

Supplementary Material — Summarizing the Solution Space in Tumor Phylogeny Inference by Multiple Consensus Trees

Nuraini Aguse^{1,*} Yuanyuan Qi^{1,*}

Mohammed El-Kebir^{1,†}

¹Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801

Contents

A	Supplementary Text	2
A.1	Combinatorial Structure and Complexity	2
A.1.1	Combinatorial Characterization of Optimal Solutions	2
A.1.2	Complexity	3
A.2	Methods	6
A.2.1	Mixed Integer Linear Program	6
A.3	Supplementary Results	6

*Joint first authors.

†To whom correspondence should be addressed.

A Supplementary Text

A.1 Combinatorial Structure and Complexity

A.1.1 Combinatorial Characterization of Optimal Solutions

(Main Text) Theorem 1. Given input trees $\mathcal{T} = \{T_1, \dots, T_n\}$, there exists a consensus tree R with minimum distance $d(\mathcal{T}, R) = \sum_{i=1}^n d(T_i, R)$ that is a maximum weight spanning arborescence in the parent-child graph $G_{\mathcal{T}}$.

(Main Text) Lemma 1. There exists an optimal consensus tree R to SCT instance \mathcal{T} where each edge (u, v) of R occurs in an input tree.

Proof. By Theorem 1, a maximum weight spanning arborescence is an optimal solution to \mathcal{T} . Consider such a maximum weight spanning arborescence R' . By construction, the edge set of the parent-child graph $G_{\mathcal{T}}$ equals the union of all edges in the set \mathcal{T} of input trees. As such, R' does not contain an edge (u, v) that is not present in \mathcal{T} . \square

(Main Text) Lemma 2. There exists an optimal consensus tree R to SCT instance \mathcal{T} where if an edge (u, v) is present in all trees \mathcal{T} then (u, v) is an edge of the consensus tree R .

Proof. Consider an edge (u, v) that is present in all input trees \mathcal{T} . By Theorem 1 a maximum weight spanning arborescence is an optimal solution to \mathcal{T} . Consider such a maximum weight arborescence R' . We distinguish two cases. First, v is not the root of R' . Let u' be the unique parent of v in R' . By construction, the edge set of the parent-child graph $G_{\mathcal{T}}$ equals the union of all edges in the set \mathcal{T} of input trees. We have that u is the parent of v in each input tree $T_i \in \mathcal{T}$. As such, $u' = u$.

Second, v is the root of R' . Let w be the unique parent of u in R' . We construct a new tree $R'' = (V, E(R''))$, where we remove the edge (w, u) and introduce the edge (u, v) . That is, $E(R'') = E(R') \setminus \{(w, u)\} \cup \{(u, v)\}$. Let $\ell(R)$ be the sum of weights of all its edges in $G_{\mathcal{T}}$, i.e. $\ell(R) = \sum_{(u,v) \in E(R)} \ell(u,v)$. By construction of the parent-child graph $G_{\mathcal{T}}$, we have that $\ell(u, v) = n \geq \ell(w, u)$. Therefore $\ell(R') \leq \ell(R'')$. Since R' is an optimal tree, we know that R'' is also an optimal tree. Moreover, R'' contains the edge (u, v) , thereby proving the lemma. \square

(Main Text) Lemma 3. The total distance $d(\mathcal{T}, R) = \sum_{i=1}^n d(T_i, R)$ of any spanning arborescence $R = (V, E_R)$ of parent-child graph $G_{\mathcal{T}}$ to input trees $\mathcal{T} = \{T_1, \dots, T_n\}$ equals $2[n(m-1) - \ell(R)]$.

Proof. Let $\mathbf{1}_T(u, v)$ be an indicator function, where $\mathbf{1}_T(u, v) = 1$ if (u, v) is an edge of T , and 0 otherwise. Using that $|E_i| = |E_R| = m-1$ for any input tree $T_i = (V, E_i) \in \mathcal{T}$, we have

$$d(T_i, R) = (|E_i| - |E_i \cap E_R|) + (|E_R| - |E_i \cap E_R|) \quad (1)$$

$$= 2(m-1) - 2 \sum_{(u,v) \in E_R} \mathbf{1}_{T_i}(u, v). \quad (2)$$

The total cost is

$$\sum_i d(T_i, R) = \sum_{i=1}^n \left[2(m-1) - 2 \sum_{(u,v) \in E_R} \mathbf{1}_{T_i}(u, v) \right] \quad (3)$$

$$= 2n(m-1) - 2 \sum_{i=1}^n \sum_{(u,v) \in E_R} \mathbf{1}_{T_i}(u, v) \quad (4)$$

$$= 2n(m-1) - 2 \sum_{(u,v) \in E_R} \sum_i \mathbf{1}_{T_i}(u, v) \quad (5)$$

$$= 2(n(m-1) - \sum_{(u,v) \in E_R} \ell(u, v)) \quad (6)$$

$$= 2[n(m-1) - \ell(R)]. \quad (7)$$

\square

(Main Text) Proposition 1. Given a clustering $\sigma : [n] \rightarrow [k]$, the MCT problem decomposes into k independent SCT problems.

Proof. Consider a clustering σ . Let \mathcal{T}_s be the set $\{T_i \in \mathcal{T} \mid \sigma(i) = s\}$ of input trees assigned to cluster s . Let $\mathcal{R} = \{R_1, \dots, R_k\}$ be optimal solutions to the k SCT instances $\{\mathcal{T}_1, \dots, \mathcal{T}_k\}$ and let $\mathcal{R}' = \{R'_1, \dots, R'_k\}$ be an optimal solution to the MCT instance \mathcal{T} constrained to clustering σ . We claim that $d(\mathcal{T}_s, R_s) = d(\mathcal{T}_s, R'_s)$ for all clusters s .

Suppose for a contradiction that there exists a cluster s where $d(\mathcal{T}_s, R_s) \neq d(\mathcal{T}_s, R'_s)$. If $d(\mathcal{T}_s, R_s) < d(\mathcal{T}_s, R'_s)$ then \mathcal{R}' is not an optimal set of consensus trees, yielding a contradiction. If $d(\mathcal{T}_s, R_s) > d(\mathcal{T}_s, R'_s)$ then R_s is not an optimal consensus tree to \mathcal{T}_s , which contradicts Theorem 1. Hence, $d(\mathcal{T}_s, R_s) = d(\mathcal{T}_s, R'_s)$ for all clusters s . \square

(Main Text) Proposition 2. The minimum total distance of an MCT instance (\mathcal{T}, k) is monotonically decreasing with increasing number k of clusters.

Proof. Let $\mathcal{T} = \{T_1, \dots, T_n\}$ be a set of input trees. Consider $k < n$ number of clusters. Let (\mathcal{R}, σ) be the optimal solution for MCT instance (\mathcal{T}, k) , with total distance $d(\mathcal{T}, \mathcal{R}, \sigma)$. Since $k < n$, there must exist a cluster s with $|\mathcal{T}_s| > 1$ trees. By definition, the input trees in \mathcal{T}_s must be distinct from one another. Hence, there exists an input tree $T_i \in \mathcal{T}_s$ that differs from the consensus tree R_s . Consider a new clustering σ' where $\sigma'(j) = \sigma(j)$ if $j \neq i$ and $\sigma'(i) = k + 1$. Thus, σ' contains an additional cluster $k + 1$. Choosing the consensus tree R_{k+1} of this cluster to be equal to T_i results in a distance $d(T_i, R_{k+1})$ of 0. The distance of $d(\mathcal{T}_s \setminus \{T_i\}, R_s)$ similarly decreases as T_i is distinct from R_s . Hence, $d(\mathcal{T}, \mathcal{R}, \sigma) > d(\mathcal{T}, \mathcal{R} \cup R_{k+1}, \sigma')$. \square

A.1.2 Complexity

(Main Text) Lemma 4. The cost of a clustering $\sigma : [n] \rightarrow [k]$ that partitions \mathcal{T} into parts of sizes n_1, \dots, n_k is at least $2[(c-1) \cdot |E(H)| - \sum_{s=1}^k \binom{n_s}{2}]$. This bound is tight if and only if the input trees \mathcal{T}_s assigned to each cluster s encode a clique in the undirected graph H .

Proof. Recall that each input tree $T_i \in \mathcal{T}$ corresponds to vertex v_i of the undirected graph H . Let R_1, \dots, R_k be the corresponding optimal consensus trees of clustering σ .

Consider cluster $s \in [k]$. Let n_s be the number of input trees assigned to cluster s . Consider the parent-child graph $G_{\mathcal{T}_s}$ obtained from the input trees \mathcal{T}_s assigned to cluster s . This graph must contain the directed edges $\{(r, \perp), (r, v_1), \dots, (r, v_n)\}$, as these edges occur in each input tree $T_i \in \mathcal{T}$. As such, these directed edges will have weight $\ell(r, \perp) = \ell(r, v_1) = \dots = \ell(r, v_n) = n_s$. Now consider the remaining vertices of the parent-child graph $G_{\mathcal{T}_s}$ corresponding to edges $\{e_1, \dots, e_m\}$ of H . We distinguish three cases for each edge $e = (v_i, v_j)$.

1. Trees T_i and T_j are assigned to cluster s .

Each of T_i and T_j contains the directed edge (\perp, e) , as edge $e = (v_i, v_j)$ is incident to both v_i and v_j . By construction, each remaining tree $T_{i'} \in \mathcal{T}_s \setminus \{T_i, T_j\}$ does not contain the directed edge (\perp, e) but instead contains the directed edge $(v_{i'}, e)$. Thus, the parent-child graph $G_{\mathcal{T}_s}$ of cluster s contains the directed edge (\perp, e) with weight $\ell(\perp, e) = 2$, and directed edges $(v_{i'}, e)$ with weight $\ell(v_{i'}, e) = 1$ for each tree $T_{i'} \in \mathcal{T}_s \setminus \{T_i, T_j\}$.

2. Only one of trees T_i and T_j is assigned to cluster s .

Without loss of generality assume that tree T_i is assigned to cluster s (and T_j is not assigned to cluster s). By construction, tree T_i contains the directed edge (\perp, e) and each remaining tree $T_{i'} \in \mathcal{T}_s \setminus \{T_i\}$ contains the edge $(v_{i'}, e)$. Thus, the parent-child graph $G_{\mathcal{T}_s}$ of cluster s contains the directed edge (\perp, e) with weight $\ell(\perp, e) = 1$, and directed edges $(v_{i'}, e)$ with weight $\ell(v_{i'}, e) = 1$ for each tree $T_{i'} \in \mathcal{T}_s \setminus \{T_i\}$.

3. Neither T_i nor T_j is assigned to cluster s .

By construction, each tree $T_{i'} \in \mathcal{T}_s$ contains the directed edge $(v_{i'}, e)$. Thus, the parent-child graph $G_{\mathcal{T}_s}$ of cluster s contains the directed edge $(v_{i'}, e)$ with weight $\ell(v_{i'}, e) = 1$ for each tree $T_{i'} \in \mathcal{T}_s$.

Consider the consensus tree $R_s = (U, A'_s)$ for cluster s obtained from the optimal solution (\mathcal{R}, σ) . By Theorem 1, R_s is a spanning tree of the parent-child graph of cluster s with maximum weight $\ell(R_s)$. What is the largest value that $\ell(R_s)$ can attain? To answer this question, observe that vertices $\{\perp, v_1, \dots, v_n\}$ each have a unique parent r in the parent-child graph $G_{\mathcal{T}_s}$, with a corresponding edge weight of n_s . This amounts to a total weight of $n_s(n+1)$. Consider the remaining vertices $\{e_1, \dots, e_{|E(H)|}\}$. First, observe that there are no edges among these vertices in the parent-child graph. Thus, including an edge in R_s to any one vertex in $\{e_1, \dots, e_{|E(H)|}\}$ does not affect any other vertex in the same set. In other words, the maximum weight spanning tree R_s will contain for each $e \in \{e_1, \dots, e_{|E(H)|}\}$ a directed edge to e with maximum weight. Now, the maximum weight of 2 for such an edge is achieved when $e = (v_i, v_j)$ is an edge that corresponds to input trees T_i and T_j that are assigned to cluster s . Let E_s be the subset of edges of H whose two incident vertices correspond to two input trees assigned to cluster s , i.e. $E_s = \{(v_i, v_j) \in E(H) \mid \sigma(i) = \sigma(j) = s\}$. As for the remaining edges $E(H) \setminus E_s$, the maximum weight spanning tree R_s will contain for each $e \in E \setminus E_s$ a directed edge to e with weight 1. Hence,

$$\ell(R_s) = n_s(n+1) + 2|E_s| + |E \setminus E_s| \quad (8)$$

$$= n_s(n+1) + 2|E_s| + (|E(H)| - |E_s|) \quad (9)$$

$$= n_s(n+1) + |E_s| + |E(H)|. \quad (10)$$

Clearly, $|E_s| \leq \binom{n_s}{2}$; this bound is tight if and only if the vertices of H corresponding to input trees \mathcal{T}_s assigned to cluster s form a clique. Thus,

$$\ell(R_s) \leq n_s(n+1) + \binom{n_s}{2} + |E(H)|. \quad (11)$$

By Lemma 3, we have that the sum $d(\mathcal{T}_s, R_s)$ of the distances of R_s to each input tree $T_i \in \mathcal{T}_s$ equals $2[n_s(1+n+|E(H)|) - \ell(R_s)]$. Thus,

$$d(\mathcal{T}_s, R_s) = 2[n_s(1+n+|E(H)|) - \ell(R_s)] \quad (12)$$

$$\geq 2[n_s(1+n+|E(H)|) - (n_s(n+1) + \binom{n_s}{2} + |E(H)|)] \quad (13)$$

$$= 2[(n_s - 1) \cdot |E(H)| - \binom{n_s}{2}] \quad (14)$$

The cost of (\mathcal{R}, σ) equals $\sum_{s=1}^k d(\mathcal{T}_s, R_s)$, which we can bound from below as follows.

$$\sum_{s=1}^k d(\mathcal{T}_s, R_s) \geq \sum_{s=1}^k 2[(n_s - 1) \cdot |E(H)| - \binom{n_s}{2}] \quad (15)$$

$$= 2 \left[\left(\sum_{s=1}^k n_s - k \right) \cdot |E(H)| - \sum_{s=1}^k \binom{n_s}{2} \right] \quad (16)$$

$$= 2 \left[(n - k) \cdot |E(H)| - \sum_{s=1}^k \binom{n_s}{2} \right] \quad (17)$$

$$= 2 \left[(c - 1) \cdot |E(H)| - \sum_{s=1}^k \binom{n_s}{2} \right] \quad (18)$$

The bound is tight if and only if $|E_s| = \binom{n_s}{2}$ for every cluster $s \in [k]$, which follows directly from Lemma 3 and Equation (10). \square

(Main Text) Lemma 5. The cost of any clustering $\sigma : [n] \rightarrow k$ of (\mathcal{T}, k) is at least $2[(c-1) \cdot |E(H)| - \binom{c}{2}]$. This bound is tight if and only if σ contains $k-1$ singleton clusters and one cluster with c trees that encode the vertices of a clique in the undirected graph H .

Proof. We prove the lemma by showing that any optimal minimum-cost clustering $\sigma : [n] \rightarrow [k]$ has cost at least $2[(c-1) \cdot |E(H)| - \binom{c}{2}]$. Let $\mathcal{T}_1, \dots, \mathcal{T}_k$ be the sets of input trees that are assigned to the same cluster, i.e. $\mathcal{T}_s = \{T_i \mid \sigma(i) = s\}$. Let $n_s = |\mathcal{T}_s|$. By Lemma 4, the cost of (\mathcal{R}, σ) is at least $2[(c-1) \cdot |E(H)| - \sum_{s=1}^k \binom{n_s}{2}]$.

$$\sum_{s=1}^k d(\mathcal{T}_s, R_s) \geq 2[(c-1) \cdot |E(H)| - \sum_{s=1}^k \binom{n_s}{2}] \quad (19)$$

$$= 2(c-1) \cdot |E(H)| - \sum_{s=1}^k n_s(n_s - 1) \quad (20)$$

$$= 2(c-1) \cdot |E(H)| - \sum_{s=1}^k [(n_s - 1)^2 + (n_s - 1)] \quad (21)$$

$$= 2(c-1) \cdot |E(H)| - \sum_{s=1}^k (n_s - 1)^2 - [(\sum_{s=1}^k n_s) - k] \quad (22)$$

$$= 2(c-1) \cdot |E(H)| - \sum_{s=1}^k (n_s - 1)^2 - (n - k) \quad (23)$$

Observing that $(a_1 + \dots + a_t)^2 \geq a_1^2 + \dots + a_t^2$ provided that $a_i \geq 0$ for all $i \in [t]$, we obtain

$$\sum_{s=1}^k d(\mathcal{T}_s, R_s) \geq 2(c-1) \cdot |E(H)| - \left[\sum_{s=1}^k (n_s - 1) \right]^2 - (n - k) \quad (24)$$

$$= 2(c-1) \cdot |E(H)| - (n - k)^2 - (n - k) \quad (25)$$

$$= 2 \left[(c-1) \cdot |E(H)| - \frac{(n - k)(n - k - 1)}{2} \right] \quad (26)$$

Finally, plugging in $k = n - c + 1$, we obtain

$$\sum_{s=1}^k d(\mathcal{T}_s, R_s) \geq 2[(c-1) \cdot |E(H)| - \binom{c}{2}]. \quad (27)$$

For this bound to be tight, we need Equations (19) and (24) to be tight. That is, we require $\sum_{s=1}^k d(\mathcal{T}_s) = 2[(c-1) \cdot |E(H)| - \sum_{s=1}^k \binom{n_s}{2}]$ and $\sum_{s=1}^k (n_s - 1)^2 = [\sum_{s=1}^k (n_s - 1)]^2$ to be satisfied simultaneously.

By Lemma 4, we know that the first equality is satisfied if and only if the input trees \mathcal{T}_s assigned to each cluster s encode a clique in the undirected graph H . As for the second equality, we observe the following: Provided that $a_i \geq 0$ for all $i \in [t]$, we have $(a_1 + \dots + a_t)^2 = a_1^2 + \dots + a_t^2$ if and only if there exists at most one $i \in [t]$ such that $a_i > 0$. Therefore, there is at most one $s \in [k]$ such that $n_s - 1 > 0$. Without loss of generality, suppose $n_1 - 1 > 0$. Then $n_2 = \dots = n_k = 1$, and $n_1 = n - (n_2 + \dots + n_k) = n - (k - 1) = c$. In other words, the clustering σ contains $k - 1$ singleton clusters and one cluster of size c . Combining the two conditions together, we conclude that this bound is tight if and only if σ contains $k - 1$ singleton clusters and one cluster with c trees that encode the vertices of a clique in the undirected graph H . \square

(Main Text) Lemma 6. There is a clique of size c in the undirected graph H if and only if the corresponding MCT instance (\mathcal{T}, k) has an optimal solution with cost $2[(c-1) \cdot |E(H)| - \binom{c}{2}]$.

Proof. \Rightarrow : By the premise, undirected graph H contains a clique of size c . Without loss of generality, let v_1, \dots, v_c be distinct vertices that form a clique in H . Let (\mathcal{T}, k) be the MCT instance corresponding to CLIQUE instance (H, c) . We construct a clustering $\sigma : [n] \rightarrow [k]$ as follows.

$$\sigma(i) = \begin{cases} 1, & 1 \leq i \leq c, \\ i - c, & c < i \leq n. \end{cases} \quad (28)$$

We apply Theorem 1, to obtain $\mathcal{R} = \{R_1, \dots, R_k\}$ from \mathcal{T} and σ . By Lemma 5, we have that the cost of (\mathcal{R}, σ) is $2[(c-1) \cdot |E(H)| - \binom{c}{2}]$. By the same lemma, we have that the cost of any feasible solution (\mathcal{R}', σ') to the instance (\mathcal{T}, k) is at least $2[(c-1) \cdot |E(H)| - \binom{c}{2}]$. Hence, (\mathcal{R}, σ) is an optimal solution.

\Leftarrow : We prove the contrapositive: Given that there is no clique of size c in H , there is no optimal solution to MCT instance (\mathcal{T}, k) with cost $2[(c-1) \cdot |E(H)| - \binom{c}{2}]$. This follows directly from Lemma 5. That is, the absence of a clique of size c in H implies that there is no clustering σ with cost $2[(c-1) \cdot |E(H)| - \binom{c}{2}]$. \square

A.2 Methods

A.2.1 Mixed Integer Linear Program

We have the following MILP.

$$\begin{aligned}
& \min n(m-1) - \sum_{i=1}^n \sum_{s=1}^k \sum_{p=1}^m \sum_{q=1}^m w_{i,s,p,q} \\
& \text{s.t.} \quad \sum_{s=1}^k x_{i,s} = 1 && \forall i \in [n] \\
& \quad \sum_{i=1}^n x_{i,s} \geq 1 && \forall s \in [k] \\
& \quad \sum_{p=1}^m z_{s,p} = 1 && \forall s \in [k] \\
& \quad \sum_{q=1}^m y_{s,p,q} = 1 - z_{s,p} && \forall s \in [k], p \in [m] \\
& \quad y_{s,p,q} \leq b_{p,q} && \forall s \in [k], p, q \in [m] \\
& \quad \sum_{(p,q) \in \delta^-(U)} y_{s,p,q} + \sum_{p \in U} z_{s,p} \geq 1 && \forall s \in [k], U \subseteq [m] \\
& \quad w_{i,s,p,q} \leq a_{i,p,q} && \forall i \in [n], s \in [k], p, q \in [m] \\
& \quad w_{i,s,p,q} \leq x_{i,s} && \forall i \in [n], s \in [k], p, q \in [m] \\
& \quad w_{i,s,p,q} \leq y_{s,p,q} && \forall i \in [n], s \in [k], p, q \in [m] \\
& \quad w_{i,s,p,q} \geq 0 && \forall i \in [n], s \in [k], p, q \in [m] \\
& \quad y_{s,p,q} \leq \sum_{i=1}^n a_{i,p,q} x_{i,s} && \forall s \in [k], p, q \in [m] \\
& \quad y_{s,p,q} \geq \sum_{i=1}^n a_{i,p,q} x_{i,s} - \sum_{i=1}^n x_{i,s} + 1 && \forall s \in [k], p, q \in [m] \\
& \quad \sum_{i=1}^n x_{i,s} \geq \sum_{i=1}^n x_{i,s+1} + 1 && \forall s \in [k-1] \\
& \quad x_{i,s} \in \{0, 1\} && \forall i \in [n], s \in [k] \\
& \quad y_{s,p,q} \geq 0 && \forall s \in [k], p, q \in [m] \\
& \quad z_{s,p} \geq 0 && \forall s \in [k], p \in [m]
\end{aligned}$$

A.3 Supplementary Results

We have the following supplementary figures and tables.

Table S1: **Characteristics of simulated instances.** The table shows the number of simulated instances for varying combinations of numbers of mutation clusters and bulk samples. The mean number of trees for each combination is shown in brackets.

#bulk samples	#mutation clusters			total
	$m = 9$	$m = 11$	$m = 13$	
5	9 [30]	8 [38]	10 [138]	27 [72]
10	8 [10]	4 [9]	6 [10]	18 [9]
total	17 [20]	12 [28]	16 [89]	45 [47]

Table S2: **Number of instances solved to optimality.** For each set of instances (small, medium or large) and number $k \in \{1, \dots, 5\}$ of clusters, the table shows the number of instances solved to optimality by the mixed integer linear program (MILP 1 h) and the brute force algorithm (BF 1 h), each restricted to a running time of 1 hour. The instances that were solved to optimality by MILP include all instances solved to optimality by BF. In addition, the table shows the number of instances for which the coordinate ascent algorithm identified the same optimal solutions as the MILP using a time limit of 1 hour (CA 1 h) or restricted to 100 restarts (CA 100 r.).

	#clusters k	MILP (1 h)	BF (1 h)	CA (1 h)	CA (100 r.)
small (16)	2	16	16	16	16
	3	16	16	16	16
	4	16	16	16	16
	5	16	14	16	16
medium (15)	2	15	13	15	15
	3	13	7	13	13
	4	12	0	12	12
	5	10	0	10	10
large (14)	2	3	0	3	3
	3	0	0	0	0
	4	0	0	0	0
	5	0	0	0	0

- Table S1: Characteristics of simulated instances.
- Table S2: Number of instances solved to optimality.
- Fig. S1: The optimality gap of MILP increases with increasing number n of input trees and number k of clusters.
- Fig. S2: Set \mathcal{T} of input trees in a simulated instance with $n = |\mathcal{T}| = 9$ trees.
- Fig. S3: The identified number k of clusters increases with increasing number n of input trees.
- Fig. S4: Set \mathcal{T} of input trees in lung cancer patient CRUK0013 with $n = |\mathcal{T}| = 8$ trees.
- Fig. S5: Set \mathcal{T} of input trees in lung cancer patient CRUK0037 with $n = |\mathcal{T}| = 17$ trees.

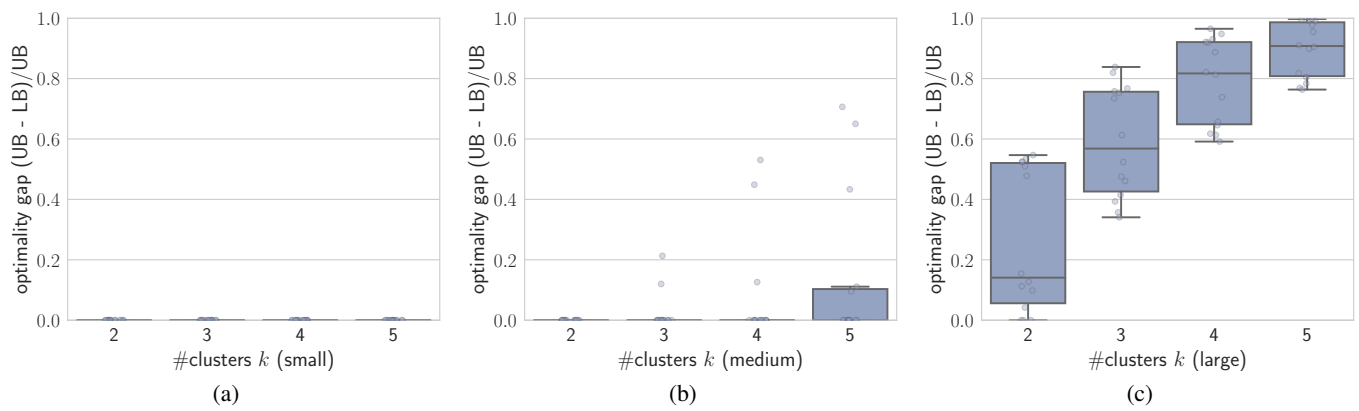
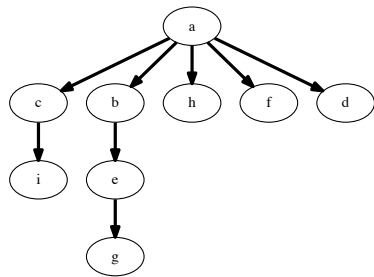
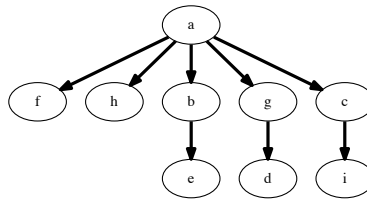


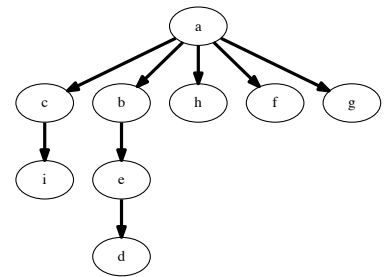
Figure S1: **The optimality gap of MILP increases with increasing number n of input trees and number k of clusters.** LB indicates lower bound of MILP, corresponding to a feasible (integral) solution, whereas UB indicates the upper bound which is the objective value of the linear programming relaxation. The optimality gap is defined as $(UB-LB)/LB$. Thus, an optimality gap of 0 indicates an optimal solution (i.e. $LB=UB$). All small instances (a) were solved to optimality, as opposed to medium (b) and large (c) instances.



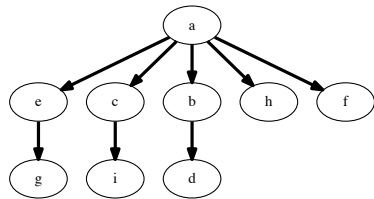
(a)



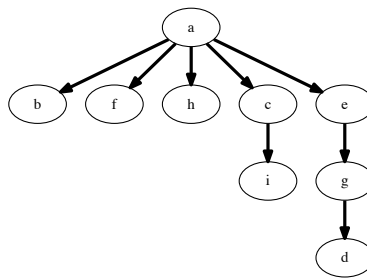
(b)



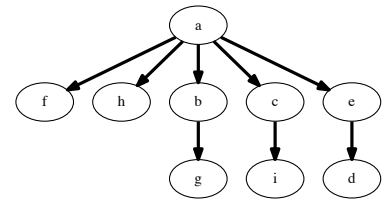
(c)



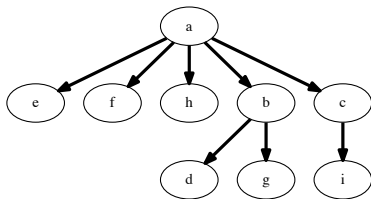
(d)



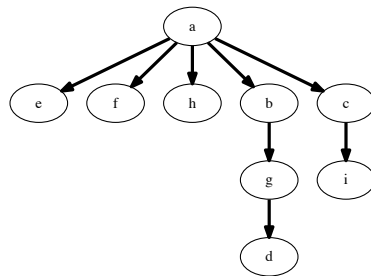
(e)



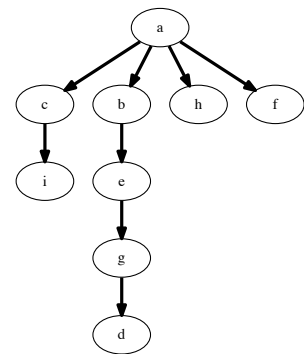
(f)



(g)



(h)



(i)

Figure S2: Set \mathcal{T} of input trees in a simulated instance with $n = |\mathcal{T}| = 9$ trees.

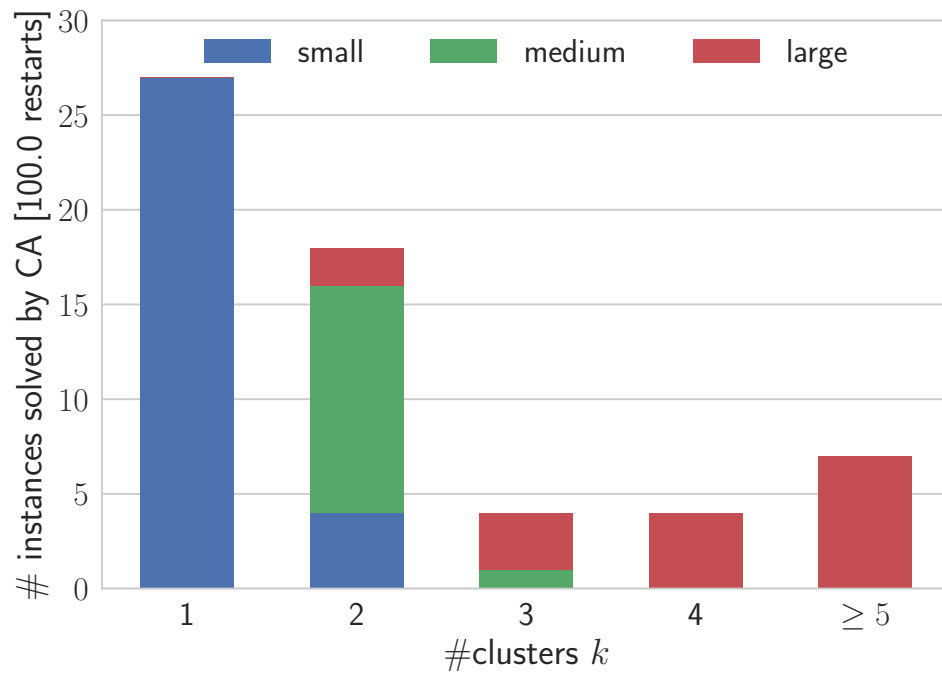


Figure S3: **The identified number k of clusters increases with increasing number n of input trees.** The number of simulated instances with k selected by BIC, colors indicate the size of the solution space. See Main Text Fig. 3a for the distribution of number of input trees for each category.

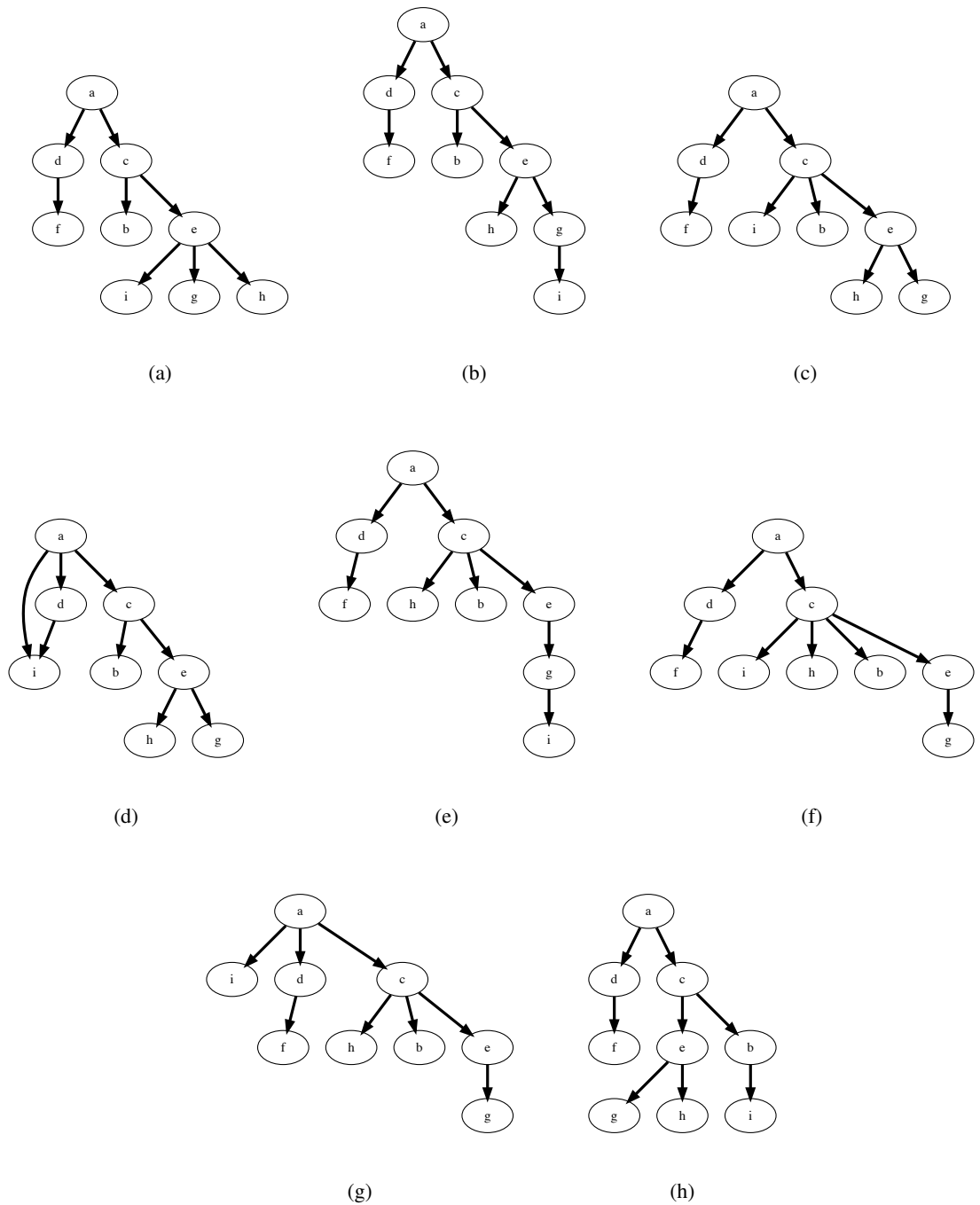
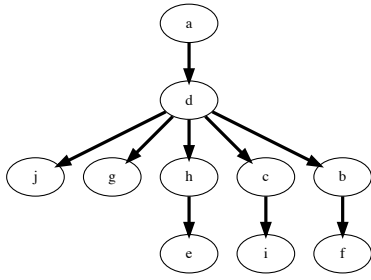
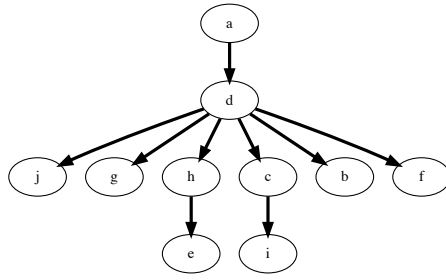


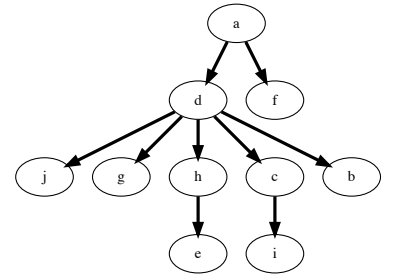
Figure S4: Set \mathcal{T} of input trees in lung cancer patient CRUK0013 with $n = |\mathcal{T}| = 8$ trees.



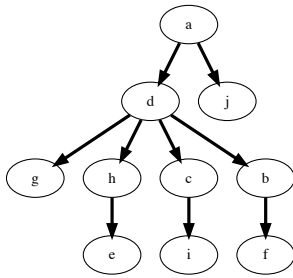
(a)



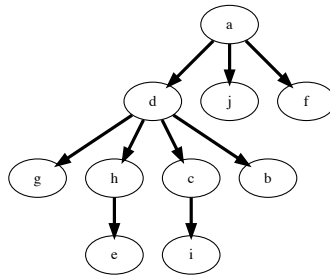
(b)



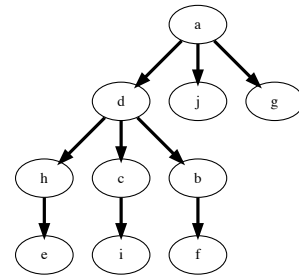
(c)



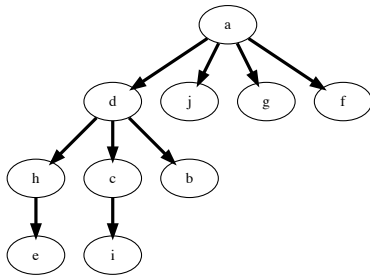
(d)



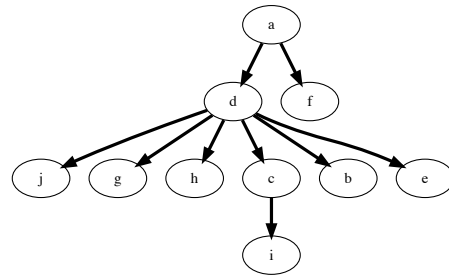
(e)



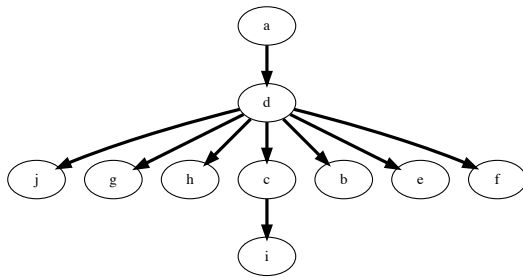
(f)



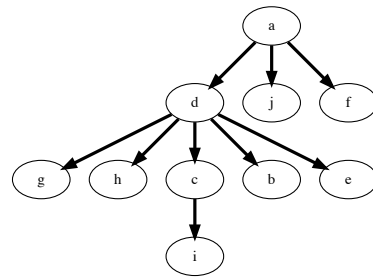
(g)



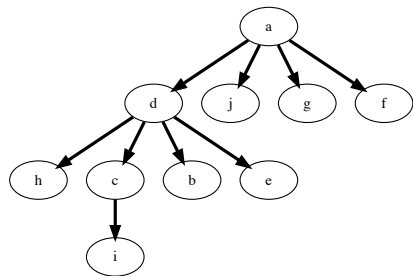
(h)



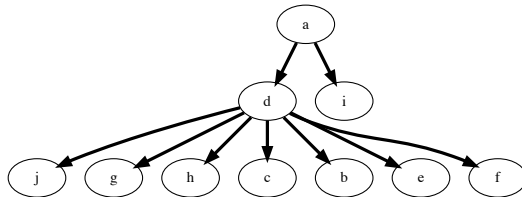
(i)



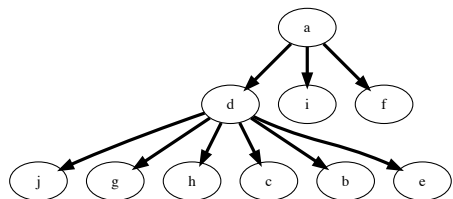
(j)



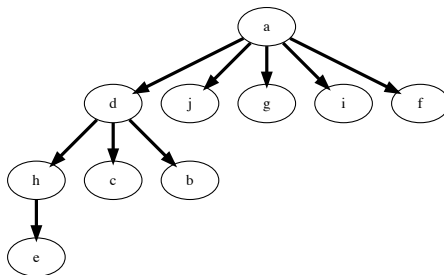
(k)



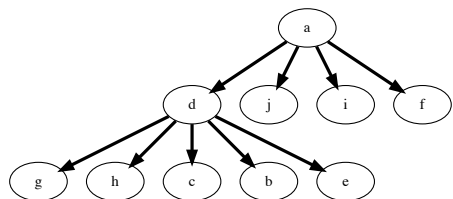
(l)



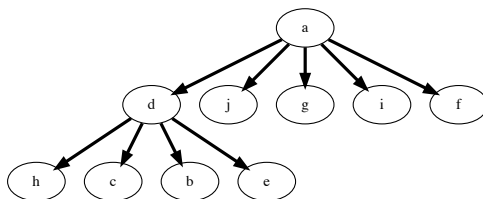
(m)



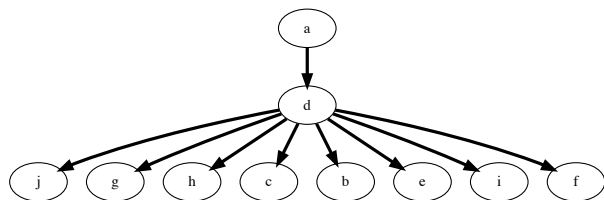
(n)



(o)



(p)



(q)

Figure S5: Set \mathcal{T} of input trees in lung cancer patient CRUK0037 with $n = |\mathcal{T}| = 17$ trees.