PCCP instance with 39 elements and 9 covering sets. (a) Universe with covering sets. (b) Precedence DAG with three components. (c) Solution using 7 selected sets (colored).

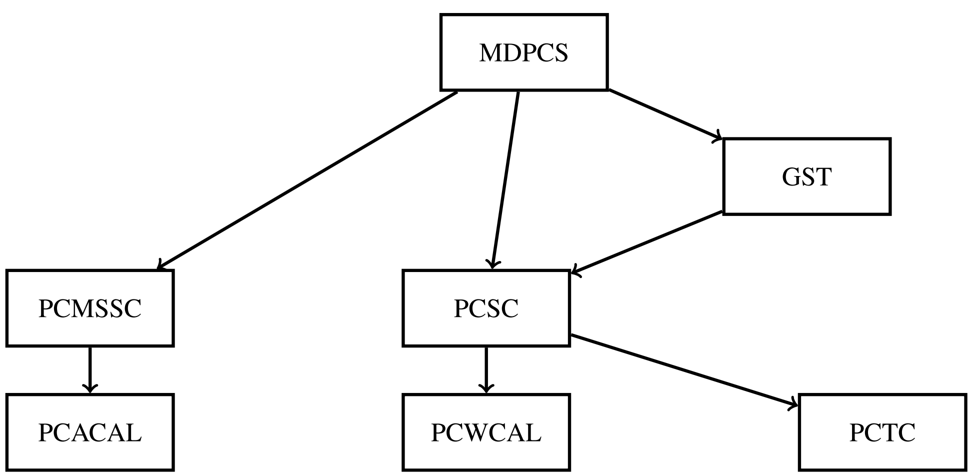



Figure 3: Relationships between covering and active learning problems, $\Pi_1 \to \Pi_2$ denotes that an approximation algorithm for problem $\Pi_1$ implies an approximation algorithm for problem $\Pi_2$.

## 2. Preliminaries

**Definition 1 (Precedence constrained set cover (PCSC))** *Given a universe $\mathcal{U}$ of $n$ items, a collection $\mathcal{S}$ of $m$ subsets of $\mathcal{U}$, a DAG $\mathcal{F} = \{\mathcal{S}, \preceq\}$ encoding precedence constraints, and a coverage requirement $K$, find a precedence-closed subfamily $\mathcal{C} \subseteq \mathcal{S}$ that covers at least $K$ items while minimizing $|\mathcal{C}|$.*

**Definition 2 (Precedence constrained min-sum set cover (PCMSSC))** *Given a universe $\mathcal{U}$ of $n$ items, a collection $\mathcal{S}$ of $m$ subsets of $\mathcal{U}$, a DAG $\mathcal{F} = \{\mathcal{S}, \preceq\}$ encoding precedence constraints, and a coverage requirement $K$, find a precedence-closed sequence of sets that covers at least $K$ items while minimizing the average time (position in sequence) at which items are covered.*

**Definition 3 (Precedence constrained test cover (PCTC))** *Given a set $\mathcal{H}$ of $n$ hypotheses, a set $\mathcal{T}$ of $m$ tests, and a DAG $\mathcal{F} = \{\mathcal{T}, \preceq\}$ encoding precedence constraints, find a precedence-closed subfamily of tests that distinguishes all pairs of hypotheses.*

**Definition 4 (Precedence constrained worst case active learning (PCWCAL))** *Given a set $\mathcal{H}$ of $n$ hypotheses, a set $\mathcal{T}$ of $m$ tests, and a DAG $\mathcal{F} = \{\mathcal{T}, \preceq\}$ encoding precedence constraints, construct a decision tree respecting precedence constraints that identifies any hypothesis from $\mathcal{H}$ while minimizing the worst-case depth of the tree.*

**Definition 5 (Precedence constrained average case active learning (PCACAL))** *Given a set $\mathcal{H}$ of $n$ hypotheses, a set $\mathcal{T}$ of $m$ tests, a DAG $\mathcal{F} = \{\mathcal{T}, \preceq\}$ encoding precedence constraints, and a probability distribution over $\mathcal{H}$, construct a decision tree respecting precedence constraints that identifies any hypothesis from $\mathcal{H}$ while minimizing the expected depth (average case cost).*

**Definition 6 (Group Steiner Tree (GST))** *Given an undirected graph $G = (V, E)$ with edge costs, a root vertex $r \in V$, and groups $g_1, \ldots, g_k \subseteq V$, find a minimum-cost tree $T$ rooted at $r$ that contains at least one vertex from each group $g_i$.*

For a subfamily $\mathcal{A} \subseteq \mathcal{F}$, we define the coverage as:

$$\mathrm{cov}(\mathcal{A}) \equiv \bigcup_{A \in \mathcal{A}} A$$

For a subset $X$ of the universe, the coverage on $X$ is:

$$\mathrm{cov}(\mathcal{A}, X) = \mathrm{cov}(\mathcal{A}) \cap X$$

The density $\Delta$ of a nonempty subfamily $\mathcal{A}$ on subset $X$ is:

$$\Delta(\mathcal{A}, X) \equiv \frac{|\mathrm{cov}(\mathcal{A}, X)|}{|\mathcal{A}|}$$

For convenience, we define $\Delta(\emptyset, X) < 0$.

**Definition 7 (Max-Density Precedence-Closed Subfamily (MDPCS))** *Given a family of $m$ sets $\mathcal{G}$, a precedence relation $\prec$, and a set of $n$ items to be covered $R \subseteq \mathrm{cov}(\mathcal{G})$, the MDPCS problem asks to find a precedence-closed subfamily $\mathcal{A} \subseteq \mathcal{G}$ that maximizes $\Delta(\mathcal{A}, R)$.*

For $S \in \mathcal{G}$, let $P[S]$ denote the minimal precedence-closed subfamily of $\mathcal{G}$ containing $S$ (i.e., the ancestors of $S$ including $S$ itself).

## 3. Active Learning via Covering Problems

We begin with the following folklore lemma concerning both worst and average case learning.

**Lemma 8** *Let $I = (\mathcal{H}, \mathcal{T}, \mathcal{F})$ be any PCAL instance. Let $\mathcal{H}' \subseteq \mathcal{H}$. Then $\texttt{OPT}(\mathcal{H}', \mathcal{T}, \mathcal{F}) \leq \texttt{OPT}(I)$.*

### 3.1. Worst Case

**Definition 9 (Coversep)** *Let $D$ be any decision tree for $\mathcal{I} = (\mathcal{H}, \mathcal{T}, \mathcal{F})$. We define a sequence of tests $P_D$ called* coversep *as follows. Initially, $P_D$ is empty and $\mathcal{H}' = \mathcal{H}$. While $|\mathcal{H}'| > |\mathcal{H}|/2$, we append to $P_D$ the test $r(D_{\mathcal{H}'})$ and update $\mathcal{H}'$ to be the set of hypotheses corresponding to the child of $D_{\mathcal{H}'}$ that contains the most hypotheses. If $\texttt{COST}(D) = \texttt{OPT}(\mathcal{I})$, then we denote $P^*(\mathcal{I}) = P_D$ (ties broken arbitrarily).*

It should be remarked that $P_D$ is well-defined, as each test in $P_D$ can have at most one child associated with more than half of the hypotheses in $\mathcal{H}'$. Since $P_D$ is a subpath of $D$, we also have the following simple observation.

**Observation 1** *Let $I$ be any instance of PCWCAL. Then $|P^*(I)| \leq \texttt{OPT}(I)$.*

We will use $|P^*(I)|$ as a lower bound on $\texttt{OPT}(I)$ in the analysis of the approximation algorithm for PCWCAL. We have the following lemma:

**Lemma 10** *Let $I = (\mathcal{H}, \mathcal{T}, \mathcal{F})$ be any PCWCAL instance. Let $S^*$ be the optimal solution for the PCSC on instance $(\mathcal{H}, \mathcal{T}, \mathcal{F}, f)$ where $f = 1/4$ and a test $t$ covers $h \in \mathcal{H}$ if $\{U_t(u)\} \leq \frac{3}{4} \cdot \{\mathcal{U}\}$. Then, $|S^*| \leq |P^*(I)|$.*

**Proof** We show that $P^*(I)$ is a feasible solution for the PCSC instance $(\mathcal{H}, \mathcal{T}, \mathcal{F}, K)$. Assume towards a contradiction that this is not the case, i. e. less than $|\mathcal{H}|/4$ are covered by $P^*(I)$. Therefore there exists $t \in P^*(I)$ and a response $\mathcal{H}'$ of size $|\mathcal{H}'| \leq |\mathcal{H}|/2$ such that hypotheses in $\mathcal{H}'$ are not covered by $P^*(I)$, otherwise the claim holds trivially, since all hypotheses are covered. Let $\mathcal{H}' \subseteq U_{t,j}$ (since $\mathcal{H}'$ is a response to a test $t$, such $U_{t,j}$ always exists). By assumption, we have that $|U_{t,j} - \mathcal{H}'| < |\mathcal{H}|/4$. Therefore, we have that $|U_{t,j}| = |\mathcal{H}'| + |U_{t,j} - \mathcal{H}'| < 3/4 \cdot |\mathcal{H}|$ which by definition means that $h$ is covered by $P^*(I)$, a contradiction. ∎

**Theorem 11** *If there is an $(\gamma, \alpha)$-bicriteria approximation algorithm for PCSC then there is an $O\left(\frac{\alpha}{\log\left(\frac{2\gamma}{2\gamma-1}\right)} \cdot \log n\right)$-approximation algorithm for PCWCAL. In particular when $\gamma = O(1)$, the approximation is $O(\alpha \cdot \log n)$.*

**Proof** The following observation follows by Lemmas 8 and 13:

**Observation 2** *Let $D_S$ be the decision tree built on tests from $S$ closed under $\mathcal{F}$. Then, $\texttt{COST}(D_S) \leq \alpha \cdot |P^*(I)|$.*

We are now ready to prove the theorem.

**Algorithm 1:** The $O(\alpha \cdot \log n)$-approximation algorithm for the PCWCAL

**procedure** $\textsc{WorstDecisionTree}(\mathcal{H}, \mathcal{T}, \mathcal{F})$

- **if** $|\mathcal{H}| = 1$ **then**
  - **return** the trivial decision tree with a single leaf corresponding to the only hypothesis in $\mathcal{H}$.
- **end**
- **foreach** $t \in \mathcal{T}$ **do**
  - Set $t$ to cover $u \in \mathcal{H}$ if for $u \in U_{t,j}$, $|U_{t,j}| \leq \frac{3}{4} \cdot |\mathcal{H}|$.
- **end**
- $S \leftarrow$ Run the $(\gamma, \alpha)$-approximation algorithm for PCSC on instance $(\mathcal{H}, \mathcal{T}, \mathcal{F}, f)$ with $f = 1/4$.
- $D \leftarrow D_S \leftarrow$ any decision tree built on tests from $S$ closed under $\mathcal{F}$.
- **foreach** $\mathcal{H}' \in \mathcal{H} - S$ **do**
  - $D' \leftarrow \textsc{WorstDecisionTree}(\mathcal{H}', \mathcal{T} - S, \mathcal{F} - S)$.
  - Attach $D'$ to the leaf of $D$ corresponding to $\mathcal{H}'$.
- **end**
- **return** $D$.

**Lemma 12** *Let $D$ be the decision tree returned by* $\textsc{WorstDecisionTree}$ *on input* $I = (\mathcal{H}, \mathcal{T}, \mathcal{F})$. *Then,* $\mathit{COST}(D) \leq \frac{\alpha}{\log\left(\frac{4\gamma}{4\gamma-1}\right)} \cdot \log n \cdot \mathit{OPT}(I)$.

**Proof** We prove the lemma by induction on $n$. The base case when $n = 1$ is trivial since the the cost of the decision tree is 0. Assume by induction that for every $I' = (\mathcal{H}', \mathcal{T}, \mathcal{F})$ such that $\mathcal{H}' \in \mathcal{H} - S$ and $n' = |\mathcal{H}'|$ we have $\mathtt{COST}(D') \leq \frac{\alpha}{\log\left(\frac{4\gamma}{4\gamma-1}\right)} \cdot \log n' \cdot \mathtt{OPT}(I')$, where $D'$ is the decision tree returned by $\textsc{WorstDecisionTree}$ on input $I'$. We have that:

$$
\begin{aligned}
\mathtt{COST}(D) \leq& \mathtt{COST}(D_S) + \max_{\mathcal{H}' \in \mathcal{H} - S} \mathtt{COST}\big(D'\big) \\
\leq& \alpha \cdot |S^*| + \max_{\mathcal{H}' \in \mathcal{H} - S} \frac{\alpha}{\log\left(\frac{4\gamma}{4\gamma-1}\right)} \cdot \log n' \cdot \mathtt{OPT}\big(I'\big) \\
\leq& \alpha \cdot |P^*(I)| + \frac{\alpha}{\log\left(\frac{4\gamma}{4\gamma-1}\right)} \cdot \log\left(\frac{(4\gamma - 1) \cdot n}{4\gamma}\right) \cdot \mathtt{OPT}(I) \\
=& \alpha \cdot \mathtt{OPT}(I) + \frac{\alpha}{\log\left(\frac{4\gamma}{4\gamma-1}\right)} \cdot \log n \cdot \mathtt{OPT}(I) - \alpha \cdot \mathtt{OPT}(I) \\
=& \frac{\alpha}{\log\left(\frac{4\gamma}{4\gamma-1}\right)} \cdot \log n \cdot \mathtt{OPT}(I)
\end{aligned}
$$

where the second inequality follows by the induction hypothesis, the third inequality follows by Lemma 8 and the fact that $|\mathcal{H}'| \leq \frac{(4\gamma-1)}{4\gamma} \cdot n$ for every $\mathcal{H}' \in \mathcal{H} - S$ and the equalities follow by rearranging terms. This concludes the proof of the lemma. ■

■

### 3.2. Average Case

It should be remarked that $C_D$ is well-defined, as each test in $C_D$ can have at most one child associated with more than half of the hypotheses in $\mathcal{H}'$. Since $C_D$ is a subpath of $D$, we also have the following simple observation.

**Observation 3** *Let $I$ be any instance of PCACAL. Then $\mathtt{COST}(C^*(I)) \leq \mathtt{OPT}(I)$.*

This allows to use $\mathtt{COST}(C^*(I))$ as a lower bound on $\mathtt{OPT}(I)$ in the analysis of the approximation algorithm for PCACAL. We have the following lemma:

**Lemma 13** *Let $I = (\mathcal{H}, \mathcal{T}, \mathcal{F})$ be any PCACAL instance. Let $S^*$ be the optimal solution for the PCMSSC on instance $(\mathcal{U}, \mathcal{T}, \mathcal{F}, f)$ with $f = 1/4$, where $\mathcal{U} = \mathcal{H}$ and a test $t$ element $u \in \mathcal{U}$ if for $u \in U_{t,j}$, $|U_{t,j}| \leq \frac{3}{4} \cdot |\mathcal{U}|$. Then, $\mathtt{COST}(S^*) \leq \mathtt{COST}(C^*(I))$.*

**Theorem 14** *If there is a $\beta$- approximation algorithm for PCMSSC then there is an $O(\beta \cdot \log n)$-approximation algorithm for PCACAL.*

**Algorithm 2:** The $O(\beta \cdot \log n)$-approximation algorithm for the PCACAL

```
procedure AVERAGEDECISIONTREE(H, T, F)
    if |H| = 1 then
        return the trivial decision tree with a single leaf corresponding to the only hypothesis in H.
    end
    foreach t ∈ T do
        Set t to cover u ∈ H if for u ∈ U_{t,j}, |U_{t,j}| ≤ 3/4 · |H|.
    end
    S ← Run the β-approximation algorithm for PCMSSC on instance (H, T, F, f) with f = 1/4.
    D ← D_S ← any decision tree built on tests from S closed under F.
    foreach H' ∈ H − S do
        D' ← AVERAGEDECISIONTREE(H', T − S, F − S).
        Attach D' to the leaf of D corresponding to H'.
    end
    return D.
```

## 4. Set covering with constraints

### 4.1. Max-Density Precedence-Closed Subfamily (MDPCS)

The key to solce PCSC and PCMSSC is to solve the MDPCS problem. An approximation algorithm for MDPCS can be used as an essential subroutine in our algorithms for PCSC and PCMSSC. By **?**, the following greedy algorithm achieves an $O(\sqrt{m})$-approximation for MDPCS:

**Algorithm 3:** The greedy algorithm for MDPCS

```
procedure MDPCS-GREEDY(𝒢, ≺, R)
    𝒜 ← 𝒢
    foreach S ∈ 𝒢 do
        if Δ(P[S], R) > Δ(𝒜, R) then
            𝒜 ← P[S]
        end
    end
    return 𝒜
```

Let $\delta = \max_{S \in \mathcal{G}} \Delta(P[S], R)$. When $\delta \geq 1$, then the approximation factor of the greedy can also be bounded by $O(\sqrt{n})$. We show that if we enforce a certain condition on the input called *$\epsilon$-shallow ancestry*, then for $\epsilon < 1$ the greedy algorithm achieves an $O(n^\epsilon)$-approximation .

For each $S \in \mathcal{G}$, let $p(S) = |P[S]|$ and $c(S, R) = |\mathrm{cov}(P[S], R)|$.

**Theorem 15** *Suppose there exists a constant $C > 0$ and $\epsilon \in (0, 1)$ such that for all $S \in \mathcal{G}$, $p(S) \leq C \cdot c(S)^\epsilon$ ($\epsilon$-shallow ancestry). Then MDPCS-Greedy provides an $O(n^\epsilon)$-approximation.*

**Proof** Let $\mathcal{A}^*$ be an optimal solution consisting of sets $S_1, \ldots, S_k$. There are two cases:

1. If $\delta \geq n^{1-\epsilon}$, then, we observe that $\Delta(\mathcal{A}, R) = \delta \geq n^{1-\epsilon}$. Since we can cover at most $n$ elements with at least one set, we have $\Delta(\mathcal{A}^*, R) \leq n$. Therefore:
$$\frac{\Delta(\mathcal{A}^*, R)}{\Delta(\mathcal{A}, R)} \leq \frac{n}{n^{1-\epsilon}} = n^\epsilon$$

2. Else, if $\delta \leq n^{1-\epsilon}$, we proceed as follows: By definition of density, for any $S \in \mathcal{G}$, $c(S) \leq \delta \cdot p(S)$. Combining this with the $\epsilon$-shallow ancestry condition, we have that for all $S \in \mathcal{G}$, $c(S) \leq \delta \cdot C \cdot c(S)^\epsilon$. Rearraning this inequality, we get that $c(S) \leq (\delta \cdot C)^{\frac{1}{1-\epsilon}}$. We have that:
$$|\mathrm{cov}(\mathcal{A}^*)| = \left| \bigcup_{j=1}^{k} \mathrm{cov}(S_j, R) \right| \leq \sum_{j=1}^{k} c(S_j) \leq k \cdot (\delta \cdot C)^{\frac{1}{1-\epsilon}}$$
Therefore:
$$\Delta(\mathcal{A}^*, R) = \frac{|\mathrm{cov}(\mathcal{A}^*, R)|}{k} \leq (\delta \cdot C)^{\frac{1}{1-\epsilon}}$$
By the greedy choice, $\Delta(\mathcal{A}, R) \geq \delta$ and by assumption $\delta \leq n^{1-\epsilon}$. Thus:
$$\frac{\Delta(\mathcal{A}^*, R)}{\Delta(\mathcal{A}, R)} \leq \frac{(\delta \cdot C)^{\frac{1}{1-\epsilon}}}{\delta} = C^{\frac{1}{1-\epsilon}} \cdot \delta^{\frac{\epsilon}{1-\epsilon}} \leq C^{\frac{1}{1-\epsilon}} \cdot (n^{1-\epsilon})^{\frac{\epsilon}{1-\epsilon}} = C^{\frac{1}{1-\epsilon}} \cdot n^\epsilon$$

Since $C$ is constant, the theorem follows. ∎

### 4.2. Precedence constrained set cover

We show the following:

**Theorem 16** *If there exists an $\gamma$ approximation algorithm for the MDPCS problem, then there exists an $(H_K + 1) \cdot \gamma$ - approximate algorithm for the PCSC problem.*

**Proof** ∎

**Algorithm 4:** The $\gamma$-greedy algorithm for PCSC

```
procedure PCSC(𝒰, 𝒮, ℱ, K)
    𝒞 ← ∅
    while |cov(𝒞,𝒰)| < K do
        𝒜 ← Run the γ-approx. algorithm for MDPCS on (𝒰 − 𝒞, 𝒮 − 𝒞, ℱ − 𝒞, m)
        if |cov(𝒞 ∪ 𝒜,𝒰)| ≥ K then
            Find the minimum budget B ∈ [|𝒜|], such that the γ-approx. algorithm for MDPCS on
            (𝒰 − 𝒞, 𝒮 − 𝒞, ℱ − 𝒞, B) returns a set ℬ with |cov(ℬ,𝒰 − 𝒞)| ≥ (K − cov(𝒞,𝒰))/α
            while |cov(𝒞,𝒰)| < K do
                ℬ ← Run the γ-approx. algorithm for MDPCS on (𝒰 − 𝒞, 𝒮 − 𝒞, ℱ − 𝒞, B)
                𝒞 ← 𝒞 ∪ ℬ
            end
            return 𝒞
        end
        foreach u ∈ cov(𝒜,𝒰 − 𝒞) do
            c(u) ← Δ(𝒜,𝒰 − 𝒞)
        end
        𝒞 ← 𝒞 ∪ 𝒜
    end
```

**Theorem 17** *If the precedence constraints form an outforest, then there exists an bicriteria $(4, O(\log n))$-approximation algorithm for PCSC which be converted to an $O(\log^2 n)$ approximation algorithm.*

### 4.3. Precedence constrained min sum set cover

By **?**:

**Theorem 18** *If there exists an $\gamma$ approximation algorithm for the MDPCS problem, then there exists an $4 \cdot \gamma$ - approximate algorithm for the PCMSSC problem.*

**Theorem 19** *If the precedence constraints form an outforest, then there exists an $(O(\log n))$-approximation algorithm for PCMSSC.*

## 5. Hardness

**Theorem 20** *PCWCAL with outforest precedence constraints is NP-hard to approximate within a factor of* $O(\log^2 n)$ *unless P = NP.*

## 6. Conclusions and Future Work

## Appendix A. My Proof of Theorem 1

This is a boring technical proof.

## Appendix B. My Proof of Theorem 2

This is a complete version of a proof sketched in the main text.