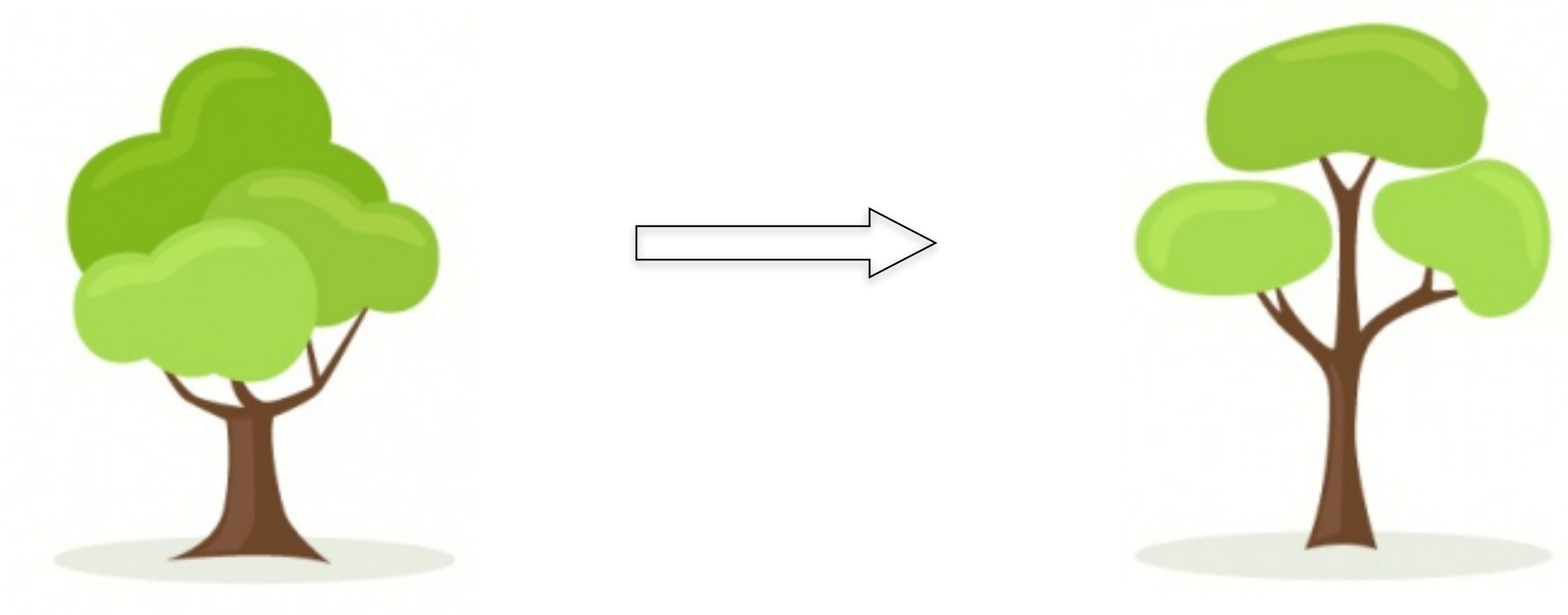
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Karl Bringmann,
Paweł Gawrychowski,
Shay Mozes,
Oren Weimann

Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

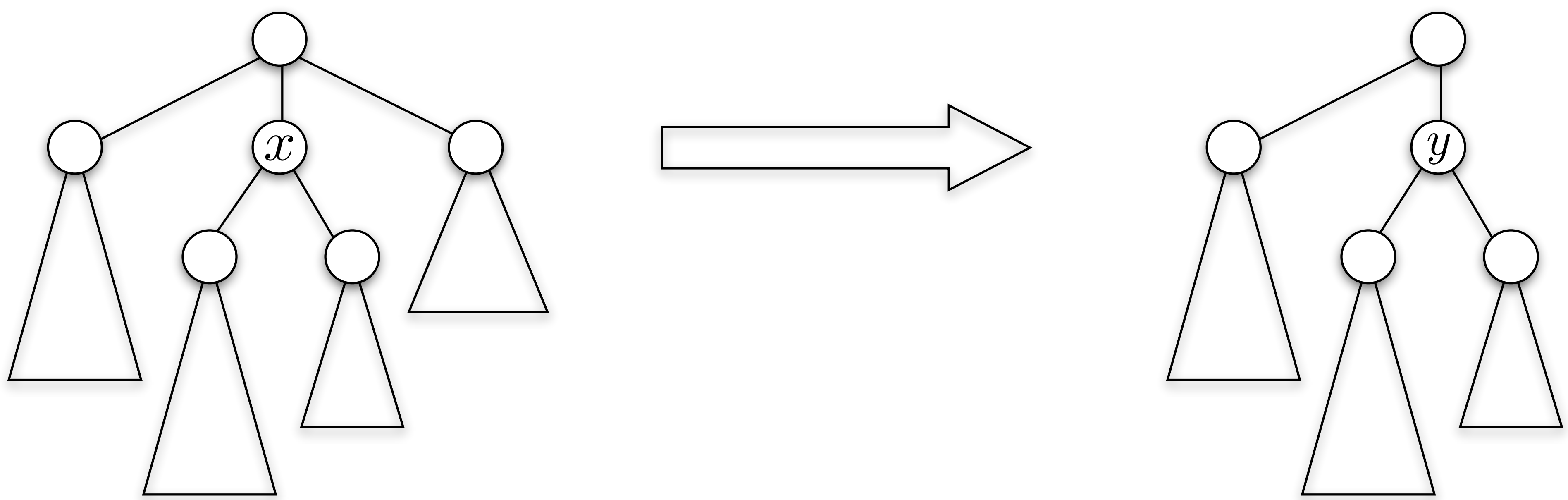
Minimum edits to transform one tree into the other



Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

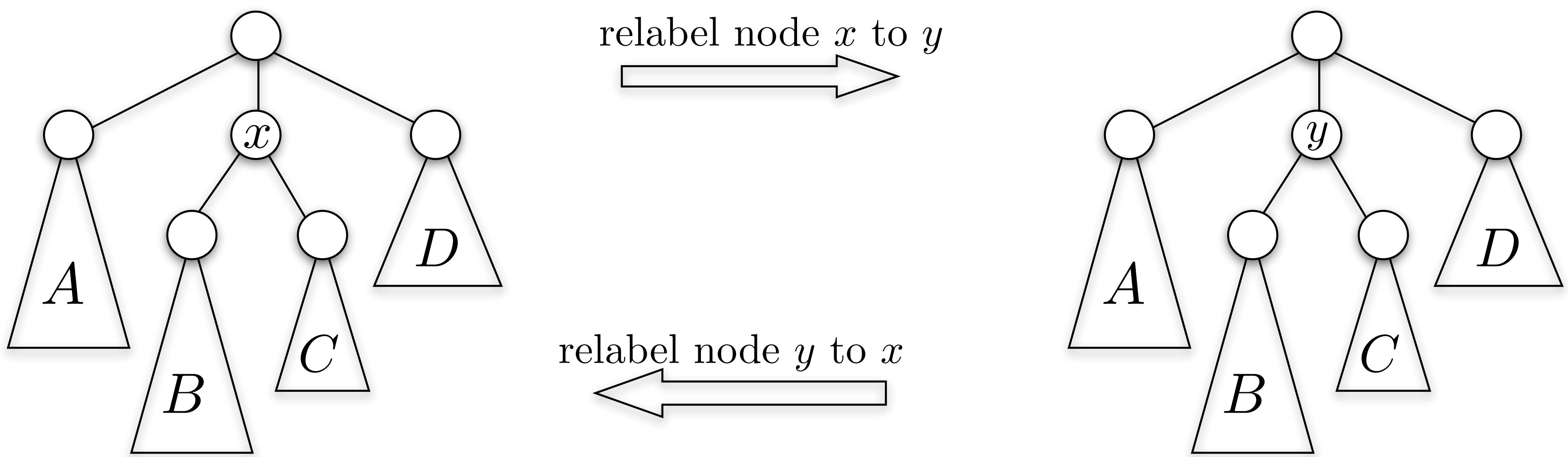
Minimum edits to transform one tree into the other

rooted, ordered trees with node labels



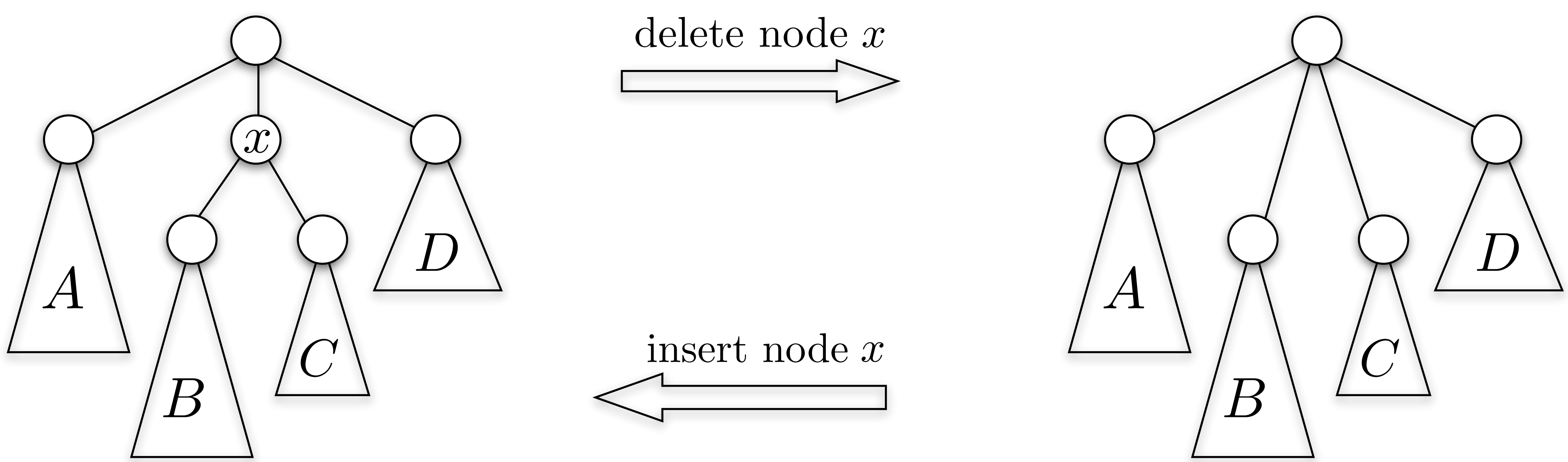
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one tree into the other



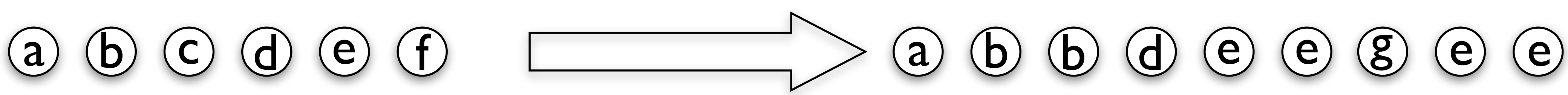
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one tree into the other



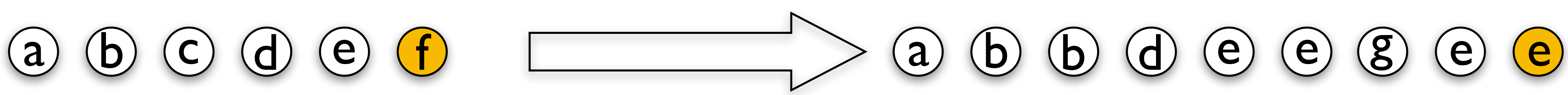
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one string into the other



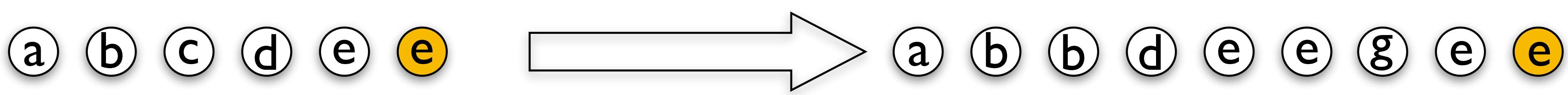
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one string into the other



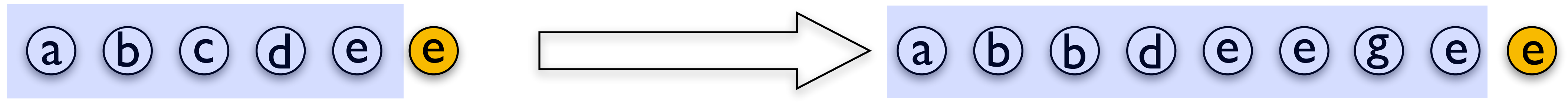
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one string into the other



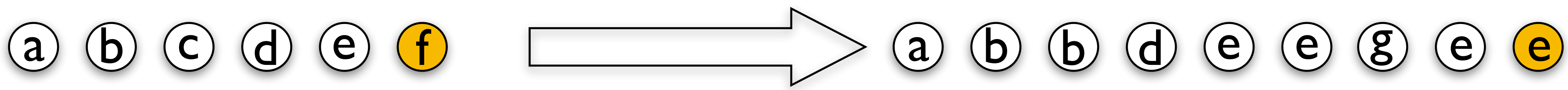
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one string into the other



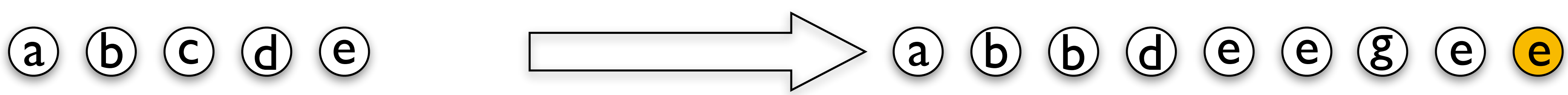
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one string into the other



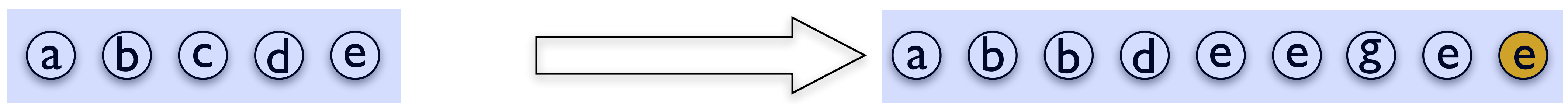
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one string into the other



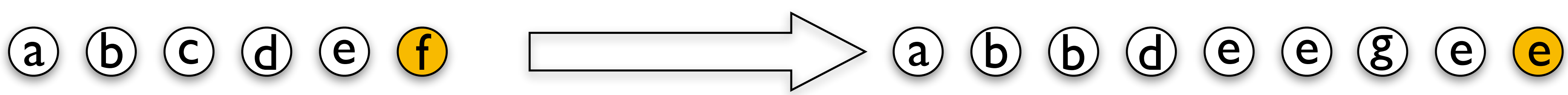
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one string into the other



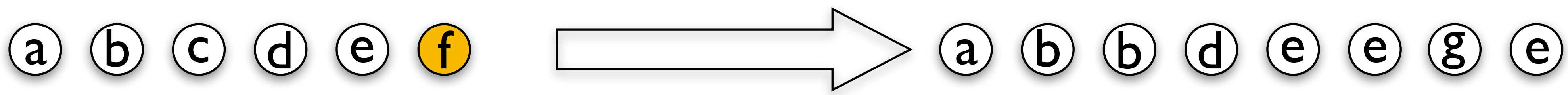
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one string into the other



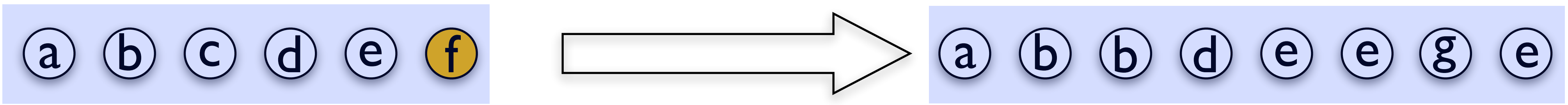
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one string into the other



Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

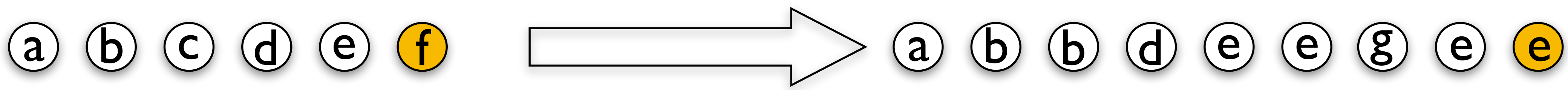
Minimum edits to transform one string into the other



Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one string into the other

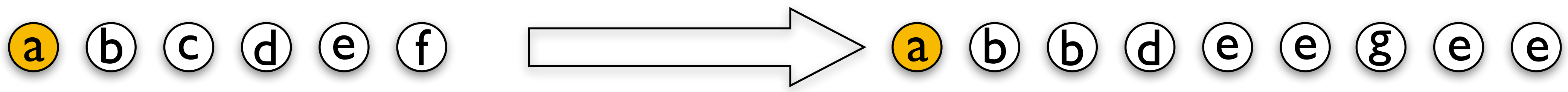
$O(n^2)$ time



Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one string into the other

$O(n^2)$ time

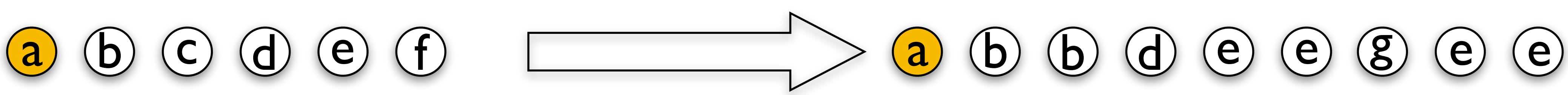


Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Minimum edits to transform one string into the other

$O(n^2)$ time

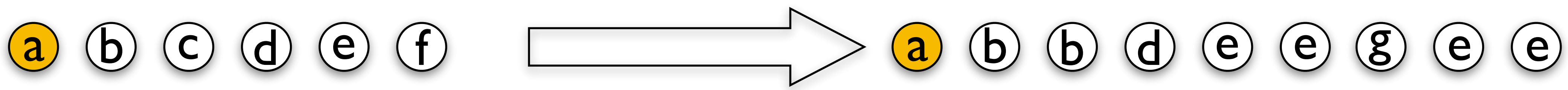
$O(n^4)$ time



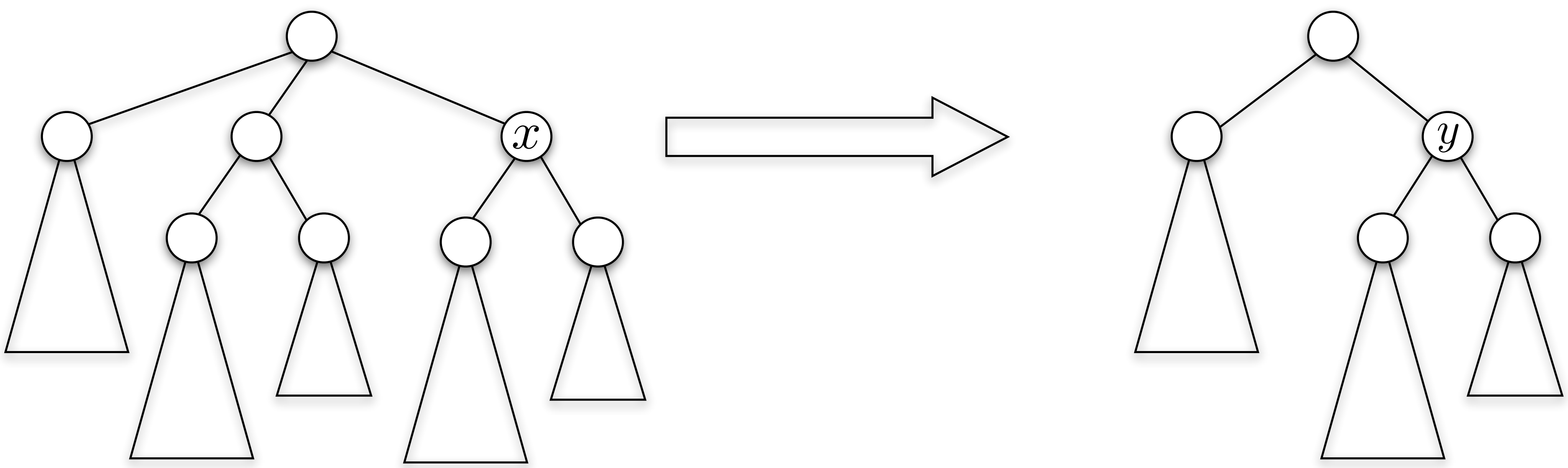
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

String Edit Distance Cannot be Computed in Strongly Subquadratic Time (unless SETH is false)

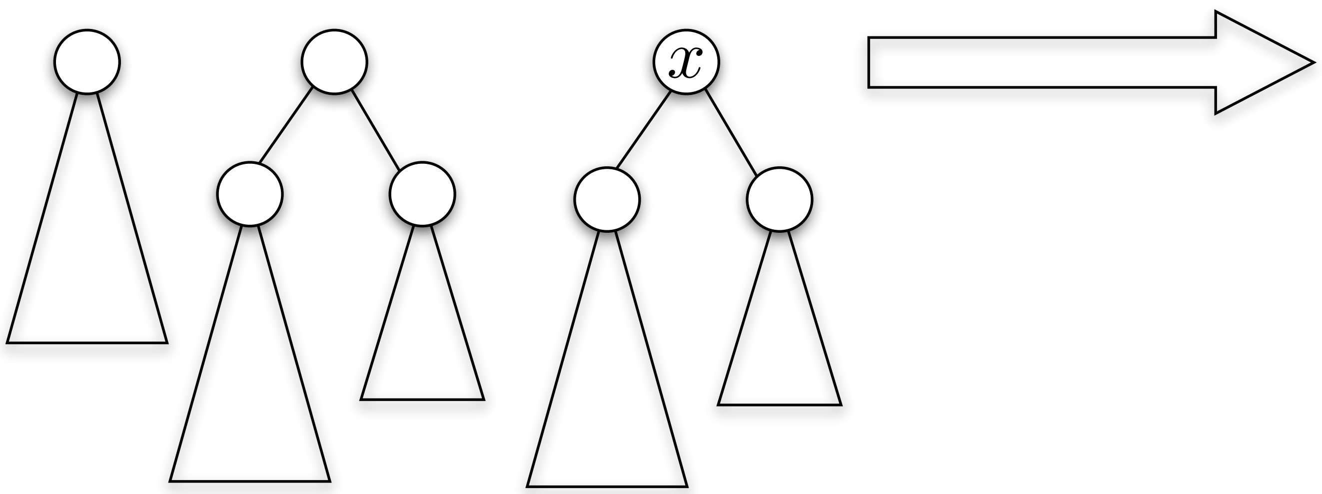
[Backurs,Indyk, STOC'15]



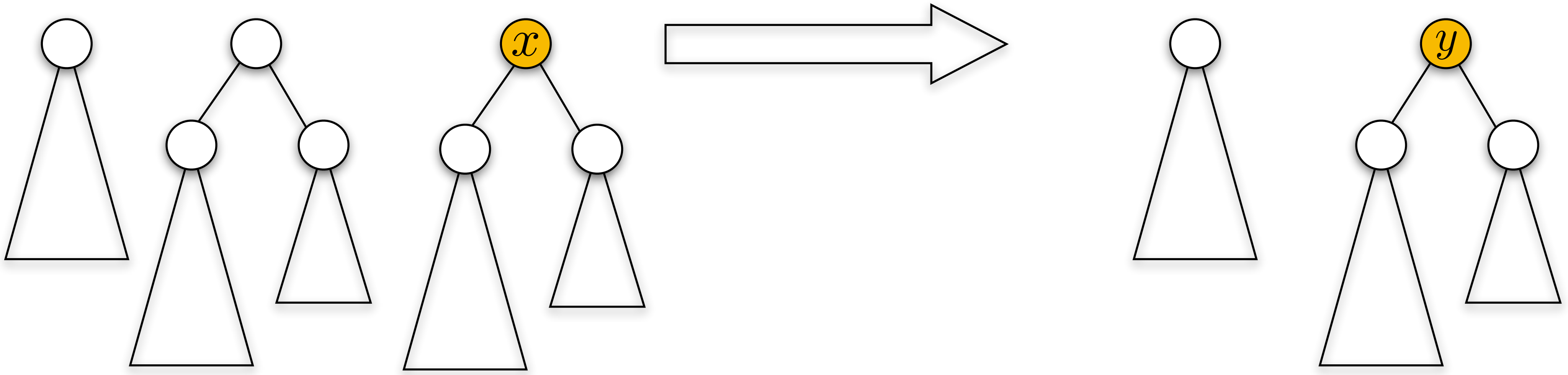
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



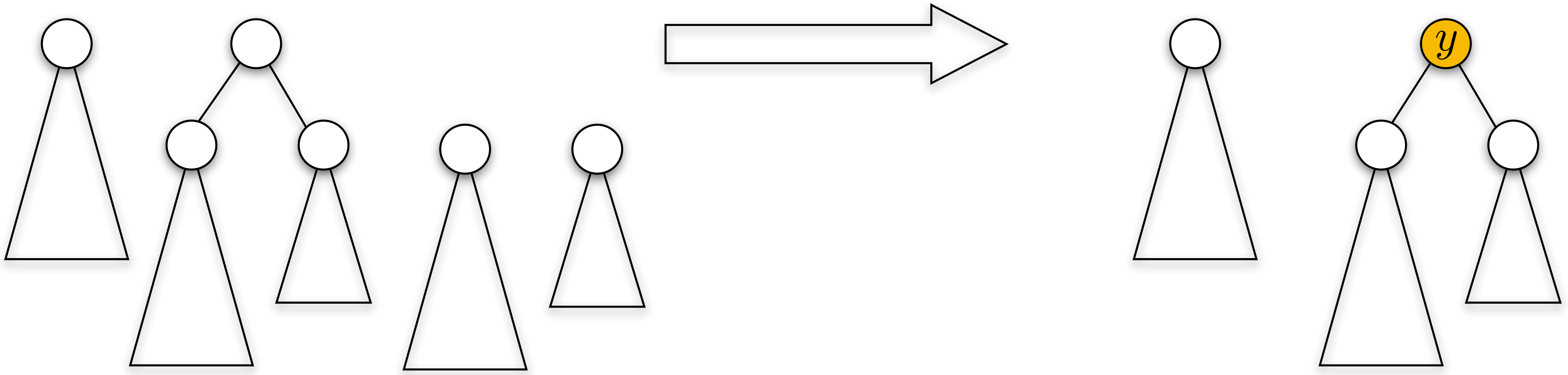
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



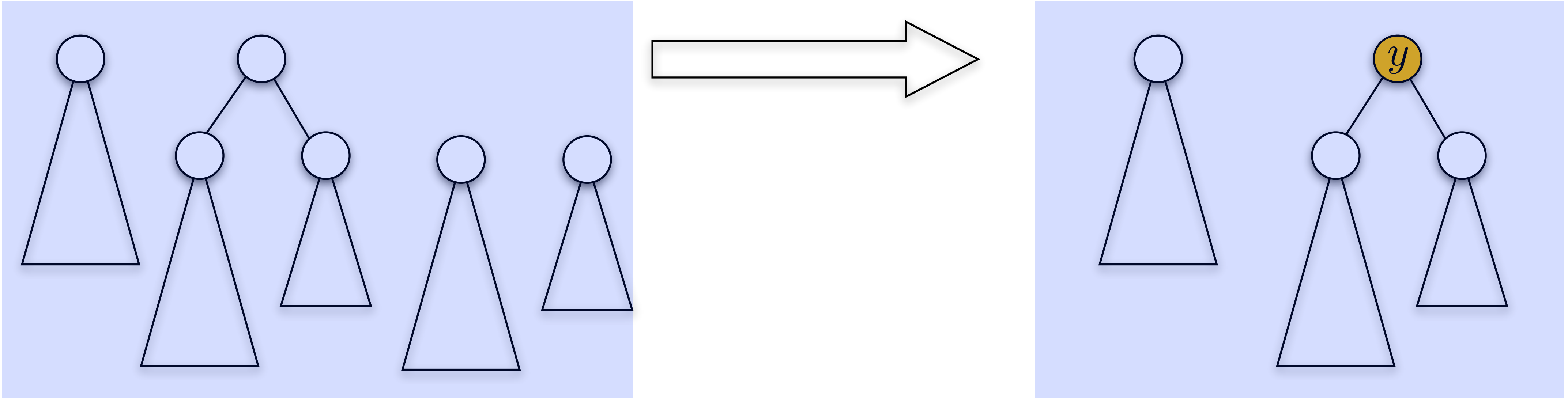
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



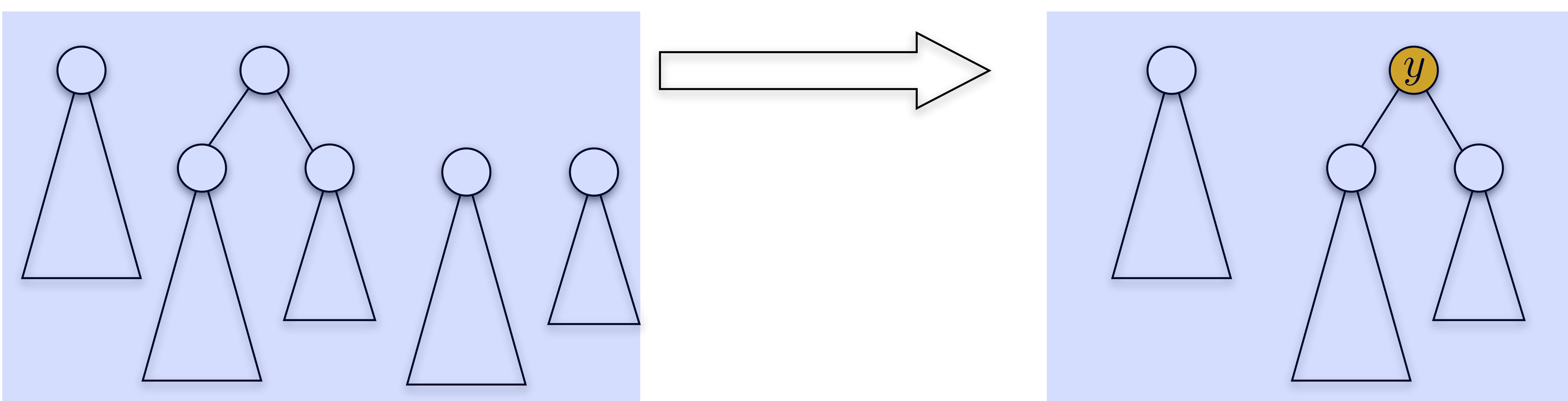
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



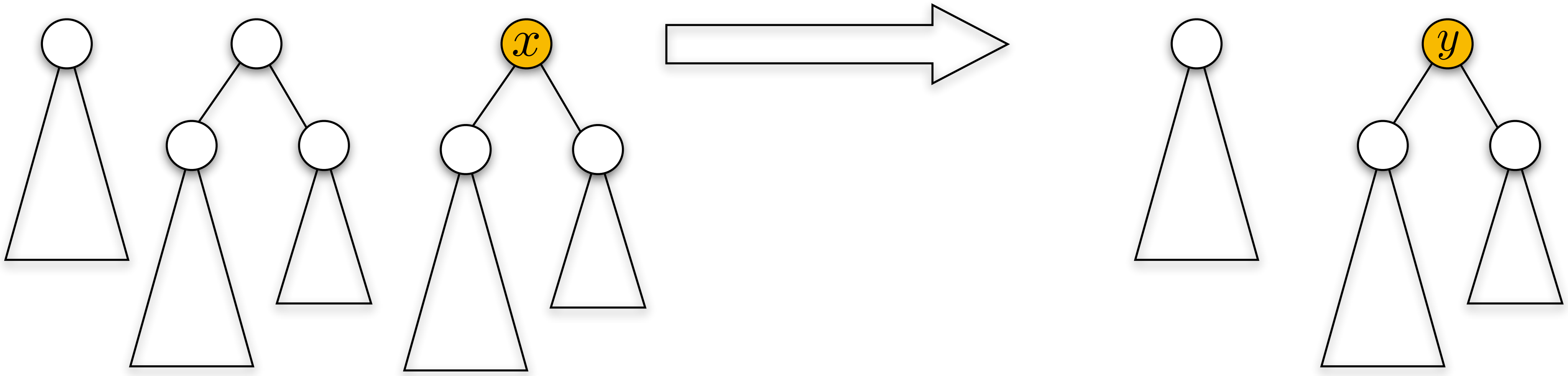
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



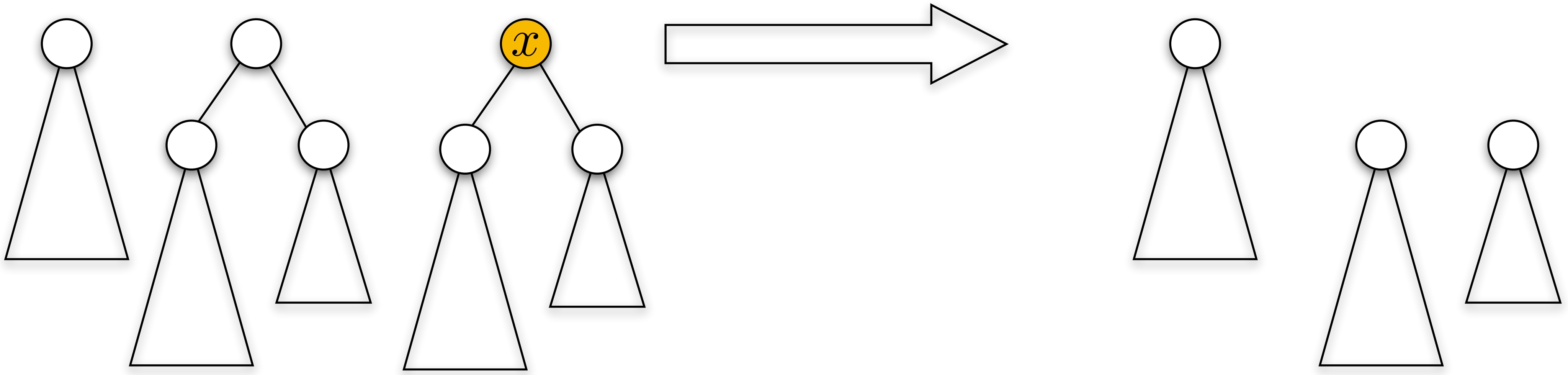
prefix in postorder traversal

prefix in postorder traversal

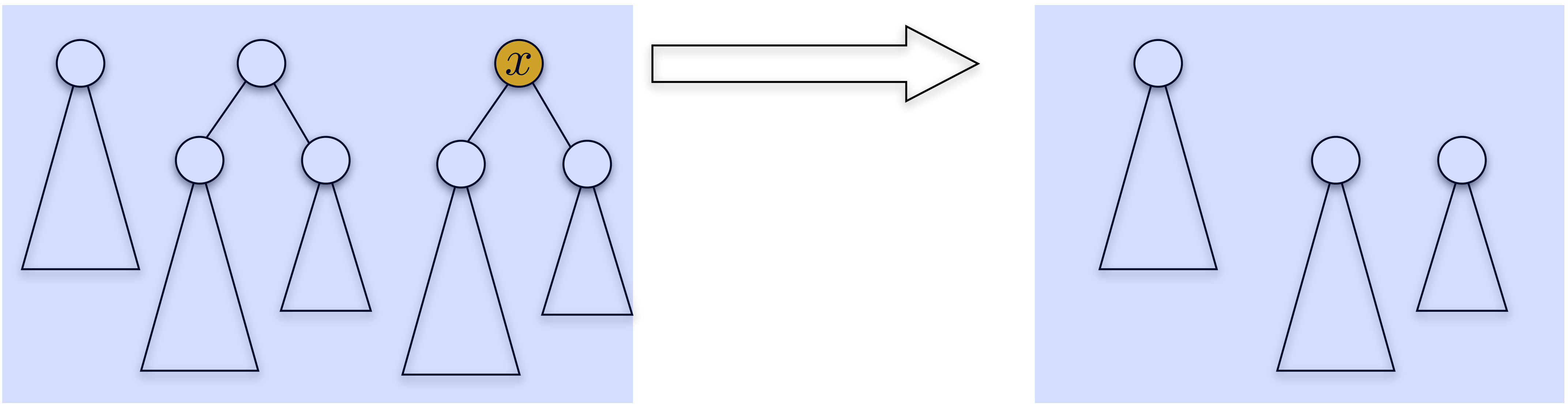
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



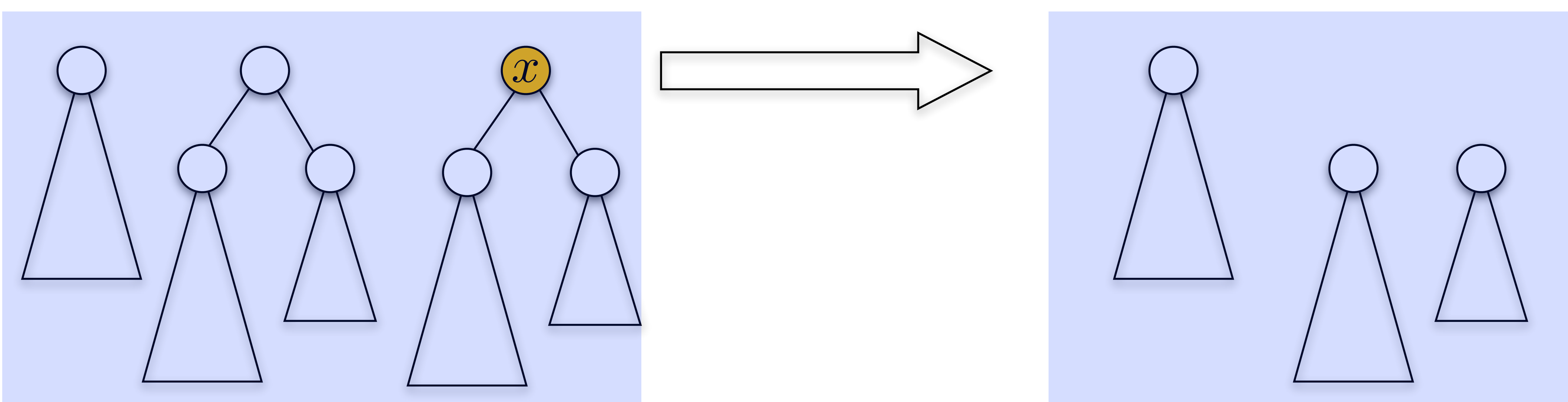
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



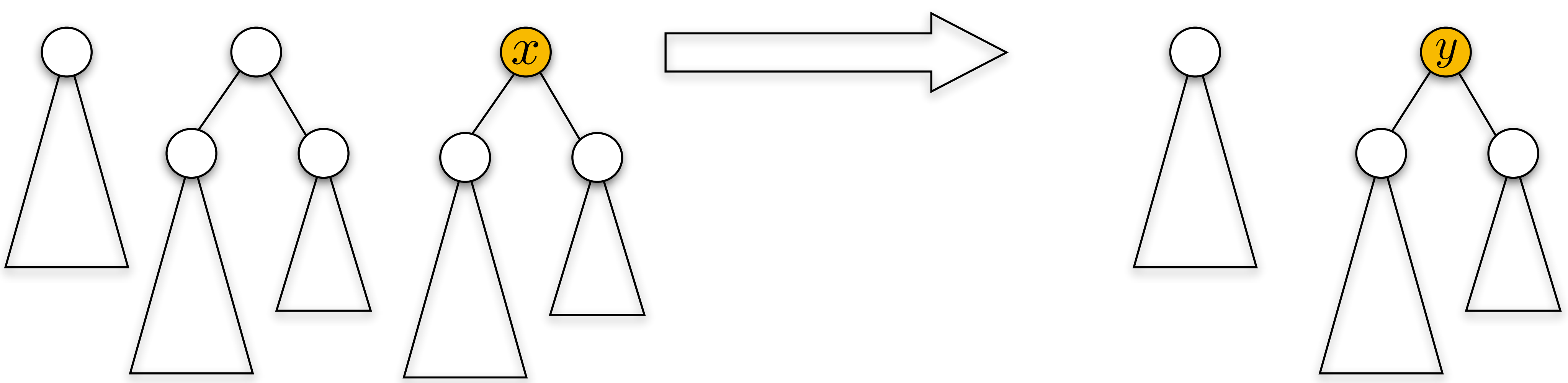
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



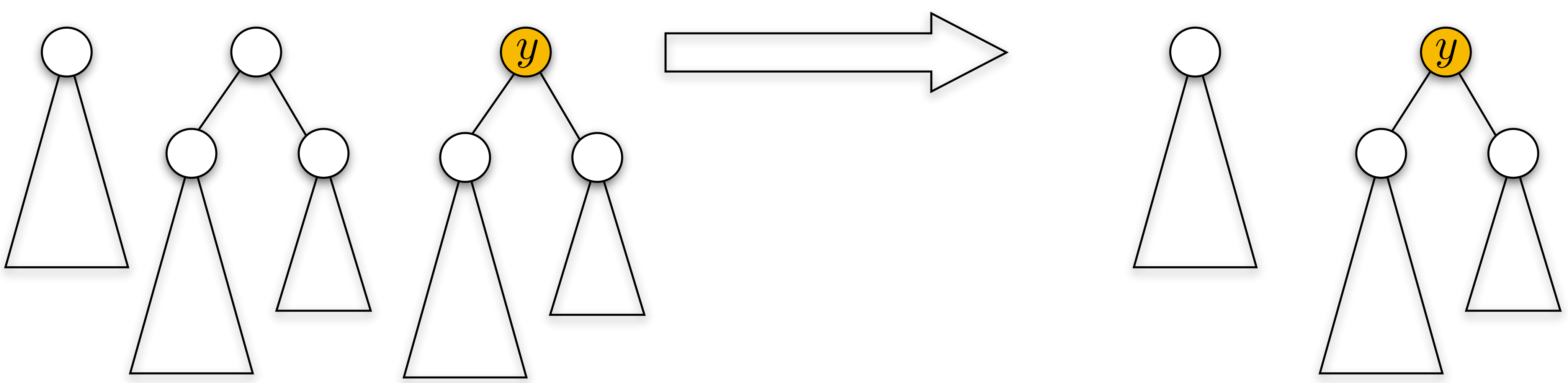
prefix in postorder traversal

prefix in postorder traversal

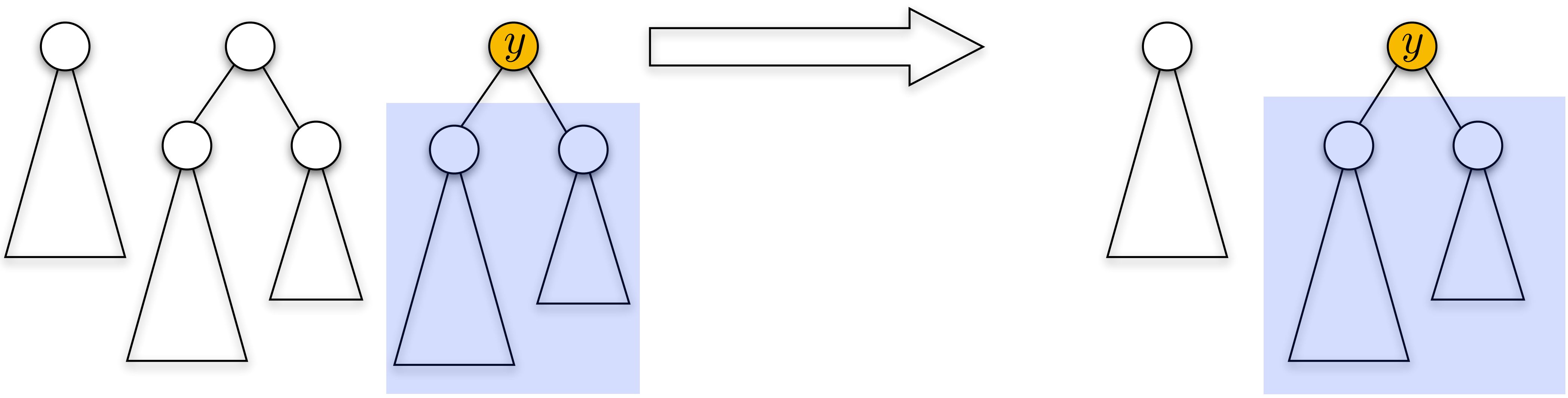
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



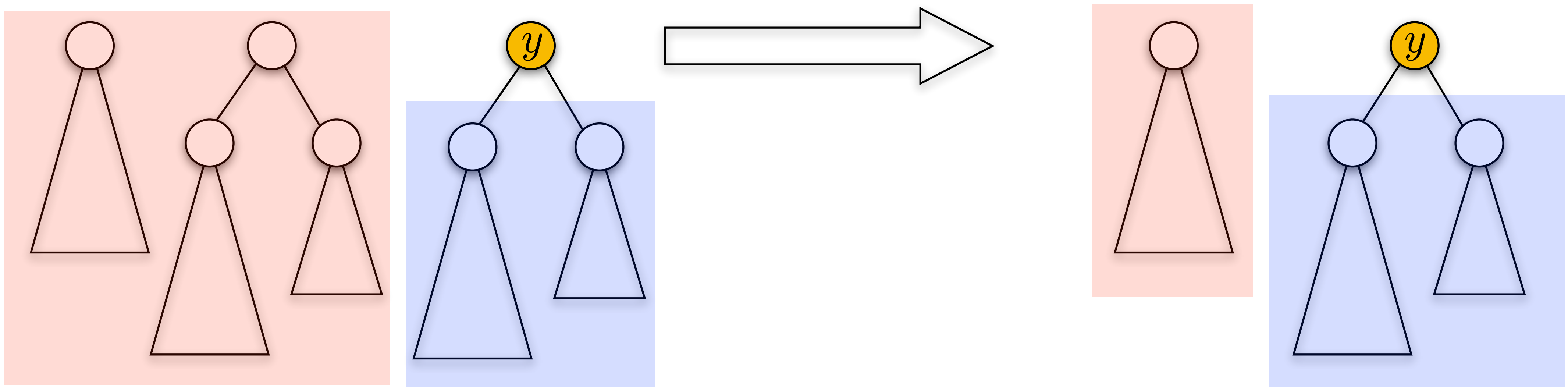
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



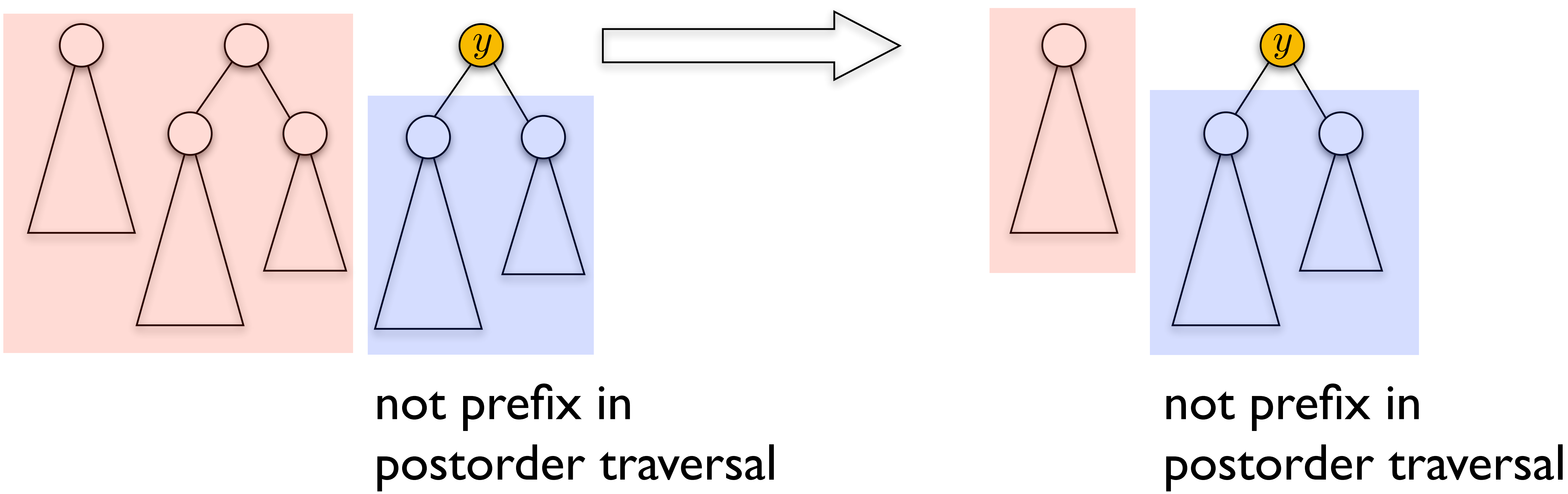
Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

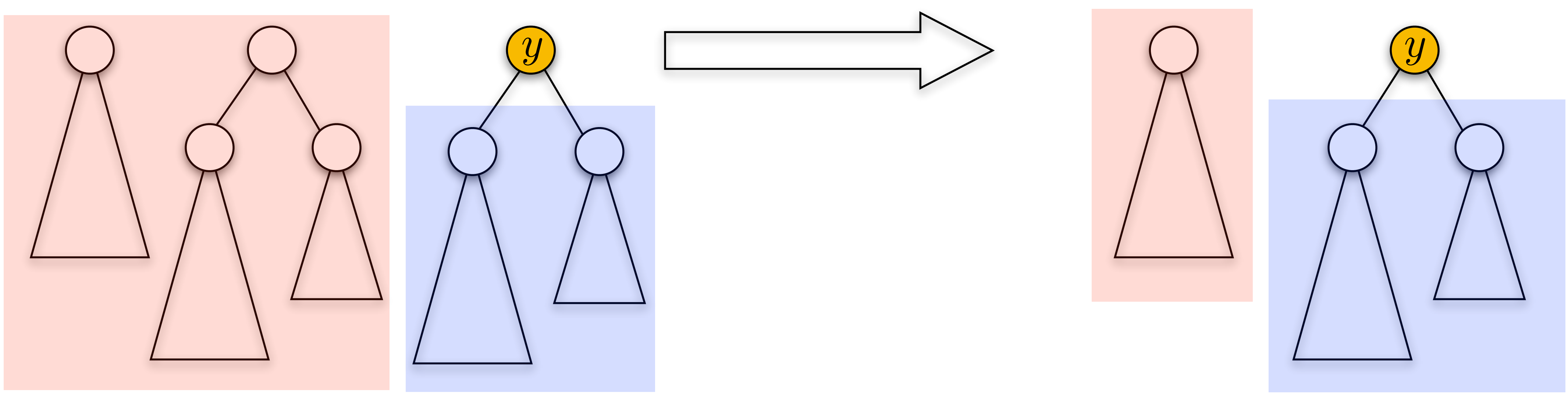


Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)



Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

$O(n^4)$ time [Shasha Zhang 1989]



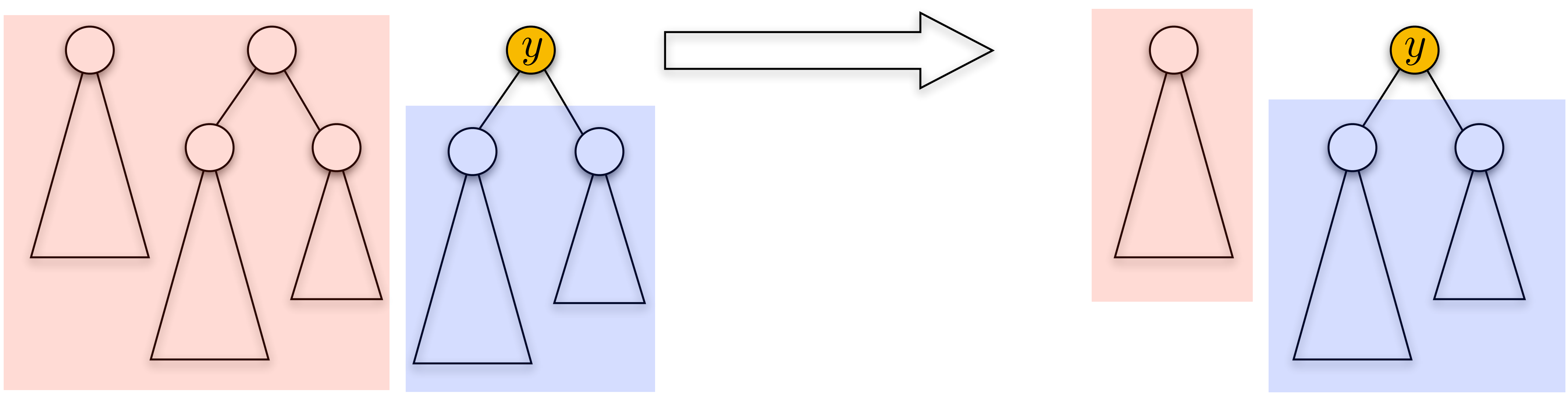
not prefix in
postorder traversal

not prefix in
postorder traversal

Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

$O(n^4)$ time [Shasha Zhang 1989]

$O(n^3 \log n)$ time [Klein 1998]



not prefix in postorder traversal

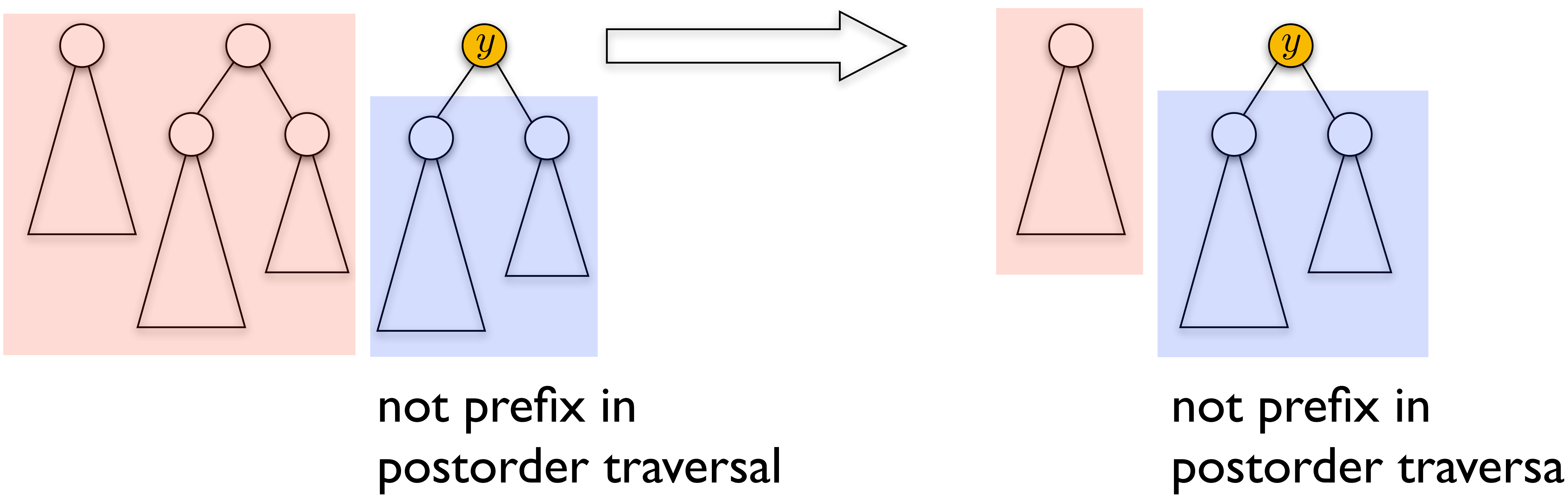
not prefix in postorder traversal

Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

$O(n^4)$ time [Shasha Zhang 1989]

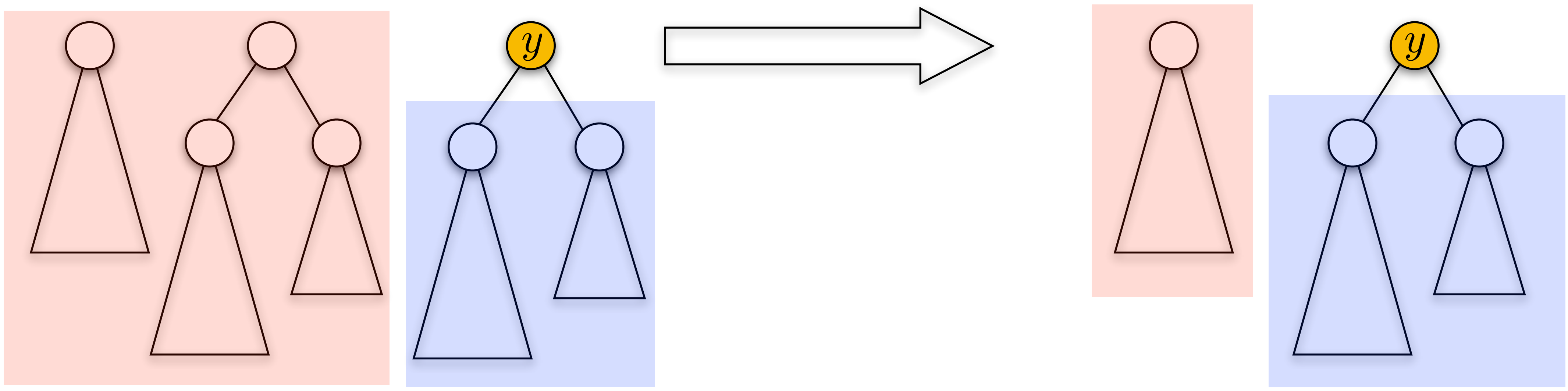
$O(n^3 \log n)$ time [Klein 1998]

$O(n^3)$ time [Demaine, Mozes, Rossman, W. 2007]



Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

$O(n^4)$ time	[Shasha Zhang 1989]
$O(n^3 \log n)$ time	[Klein 1998]
$O(n^3)$ time	[Demaine, Mozes, Rossman, W. 2007]



Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Conjecture (APSP):

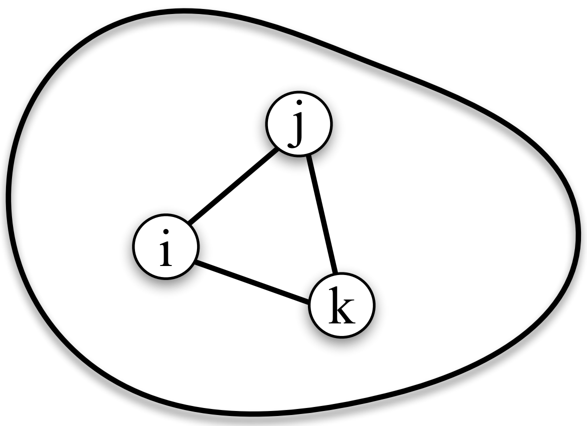
For any $\varepsilon > 0$ there exists $c > 0$, such that All Pairs Shortest Paths on n node graphs with edge weights in $\{1, \ldots, n^c\}$ cannot be solved in $O(n^{3-\varepsilon})$ time.

Tree Edit Distance Cannot be Computed in Strongly Subcubic Time (unless APSP can)

Conjecture (APSP):

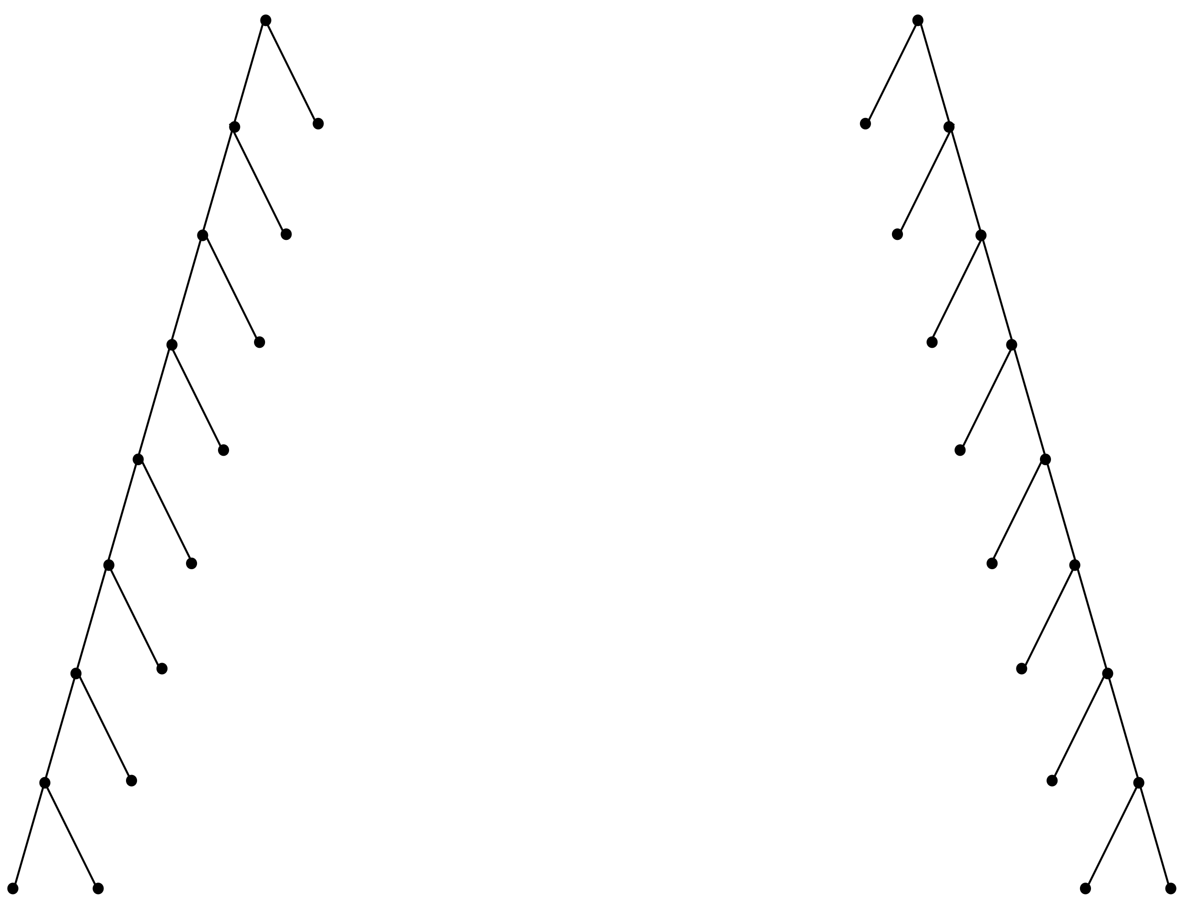
For any $\varepsilon > 0$ there exists $c > 0$, such that All Pairs Shortest Paths on n node graphs with edge weights in $\{1, \ldots, n^c\}$ cannot be solved in $O(n^{3-\varepsilon})$ time.

Equivalent to negative triangle detection [Vassilevska-Williams,Williams 2010]



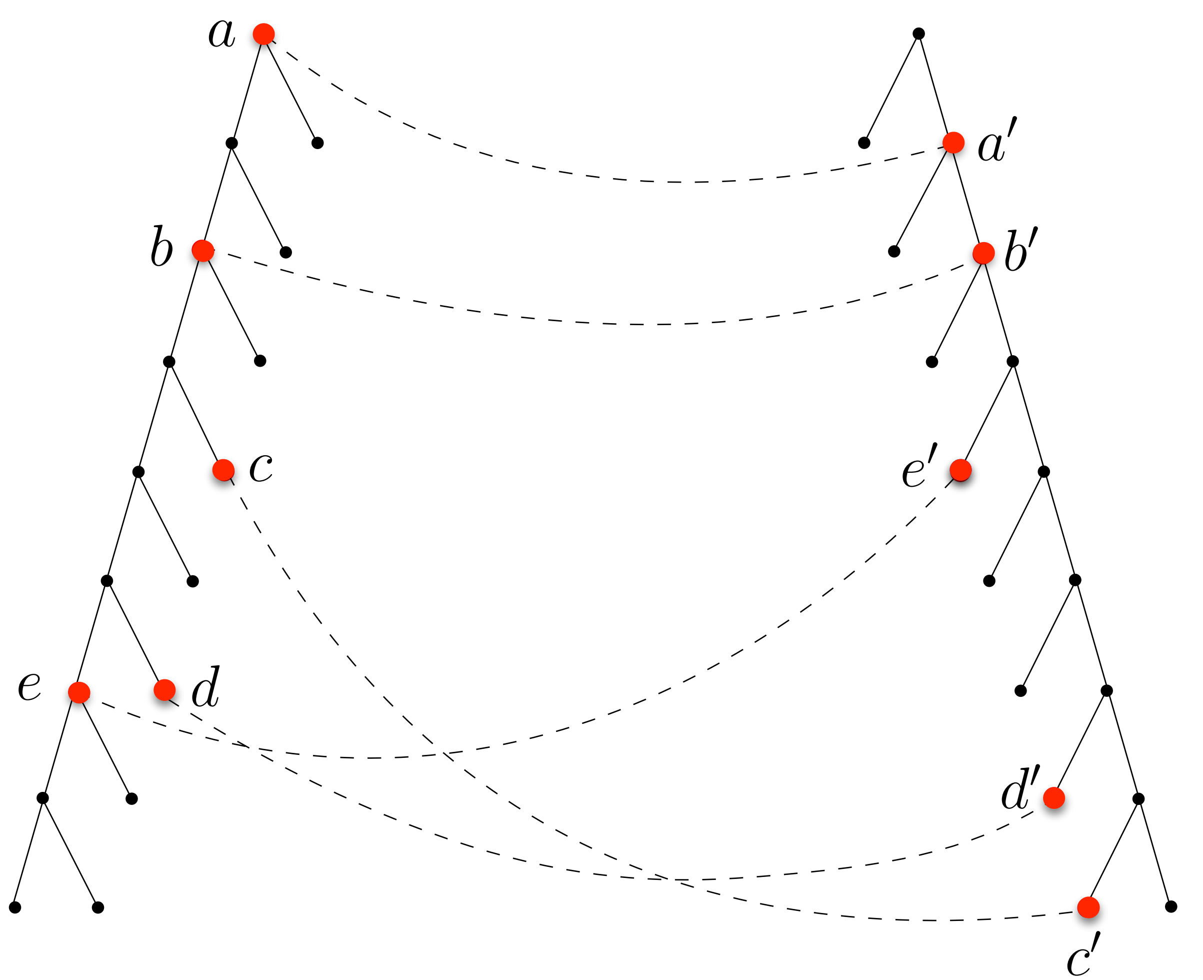
$$w(i,j)+w(j,k)+w(k,i) < 0$$

The Hard Instance of TED



The Hard Instance of TED

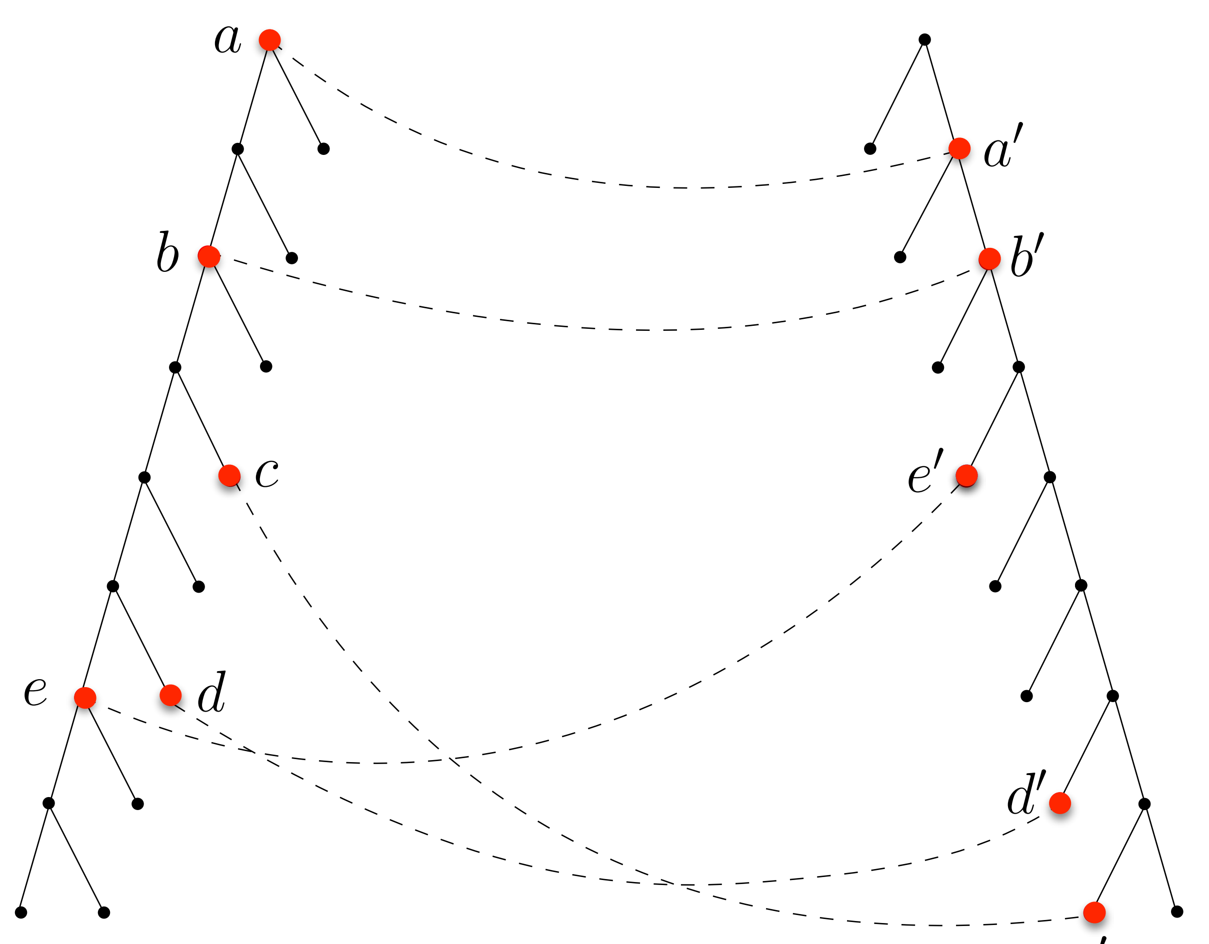
To specify a solution to TED it is enough to say which nodes are matched to which (the rest are deleted)



The Hard Instance of TED

To specify a solution to TED it is enough to say which nodes are matched to which (the rest are deleted)

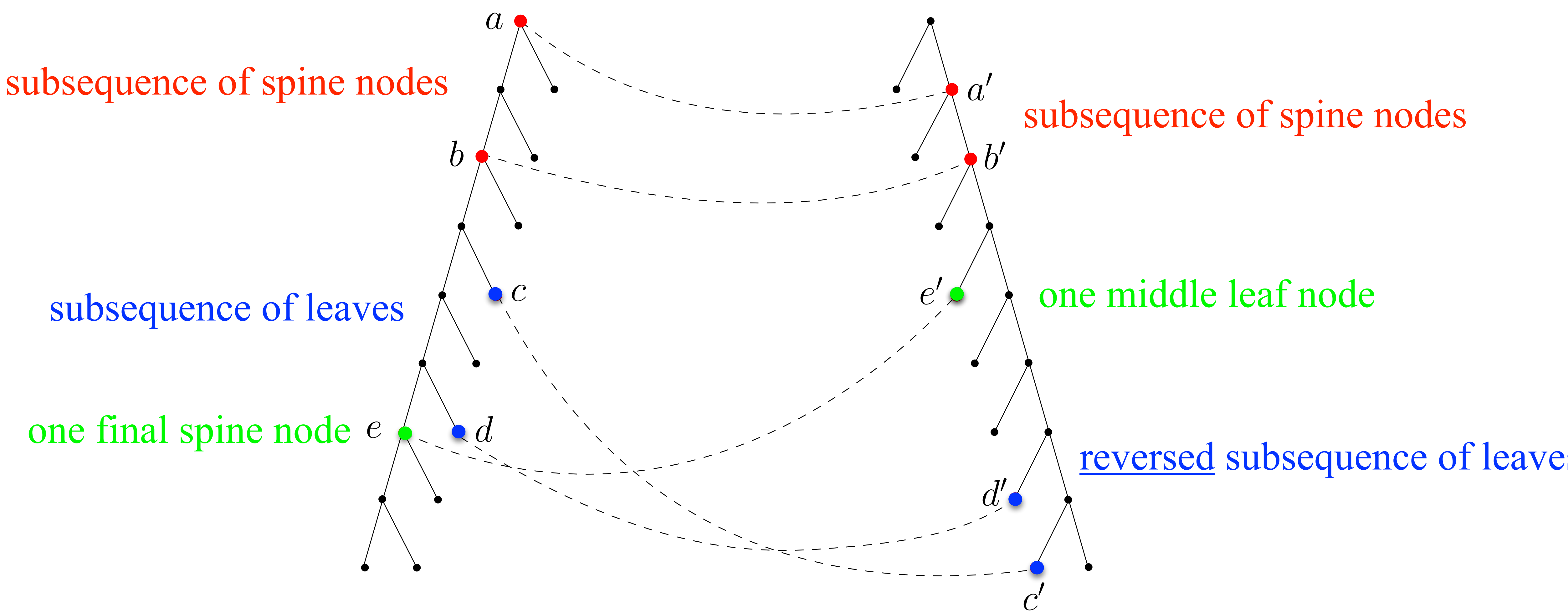
The matched pairs must have the same left-right and ancestor-descendant relations



The Hard Instance of TED

To specify a solution to TED it is enough to say which nodes are matched to which (the rest are deleted)

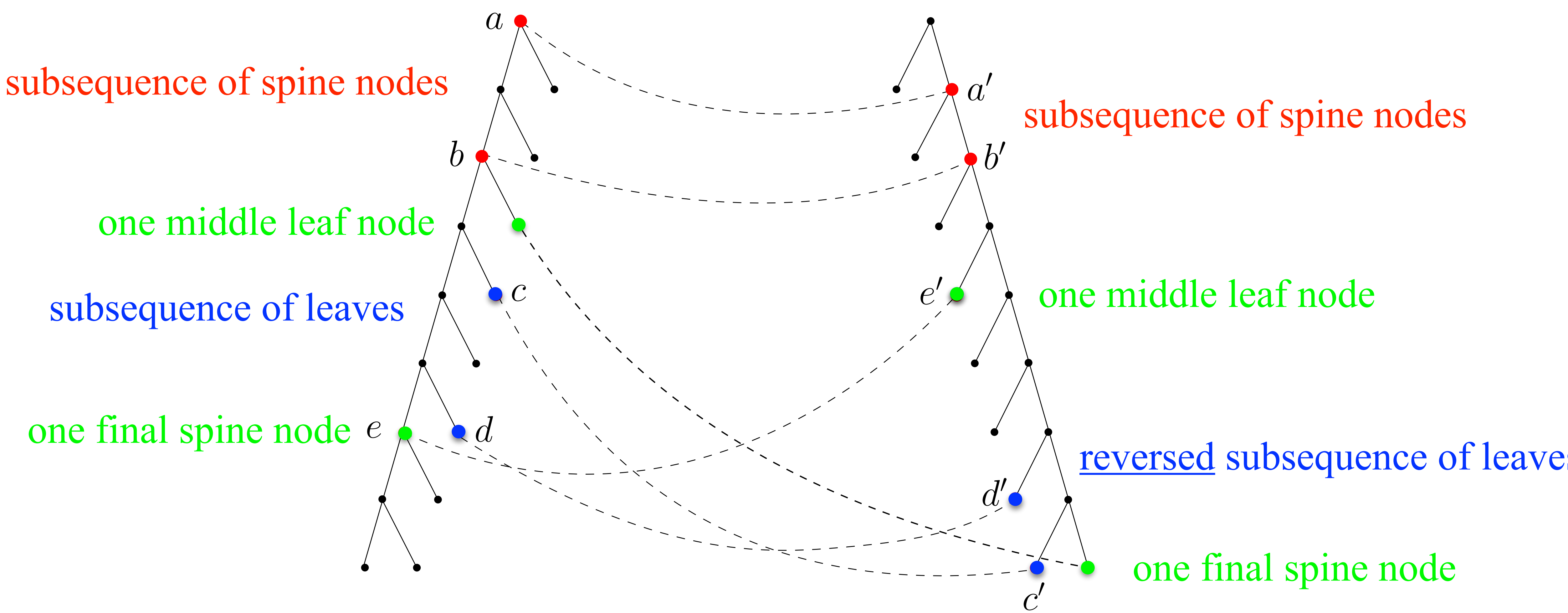
The matched pairs must have the same left-right and ancestor-descendant relations



The Hard Instance of TED

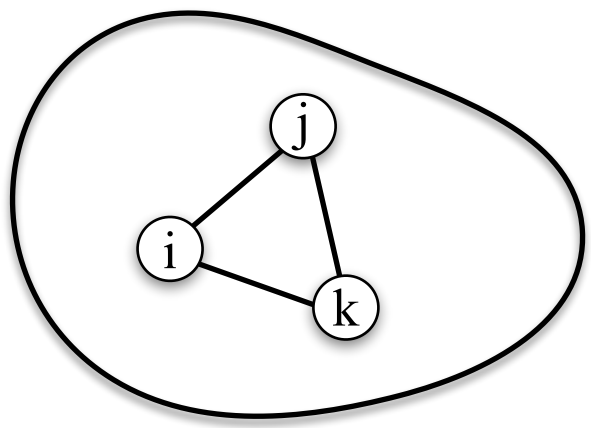
To specify a solution to TED it is enough to say which nodes are matched to which (the rest are deleted)

The matched pairs must have the same left-right and ancestor-descendant relations

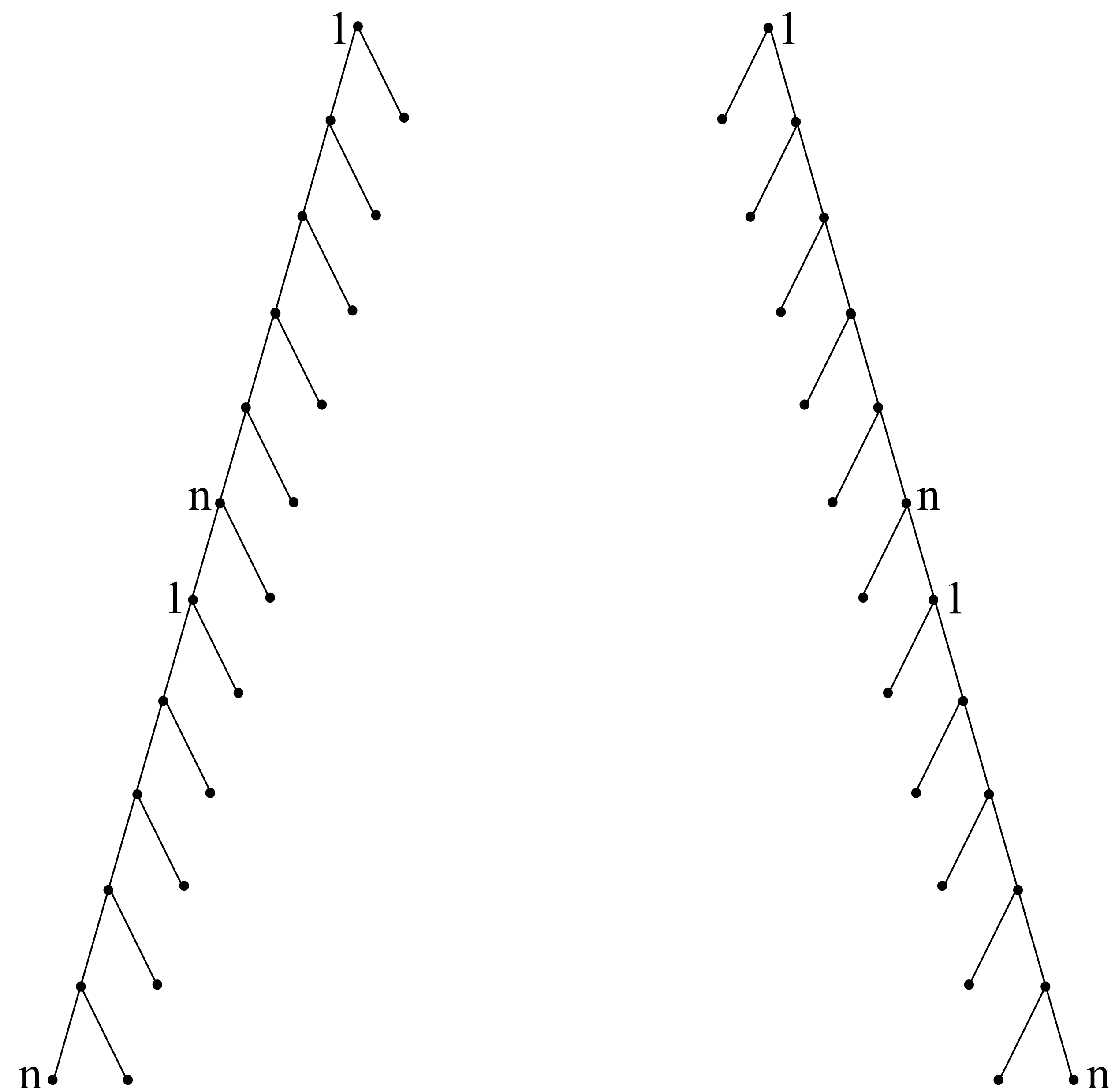


APSP ➜ TED

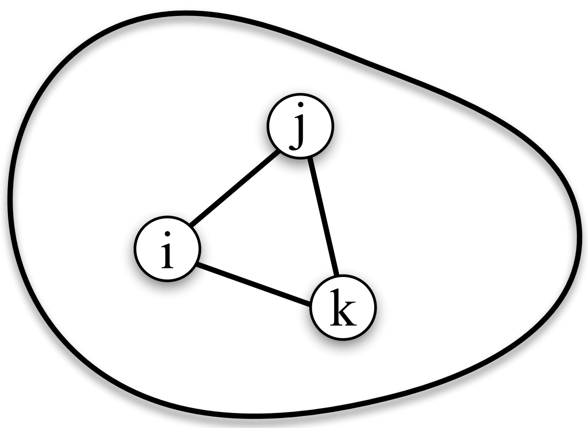
APSP ➜ TED



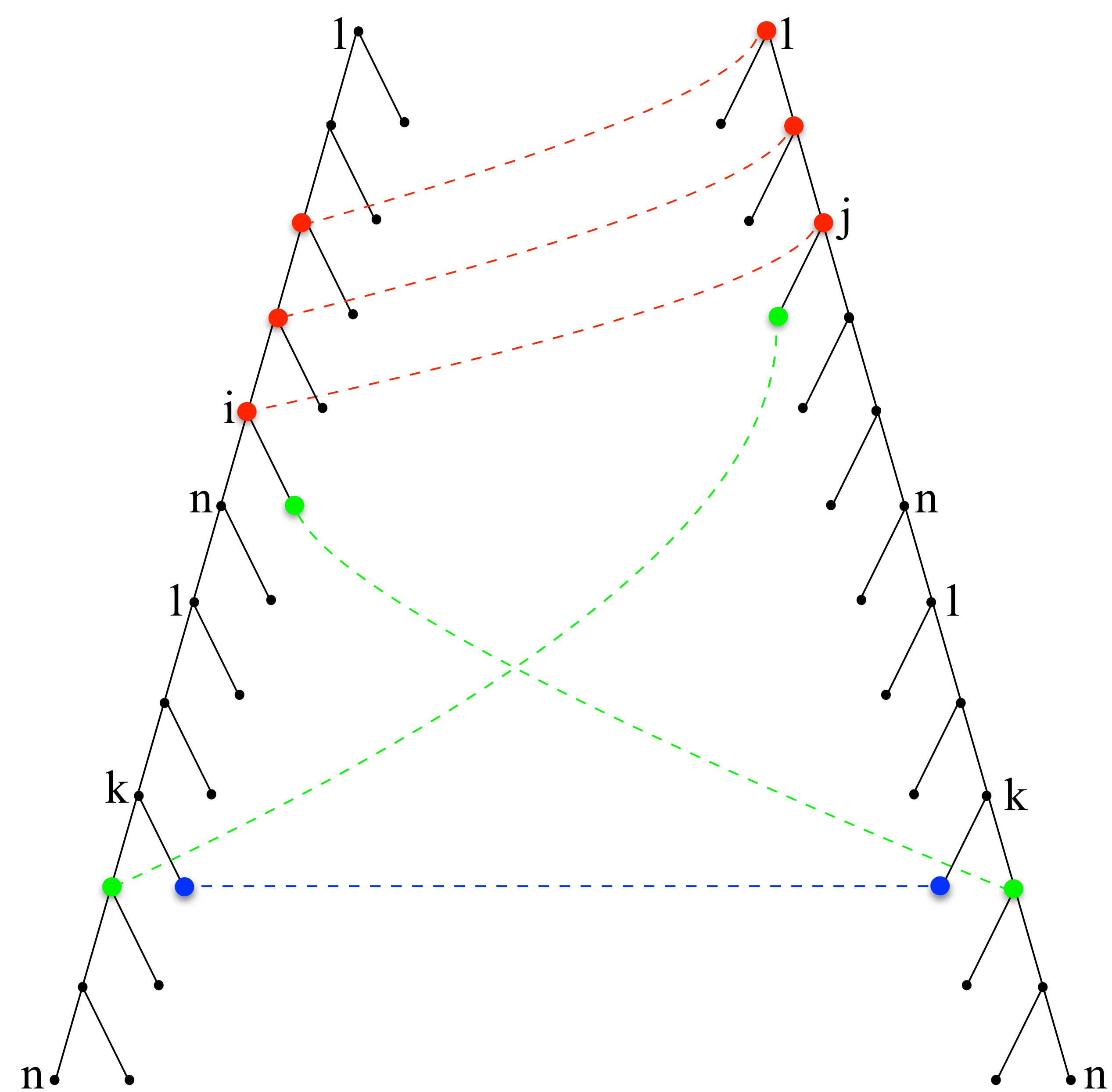
$w(i,j)+w(j,k)+w(k,i) < 0$



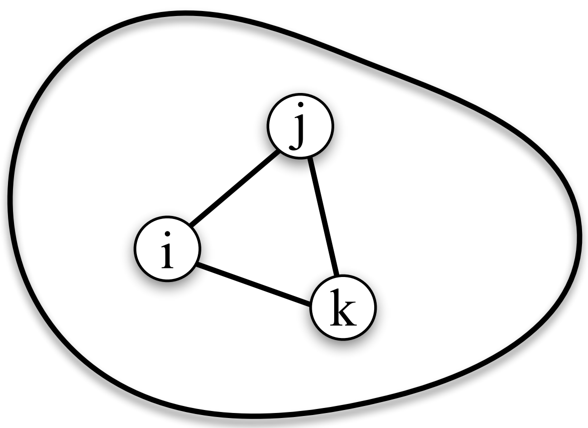
APSP ➜ TED



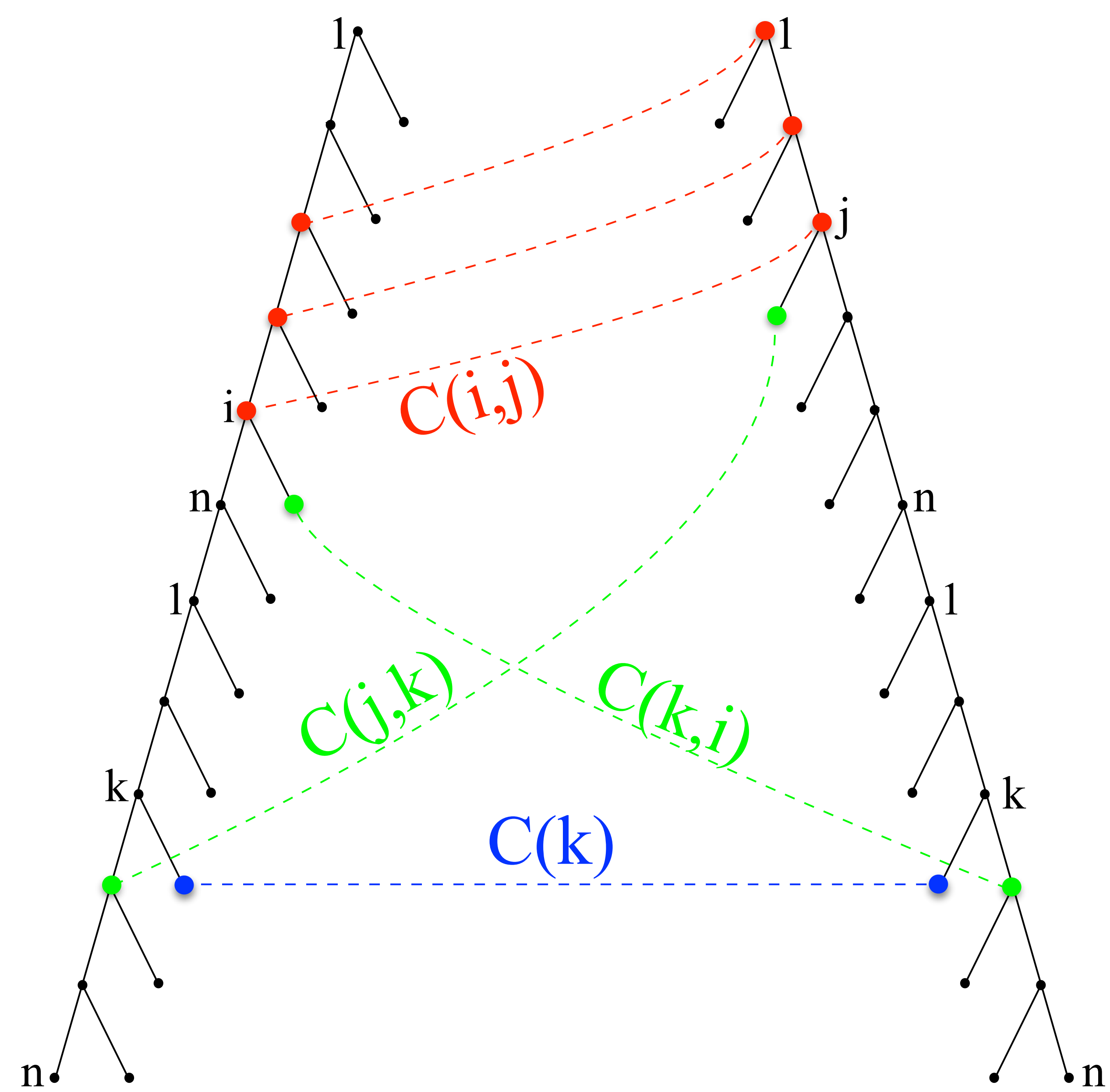
$$w(i,j)+w(j,k)+w(k,i) < 0$$



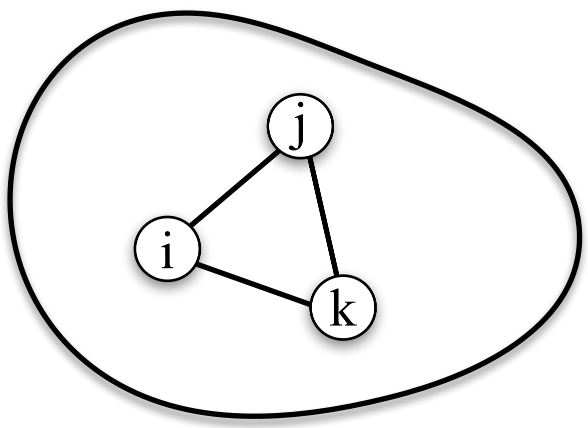
APSP ➜ TED



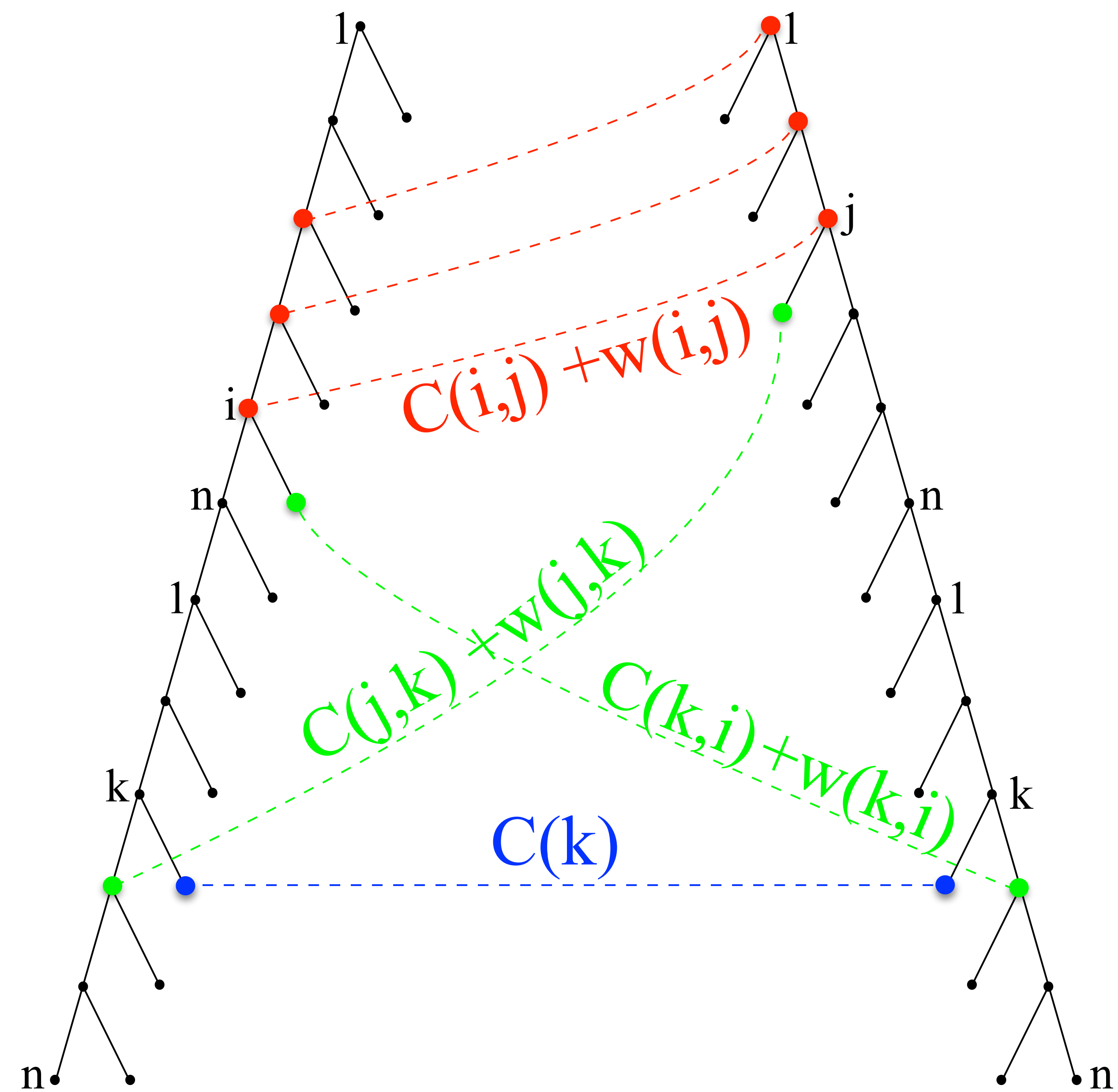
$$w(i,j)+w(j,k)+w(k,i) < 0$$



APSP ➜ TED

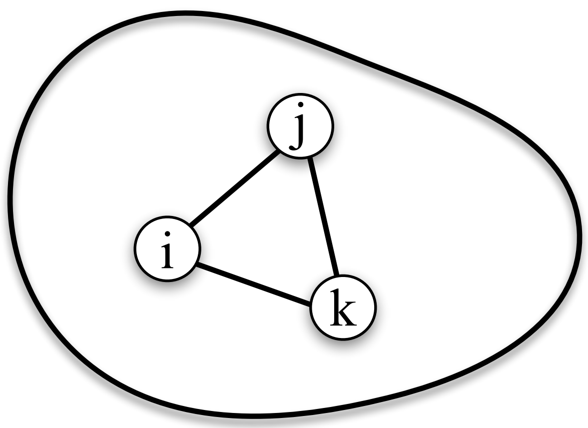


$w(i,j)+w(j,k)+w(k,i) < 0$

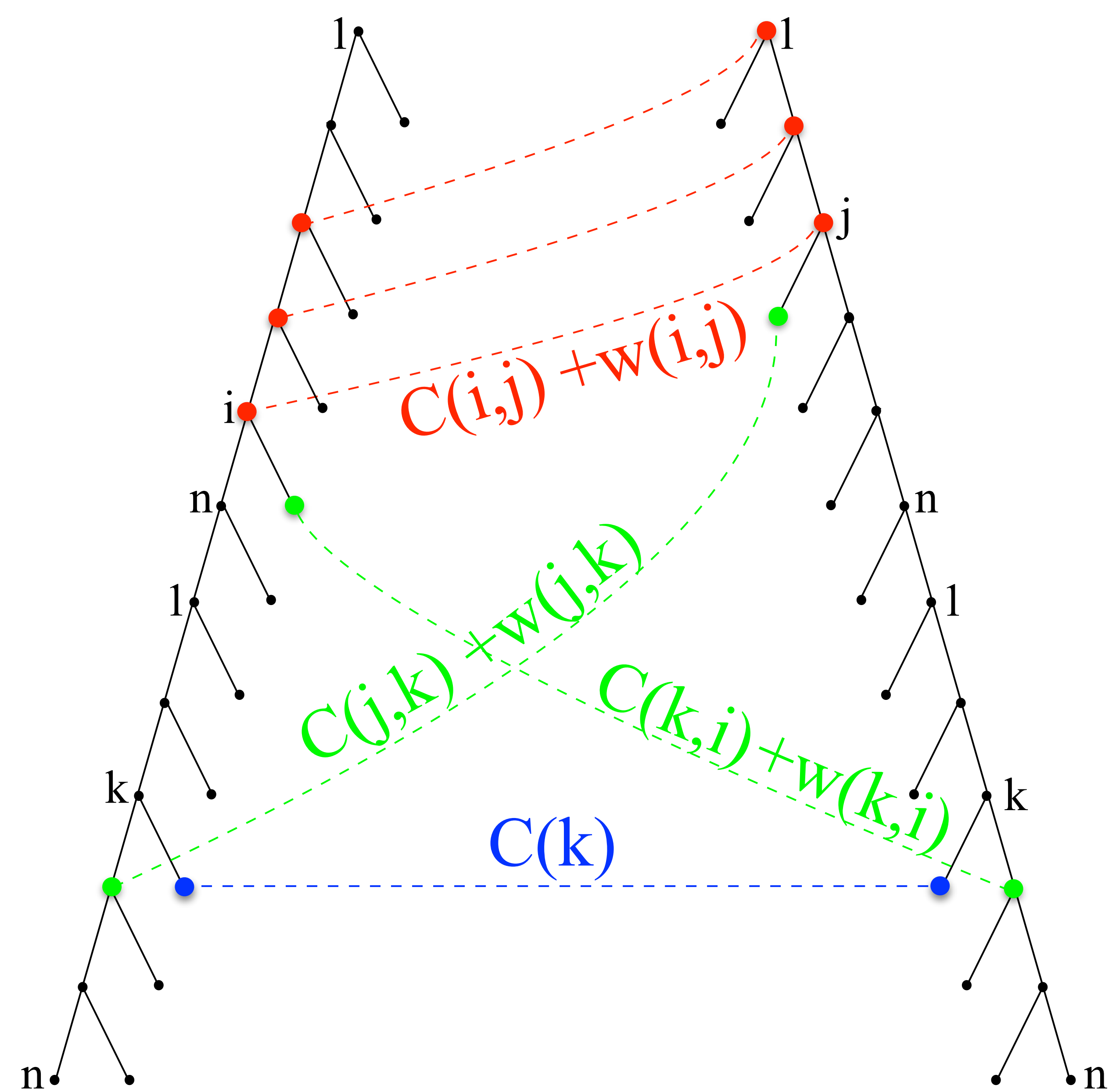


APSP ➜ TED

Large alphabet: $|\Sigma| = \Theta(n)$

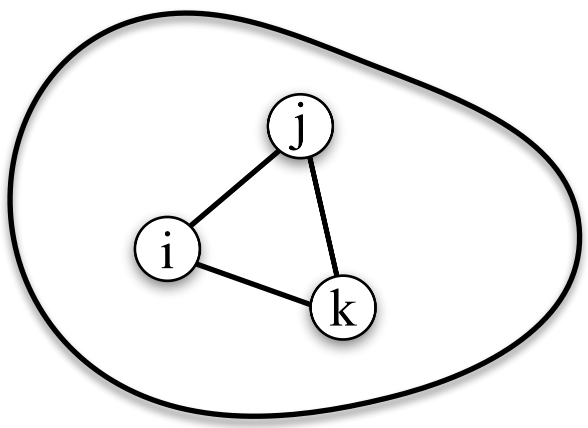


$w(i,j)+w(j,k)+w(k,i) < 0$

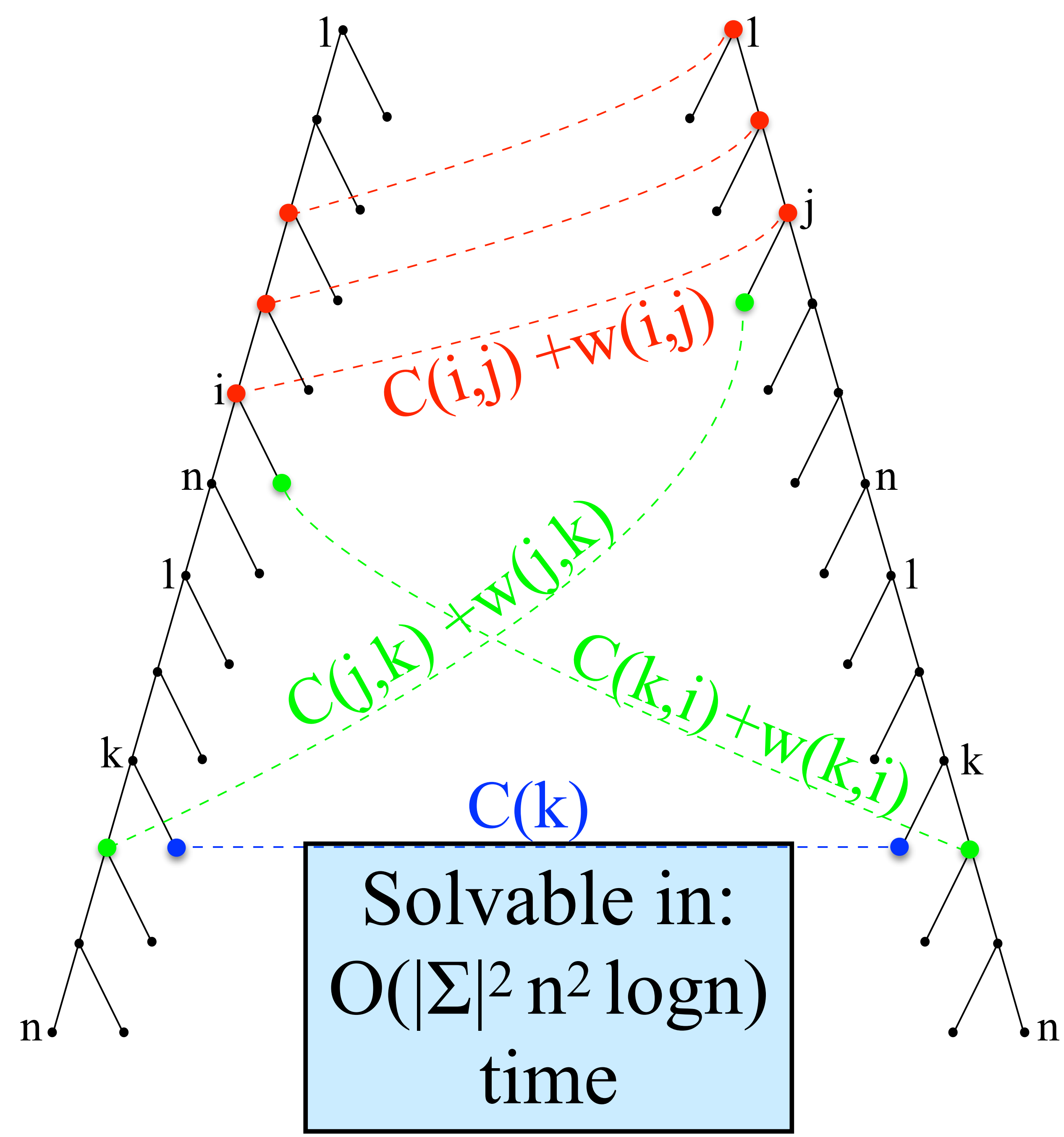


APSP ➜ TED

Large alphabet: $|\Sigma| = \Theta(n)$

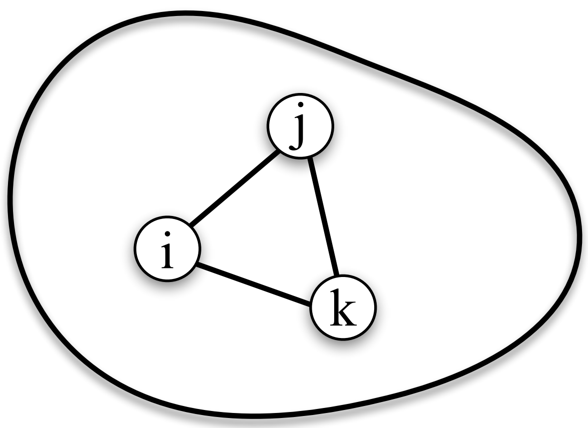


$w(i,j)+w(j,k)+w(k,i) < 0$



APSP ➜ TED

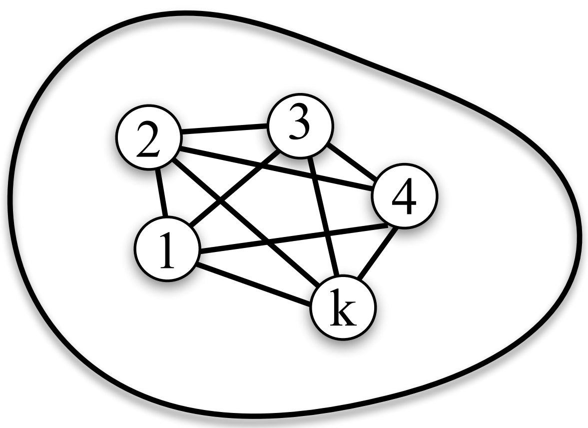
Small alphabet: $|\Sigma| = O(1)$



$w(i,j)+w(j,k)+w(k,i) < 0$

Max-weight k-Clique ➜ TED

Small alphabet: $|\Sigma| = O(1)$

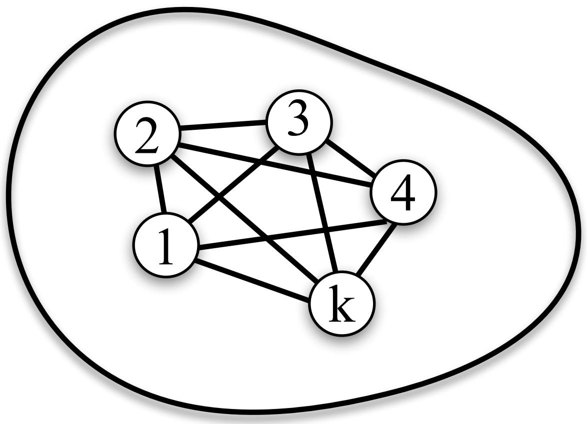


Conjecture (Max-weight k-Clique):

For any $\varepsilon > 0$ there exists $c > 0$, such that for any $k \geq 3$ finding a maximum weight k-Clique in graphs with edge weights in $\{1, \ldots, n^{ck}\}$ cannot be solved in $O(n^{k(1-\varepsilon)})$ time.

Max-weight k-Clique ➜ TED

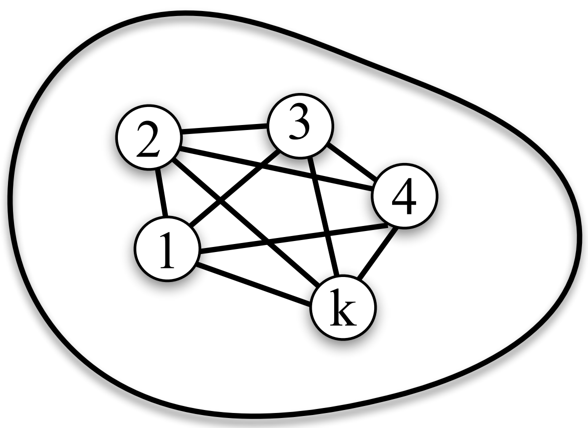
Small alphabet: $|\Sigma| = O(1)$



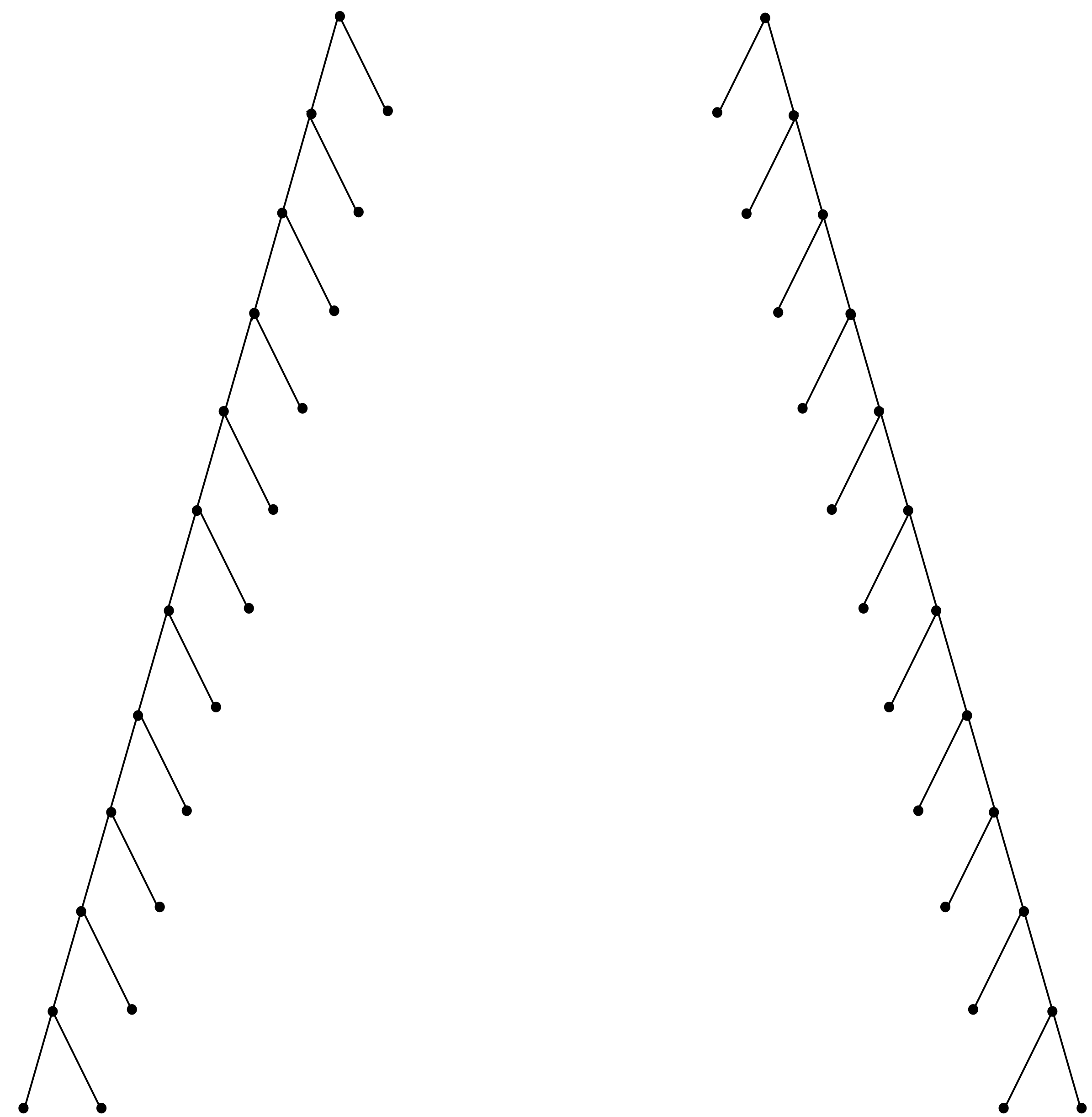
Conjecture (Max-weight k-Clique):

For any $\varepsilon > 0$ there exists $c > 0$, such that for any $k \geq 3$ finding a maximum weight k-Clique in graphs with edge weights in $\{1, \ldots, n^{ck}\}$ cannot be solved in $O(n^{k(1-\varepsilon)})$ time.

Max-weight k-Clique ➜ TED

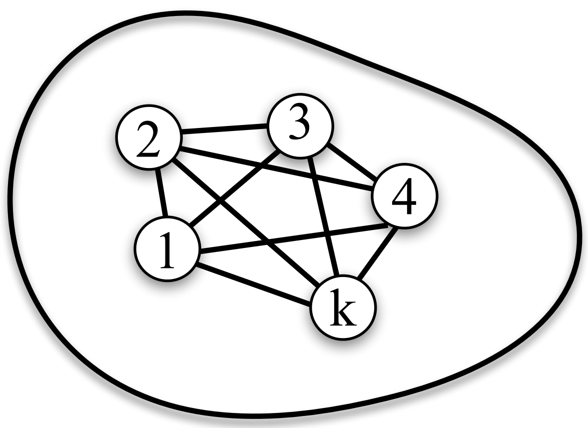


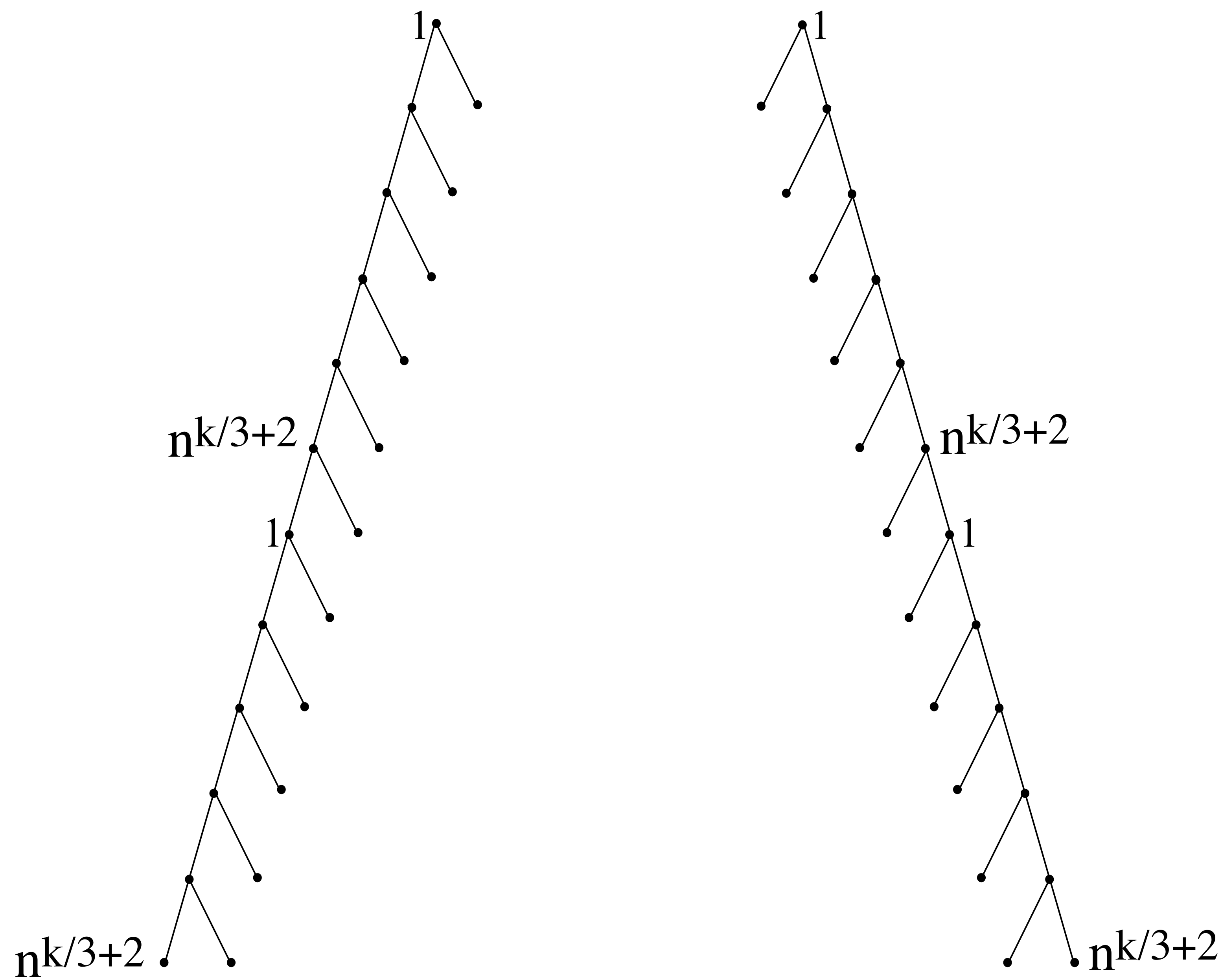
Small alphabet: $|\Sigma| = O(1)$



Max-weight k-Clique ➜ TED

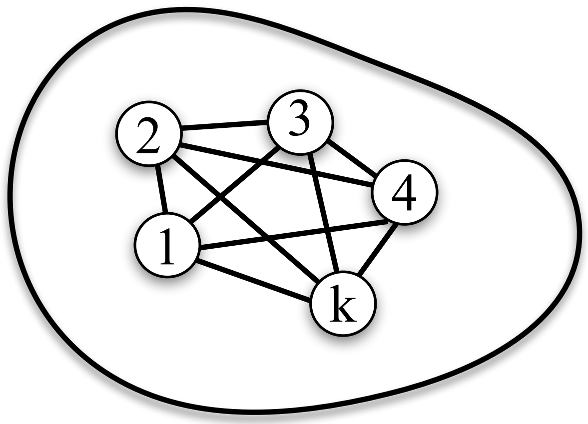
Small alphabet: $|\Sigma| = O(1)$

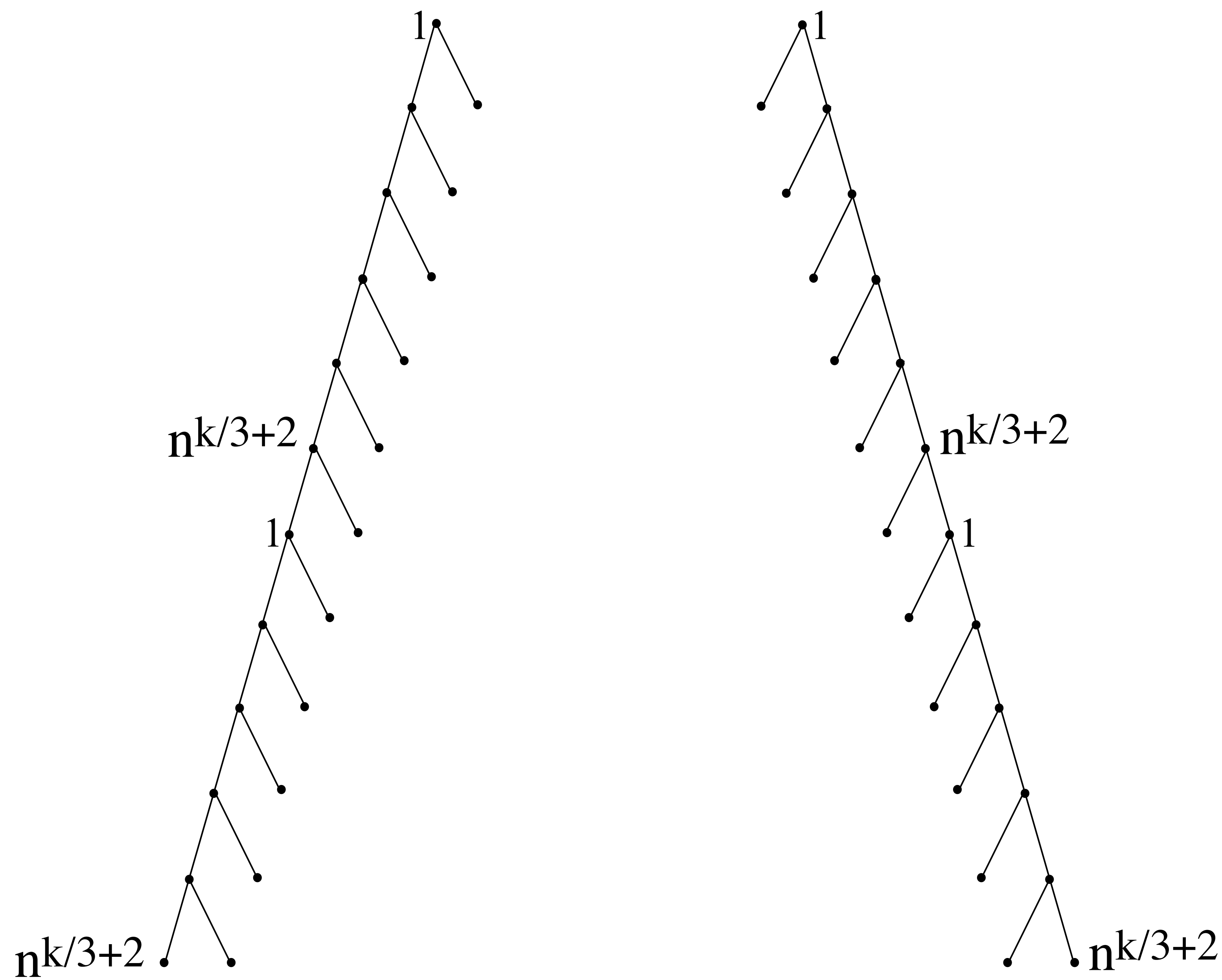




Max-weight k-Clique ➜ TED

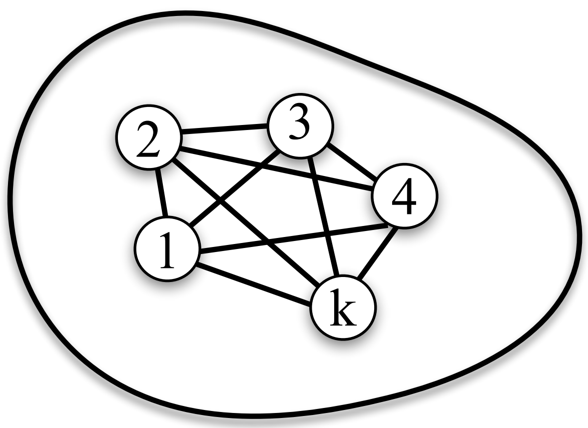
Small alphabet: $|\Sigma| = O(k)$

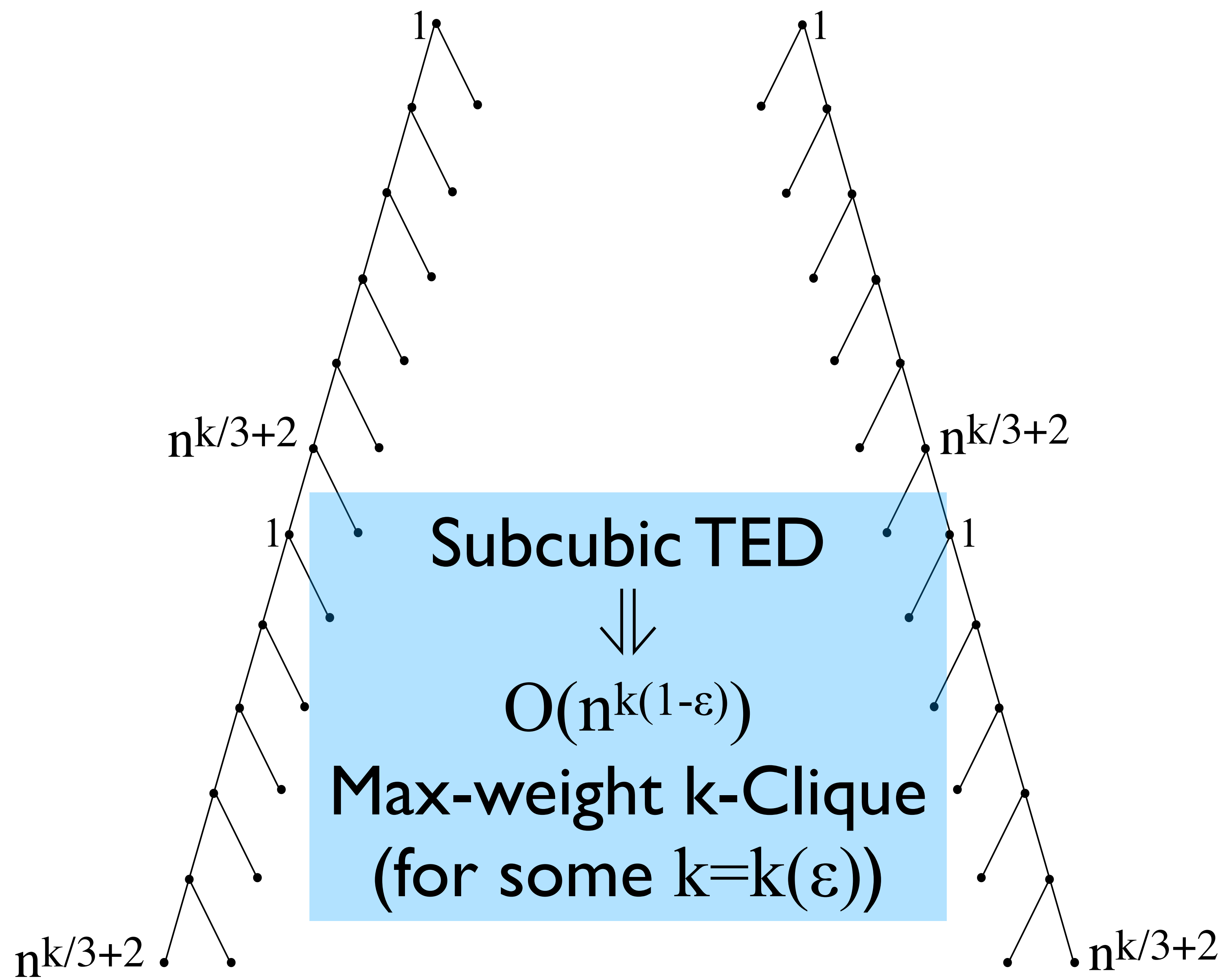




Max-weight k-Clique ➜ TED

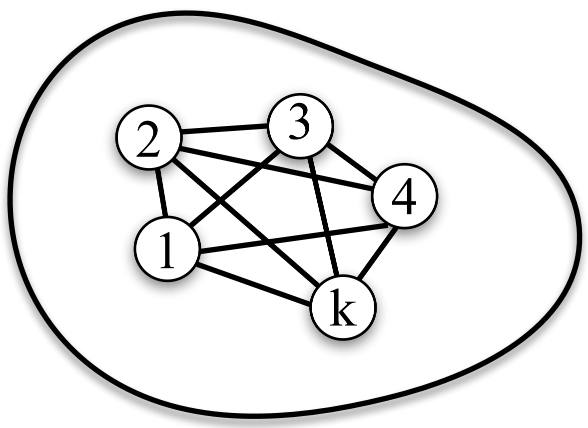
Small alphabet: $|\Sigma| = O(k)$



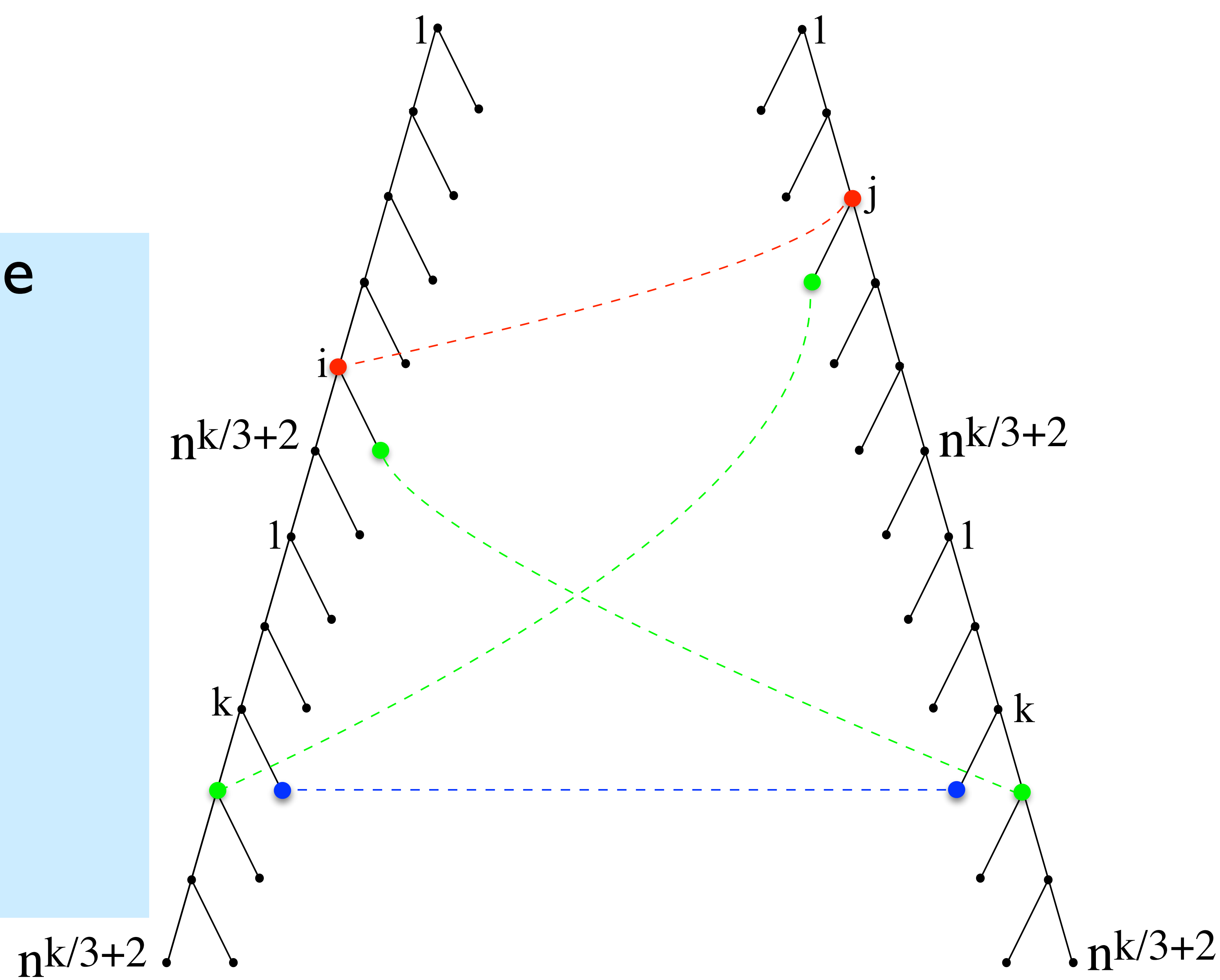


Max-weight k-Clique ➜ TED

Small alphabet: $|\Sigma| = O(k)$

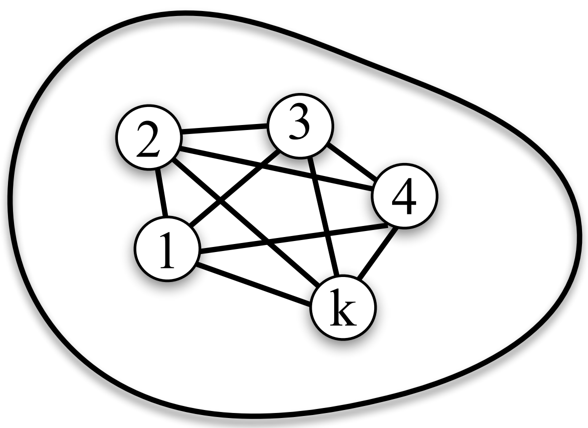


Each k/3 clique is a spine node



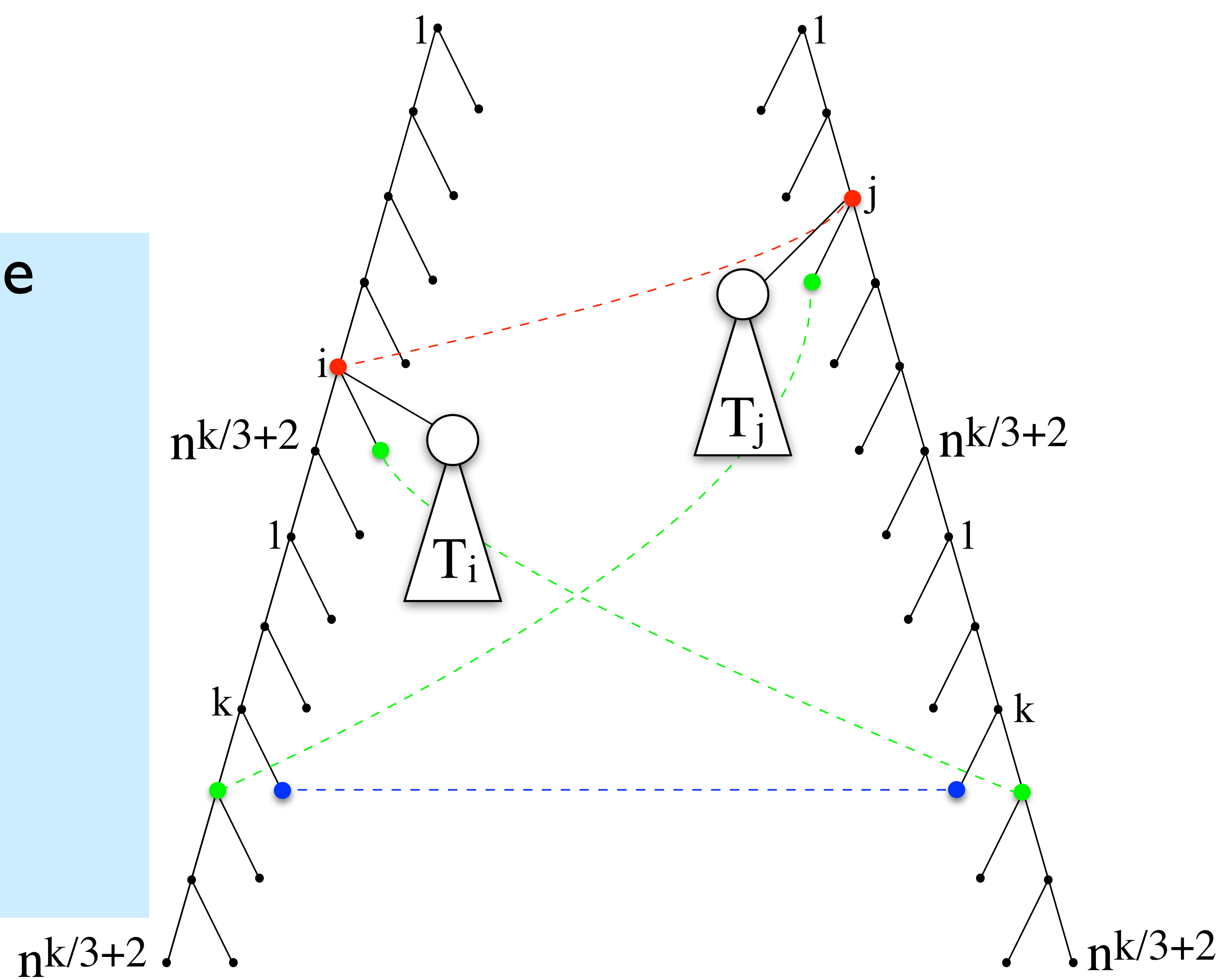
Max-weight k-Clique ➜ TED

Small alphabet: $|\Sigma| = O(k)$



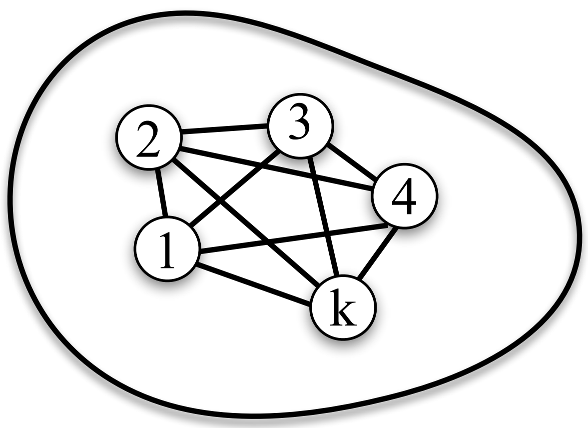
Each k/3 clique is a spine node

Simulate matching costs with small (n^2 size) gadgets T_i.



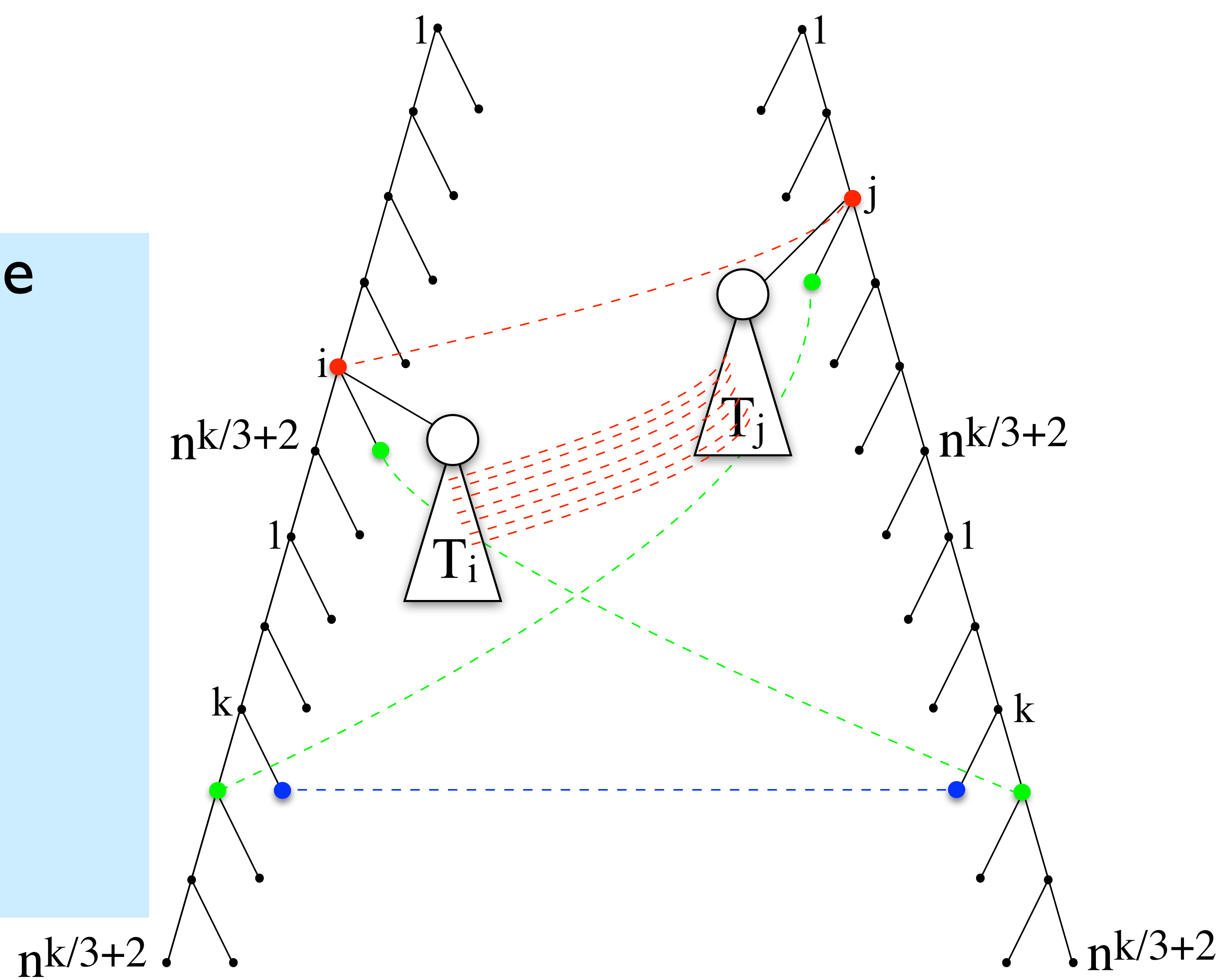
Max-weight k-Clique ➜ TED

Small alphabet: $|\Sigma| = O(k)$



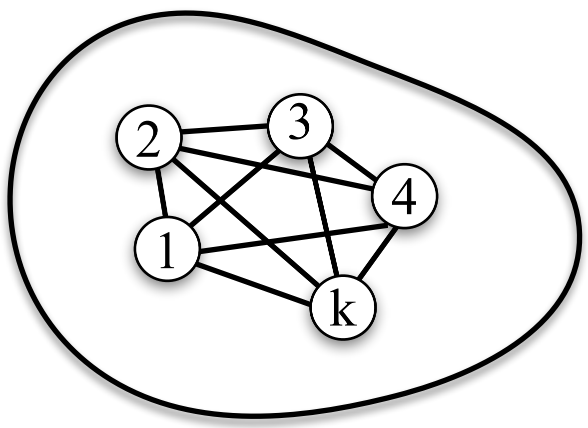
Each k/3 clique is a spine node

Simulate matching costs with small (n^2 size) gadgets T_i.



Max-weight k-Clique ➜ TED

Small alphabet: $|\Sigma| = O(k)$

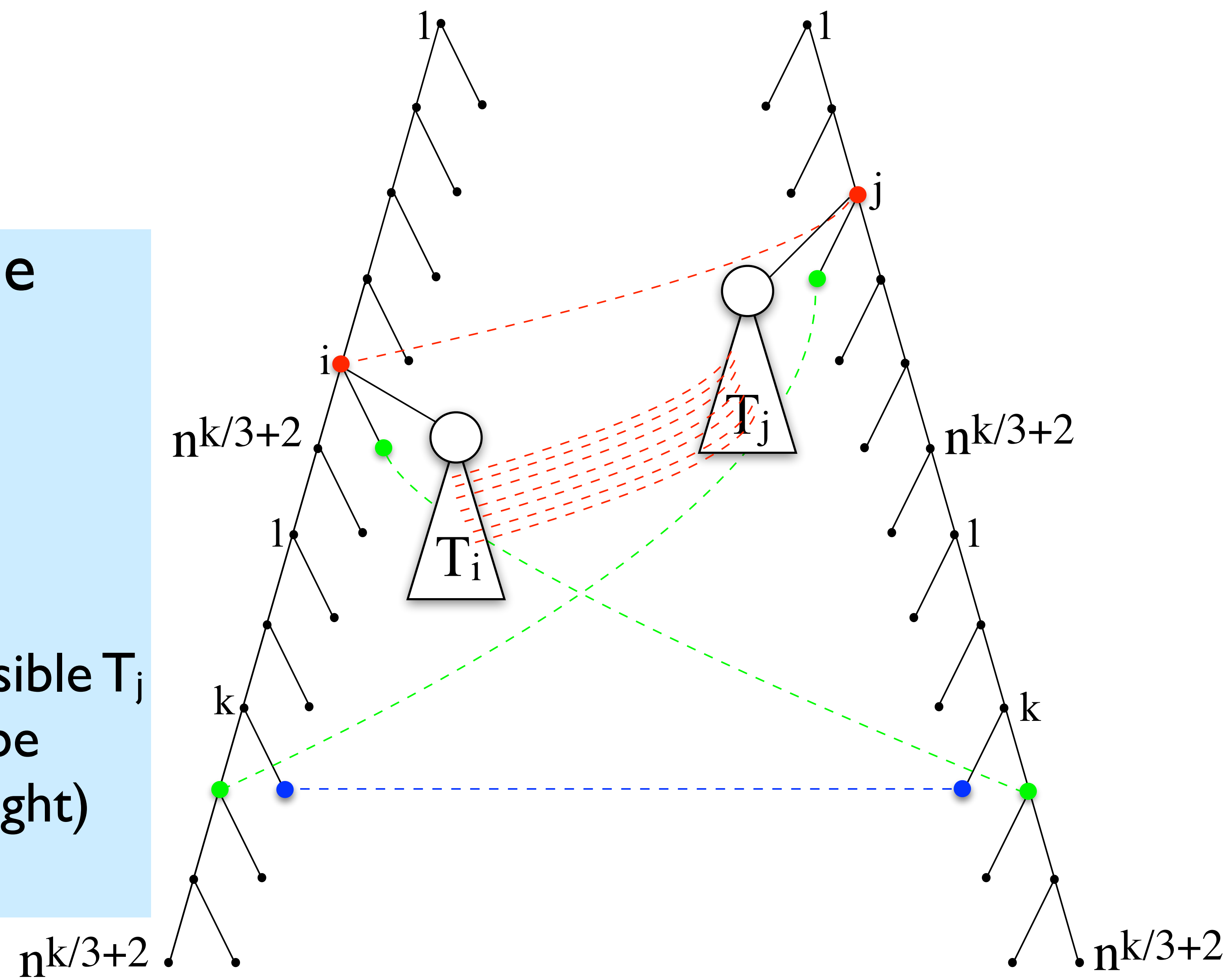


Each k/3 clique is a spine node

Simulate matching costs with small (n^2 size) gadgets T_i.

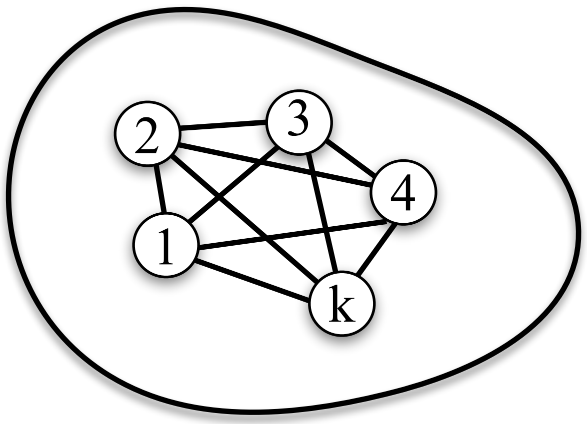
Challenging:

- T_i needs to "prepare" for any possible T_j
- we need to control which T_i can be matched to which (in APSP by height)
- constant O(k) size alphabet



Max-weight k-Clique ➜ TED

Small alphabet: $|\Sigma| = O(k)$

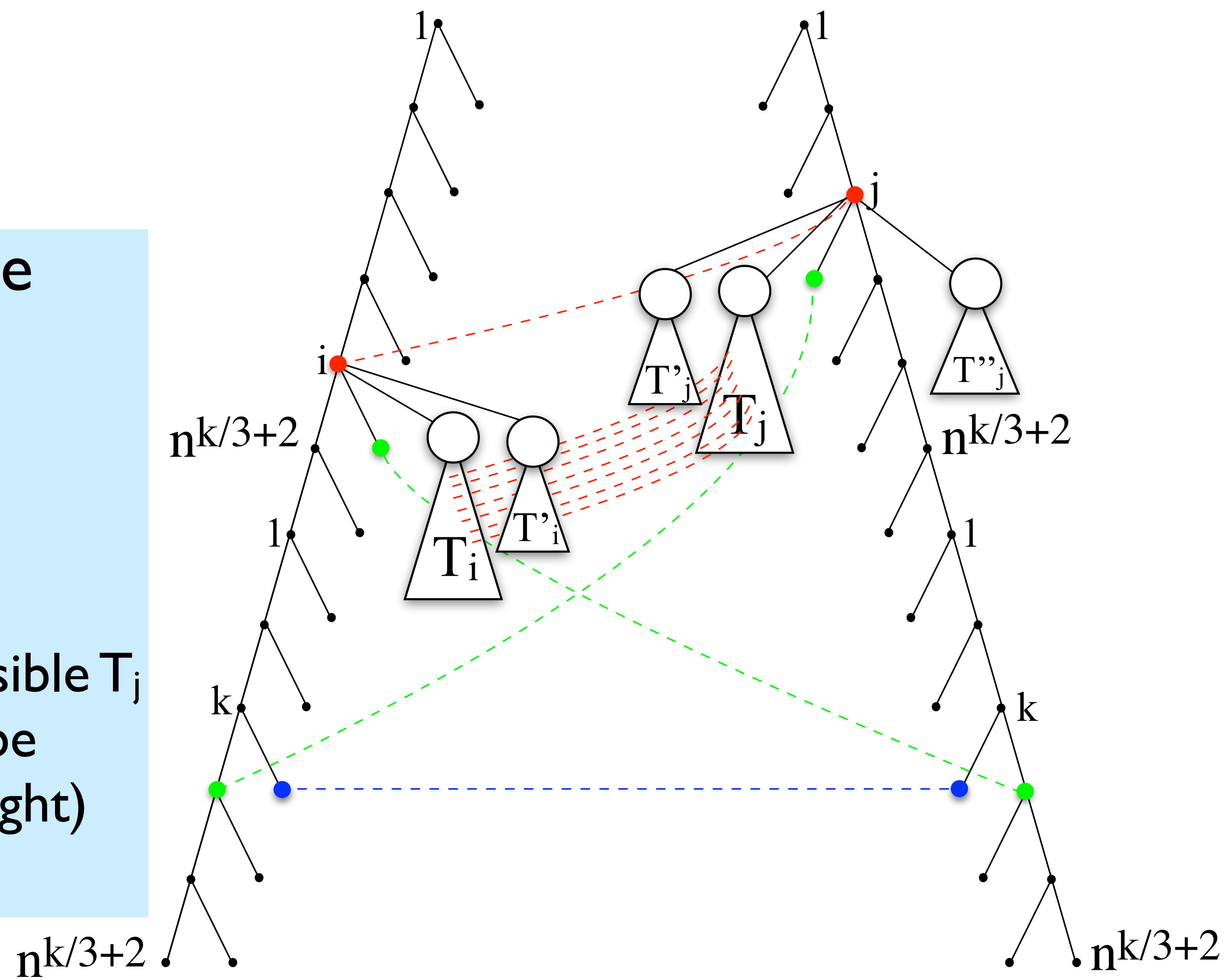


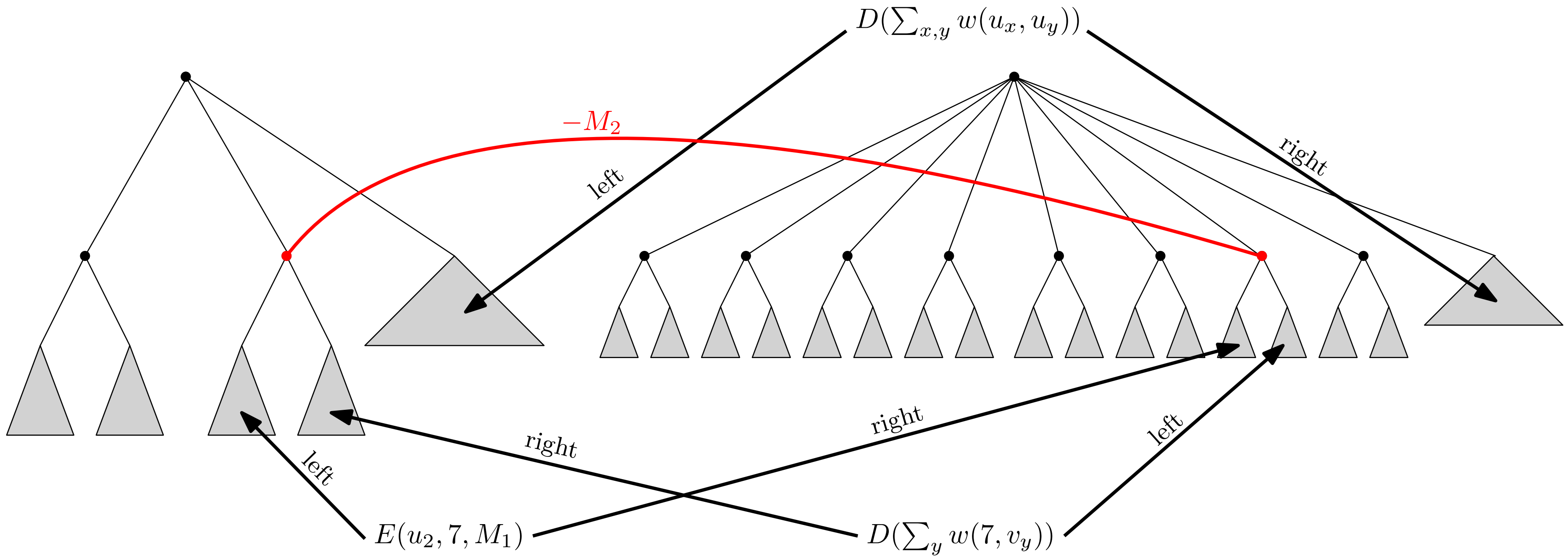
Each k/3 clique is a spine node

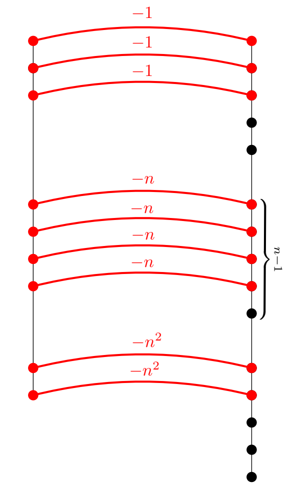
Simulate matching costs with small (n^2 size) gadgets T_i.

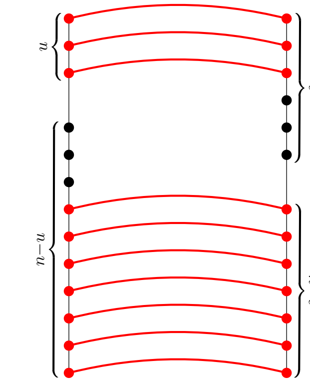
Challenging:

- T_i needs to "prepare" for any possible T_j
- we need to control which T_i can be matched to which (in APSP by height)
- constant O(k) size alphabet









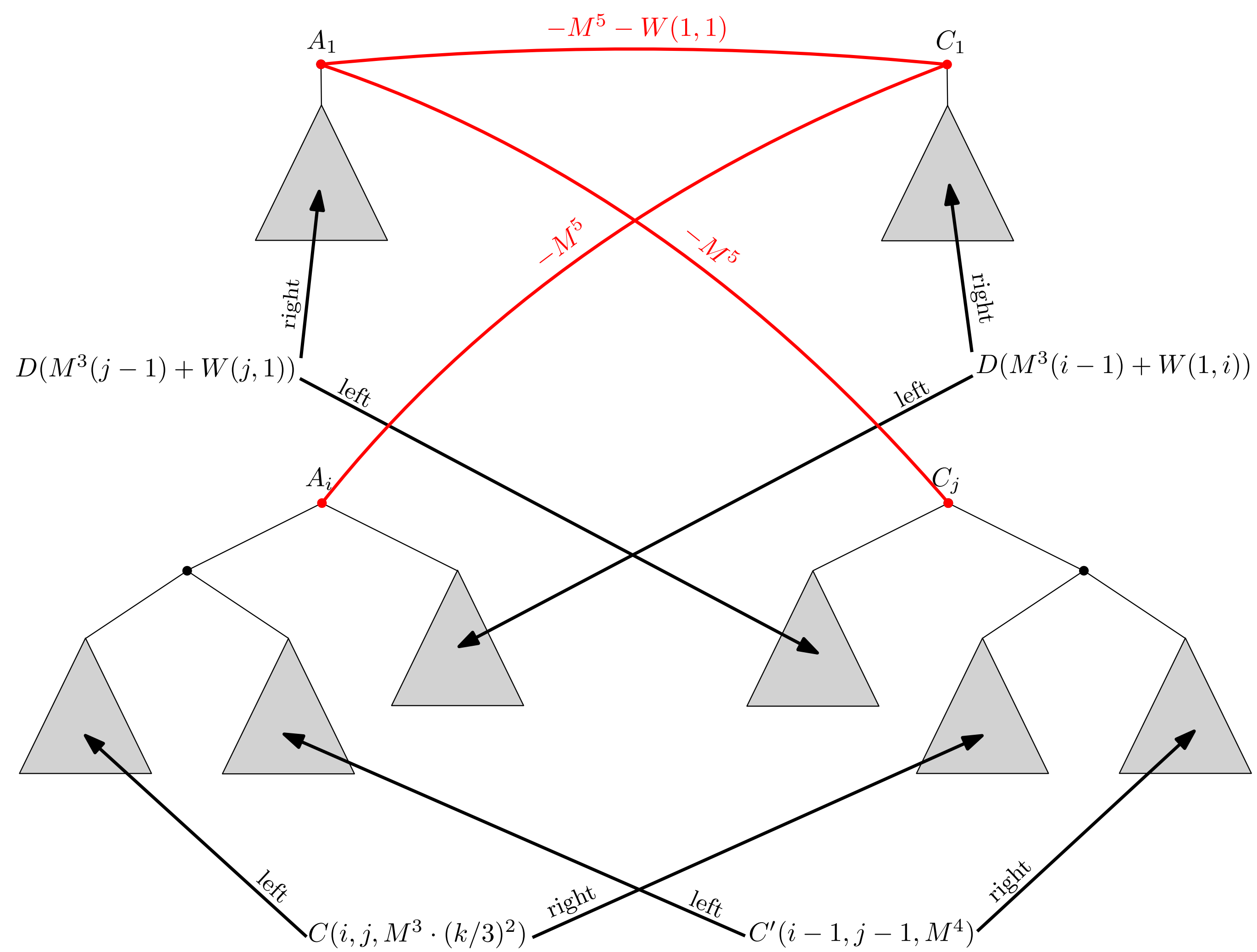
1. $c_{\text{match}}(A'_i, D_{z'}) = -M^6 - M^3(N-i) - W(i, z')$ for every $i = 1, 2, \ldots, N$ and $z' = 1, 2, \ldots, N$,
2. $c_{\text{match}}(B_z, C'_j) = -M^6 - M^3(N-j) - W(z, j)$ for every $z = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, N$.
3. $c_{\text{match}}(A_i, C_j) = -M^2 - W(j, i) + W(j-1, i-1)$ for every $i = 2, 3, \ldots, N$ and $j = 2, 3, \ldots, N$.
4. $c_{\text{match}}(A_i, C_1) = -M^5 - M^3(i-1) - W(1, i)$ for every $i = 1, 2, \ldots, N$,
5. $c_{\text{match}}(A_1, C_j) = -M^5 - M^3(j-1) - W(j, 1)$ for every $j = 1, 2, \ldots, N$.

Lemma 5. *For sufficiently large M, the total cost of an optimal matching is*

$$-M^8 \cdot 2 - M^7 \cdot 2(N-1) - M^6 \cdot 2 - M^5 - M^3 \cdot 2N + M^2 - \max_{i,j,z}\{W(i,z) + W(z,j) + W(j,i)\}.$$

Proof. Consider i, j, z maximizing $W(i,z) + W(z,j) + W(j,i)$. We may assume that $i \geq j$. Then, it is possible to choose the following matching:

1. b_k to c'_j with cost $-M^8$,
2. some nodes from the copy of I being the left child of c'_j to some spine nodes below b_z with total cost $-M^7(N-z)$,
3. a'_i to d_k with cost $-M^8$,
4. some nodes from the copy of I being d of a'_i to some spine nodes below d_z with total cost $-M^7(N-z)$,
5. b'_1 to d'_{z-1}, b'_2 to d'_{z-2}, …, b'_{z-1} to d'_1 with cost $-M^7 \cdot 2$ each,
6. a_i to c_j, a_{i-1} to c_{j-1}, …, a_{i-j+1} to c_1 with cost $-M^3 \cdot 2 + M^2$ each,
7. A'_i to D_z with cost $-M^6 - M^3(N-i) - W(i,z)$,
8. B_z to C'_j with cost $-M^6 - M^3(N-j) - W(z,j)$,
9. A_i to C_j, A_{i-1} to C_{j-1}, …, A_{i-j+2} to C_2 with costs $-M^2 - W(j,i) + W(j-1,i-1)$, $-M^2 - W(j-1,i-1) + W(j-2,i-2)$, …, $-M^2 - W(2, i-j+2) + W(1, i-j+1)$.
10. A_{i-j+1} to C_1 with cost $-M^5 - M^3(i-j) - W(1, i-j+1)$.



Open Problems

Open Problems

- TED to APSP reduction?

Open Problems

- TED to APSP reduction?
- *Largest common subforest*: unlabeled trees ($|\Sigma| = 1$)

Open Problems

- TED to APSP reduction?
- *Largest common subforest*: unlabeled trees ($|\Sigma| = 1$)
- *Levenshtein distance*: every elementary edit operation costs 1

Open Problems

- TED to APSP reduction?
- *Largest common subforest*: unlabeled trees ($|\Sigma| = 1$)
- *Levenshtein distance*: every elementary edit operation costs 1
- log shaves?

Thank You!