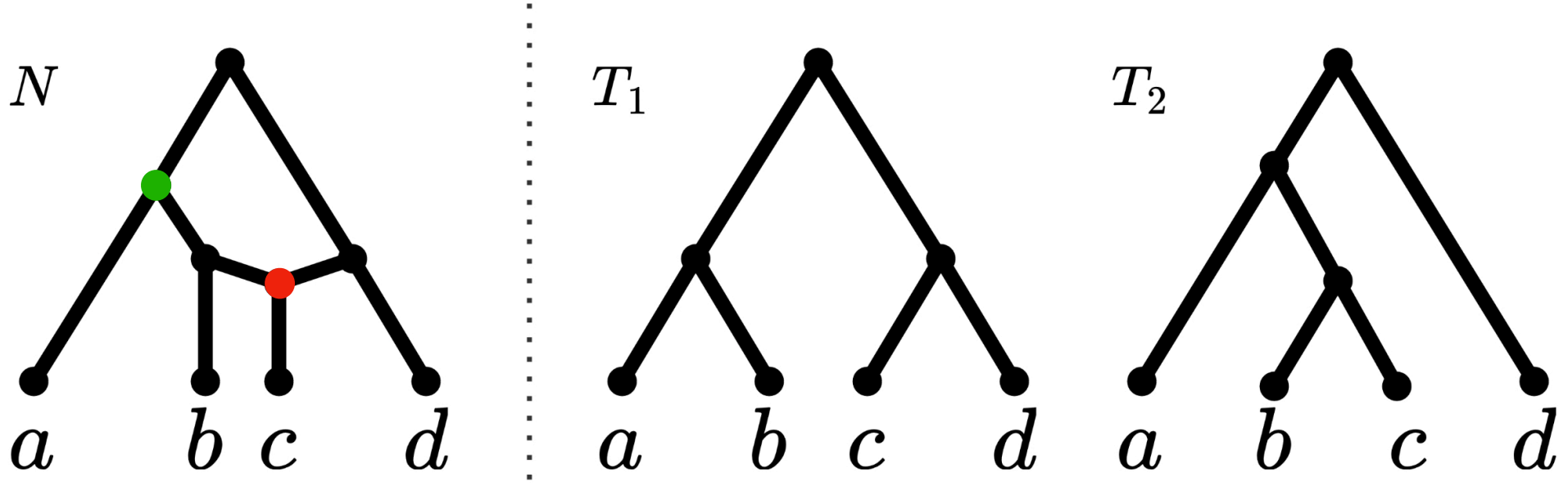
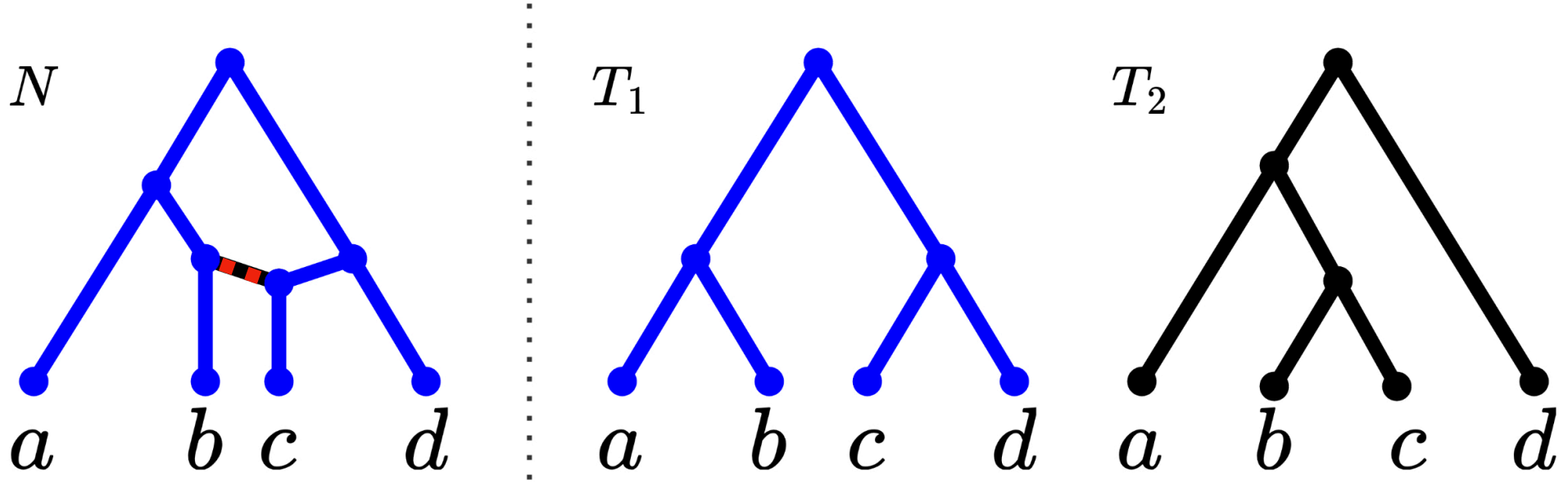
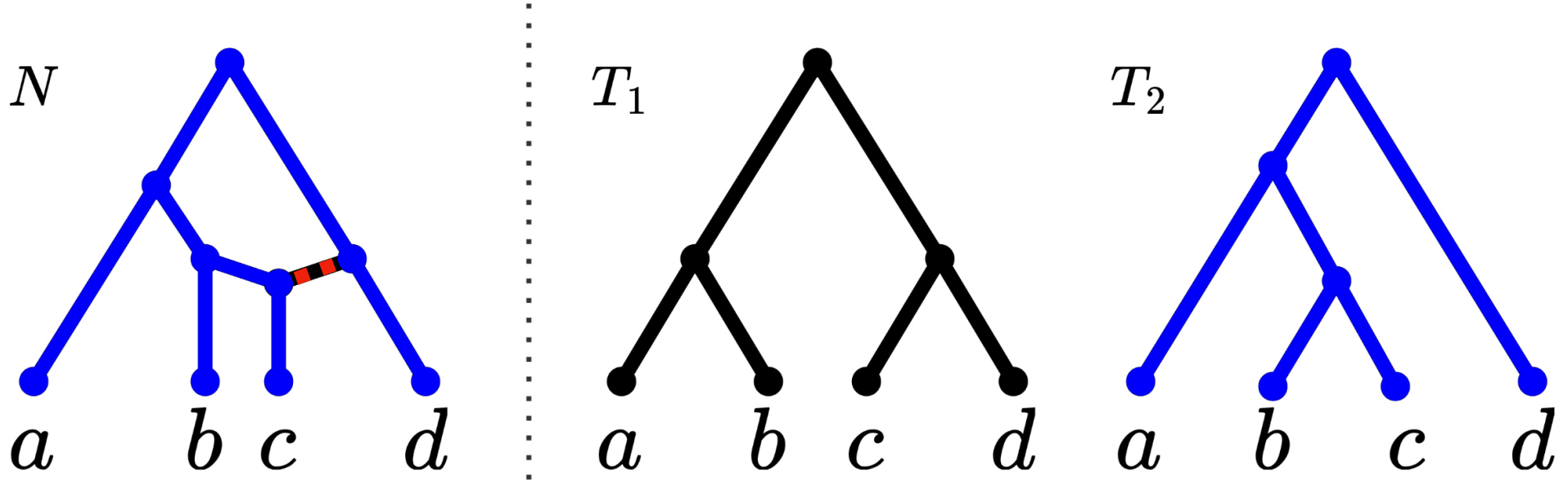
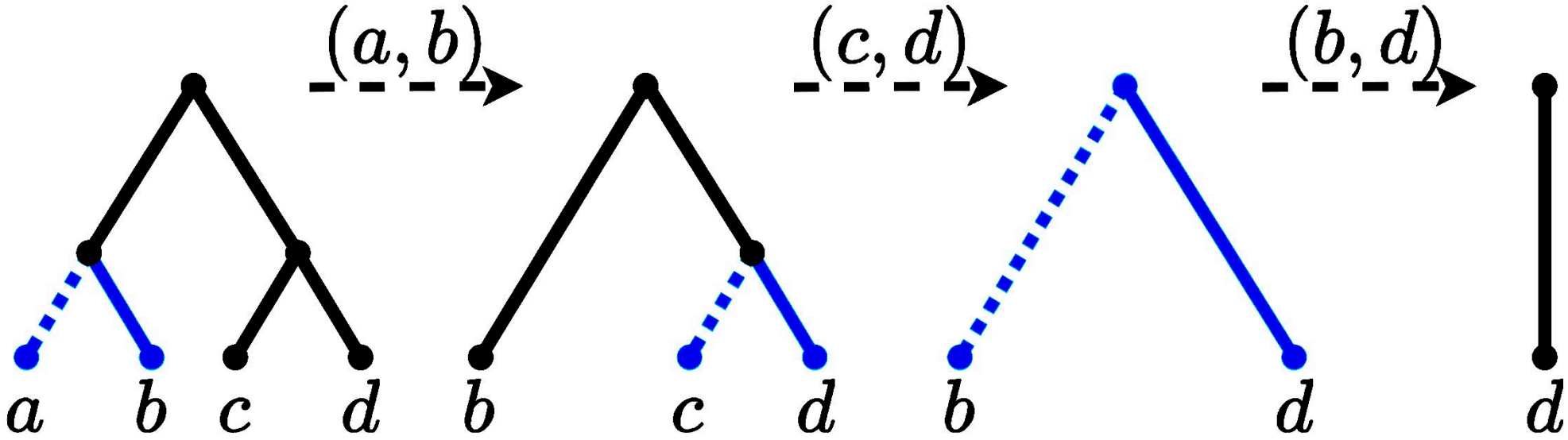
Constructing Phylogenetic Networks via Cherry Picking and Machine Learning

Giulia Bernardini, Leo van Iersel,
Esther Julien, Leen Stougie


N
a
b
c
d
T1
a
b
c
d
T2
a
b
c
d

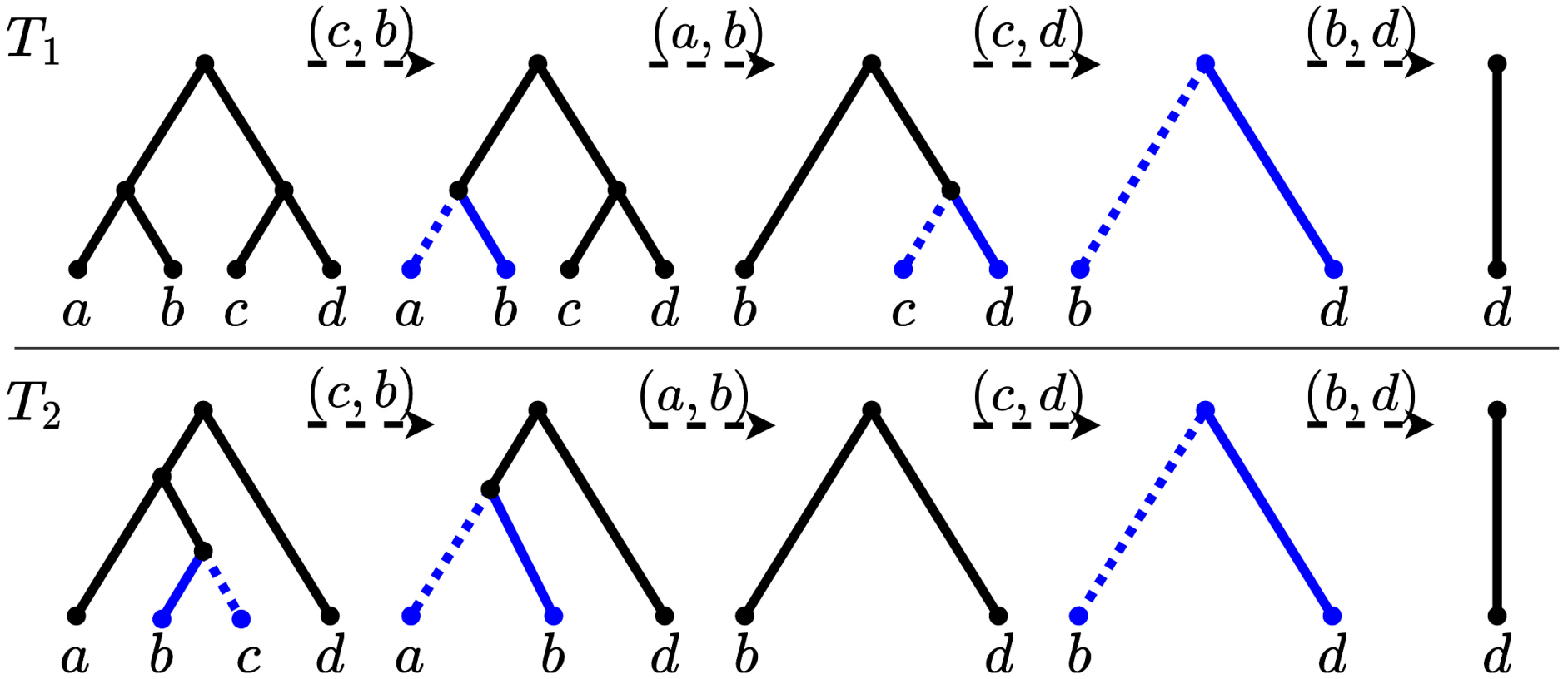

N
a
b
c
d
T1
a
b
c
d
T2
a
b
c
d

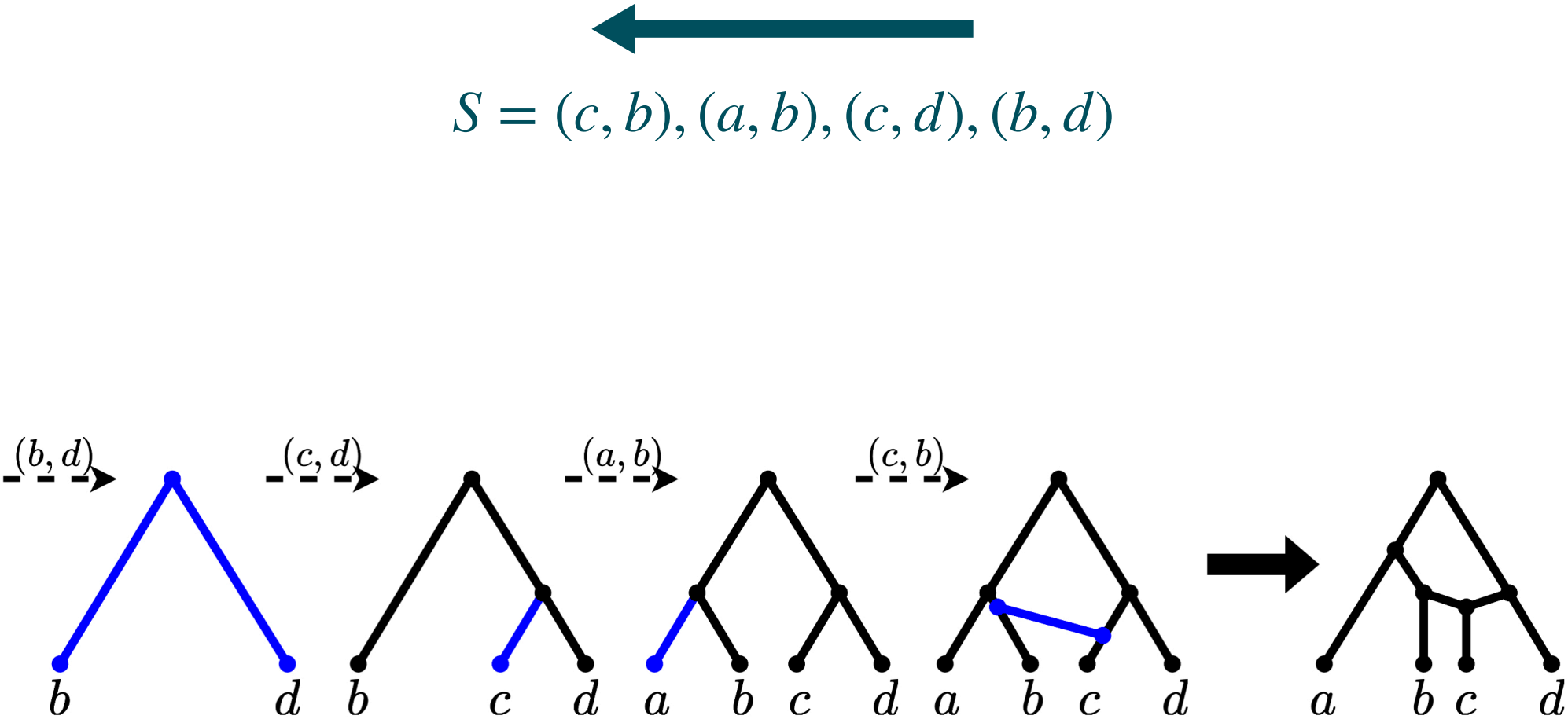

N
a
b
c
d
T1
a
b
c
d
T2
a
b
c
d



Linz, S., Semple, C. *Attaching leaves and picking cherries to characterise the hybridisation number for a set of phylogenies.* Advances in Applied Mathematics (2019)

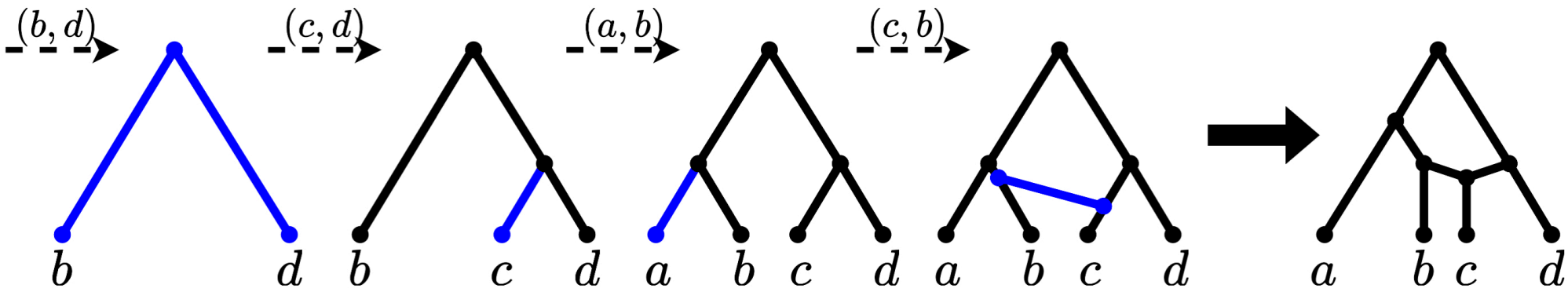
$$S = (c,b),(a,b),(c,d),(b,d)$$


T_1
(c,b)
(a,b)
(c,d)
(b,d)
a b c d
a b c d
b c d
b d
d
T_2
(c,b)
(a,b)
(c,d)
(b,d)
a b c d
a b d
b d
b d
d
T_3
T_4
⋮


$S = (c,b),(a,b),(c,d),(b,d)$
(b,d)
(c,d)
(a,b)
(c,b)
b d
b c d
a b c d
a b c d
a b c d

$$S = \underbrace{(c,b),(a,b),(c,d),(b,d)}$$

Minimize the length



$$R = |S| - n + 1$$

R: number of reticulations

n: number of leaves

Cherry-picking Heuristic Framework

$S = \emptyset$

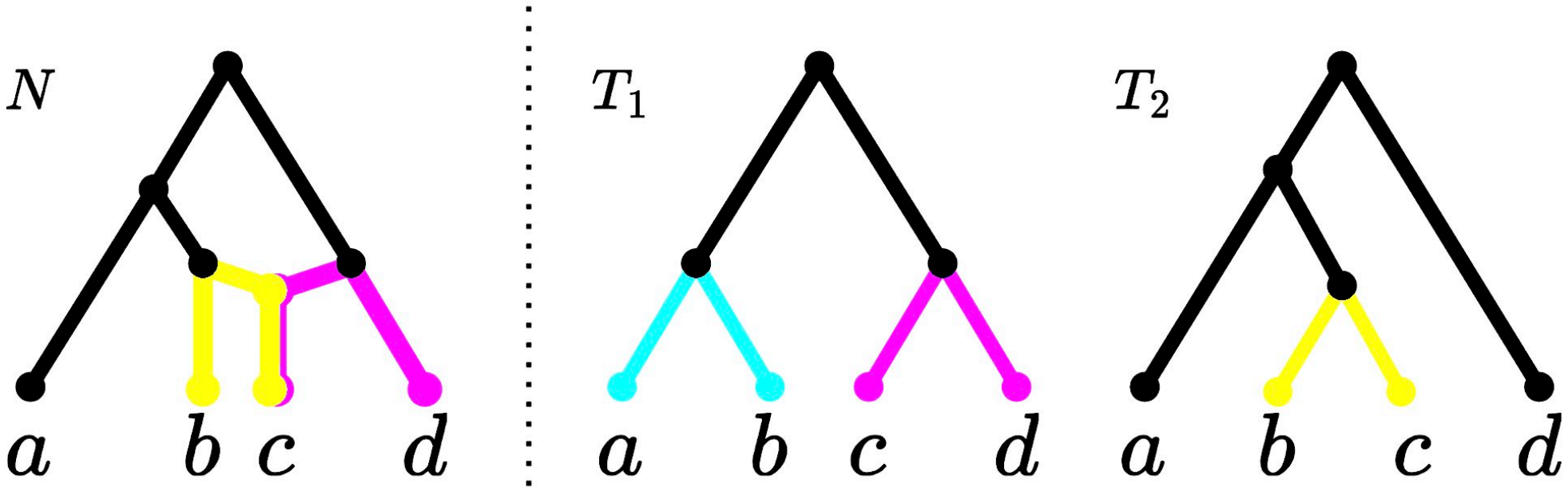
WHILE any tree not fully reduced:

CHOOSE a cherry (x, y)

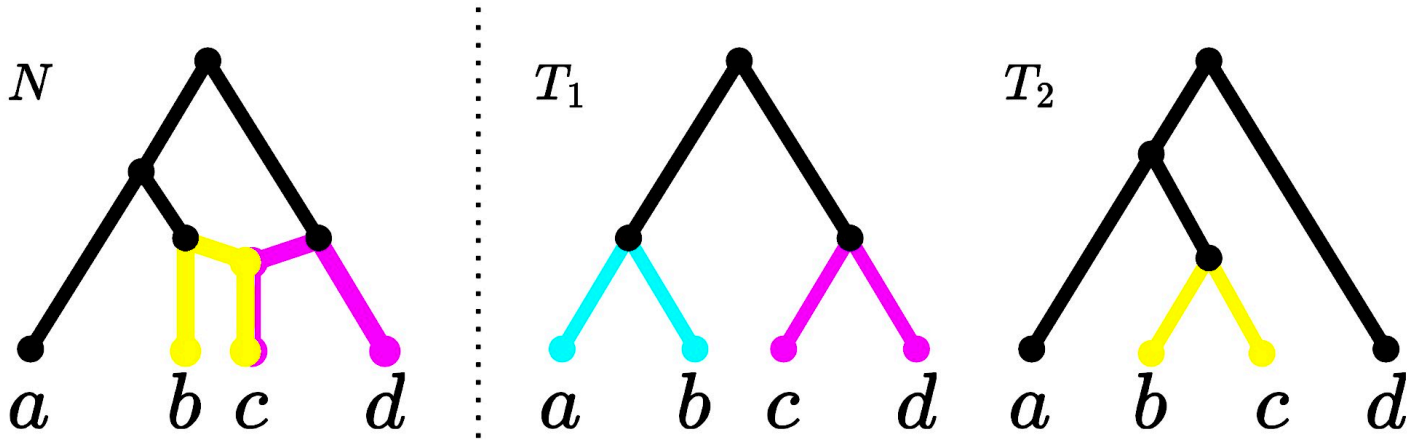
$S = S \circ (x, y)$

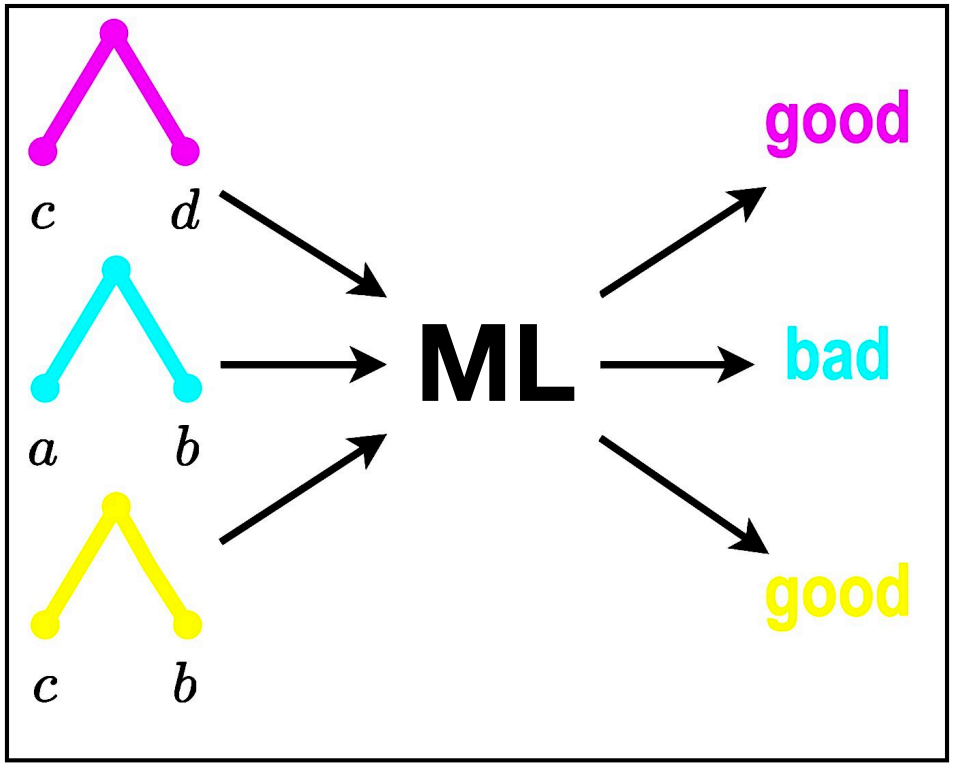
pick (x, y) in all trees

Generate network from S



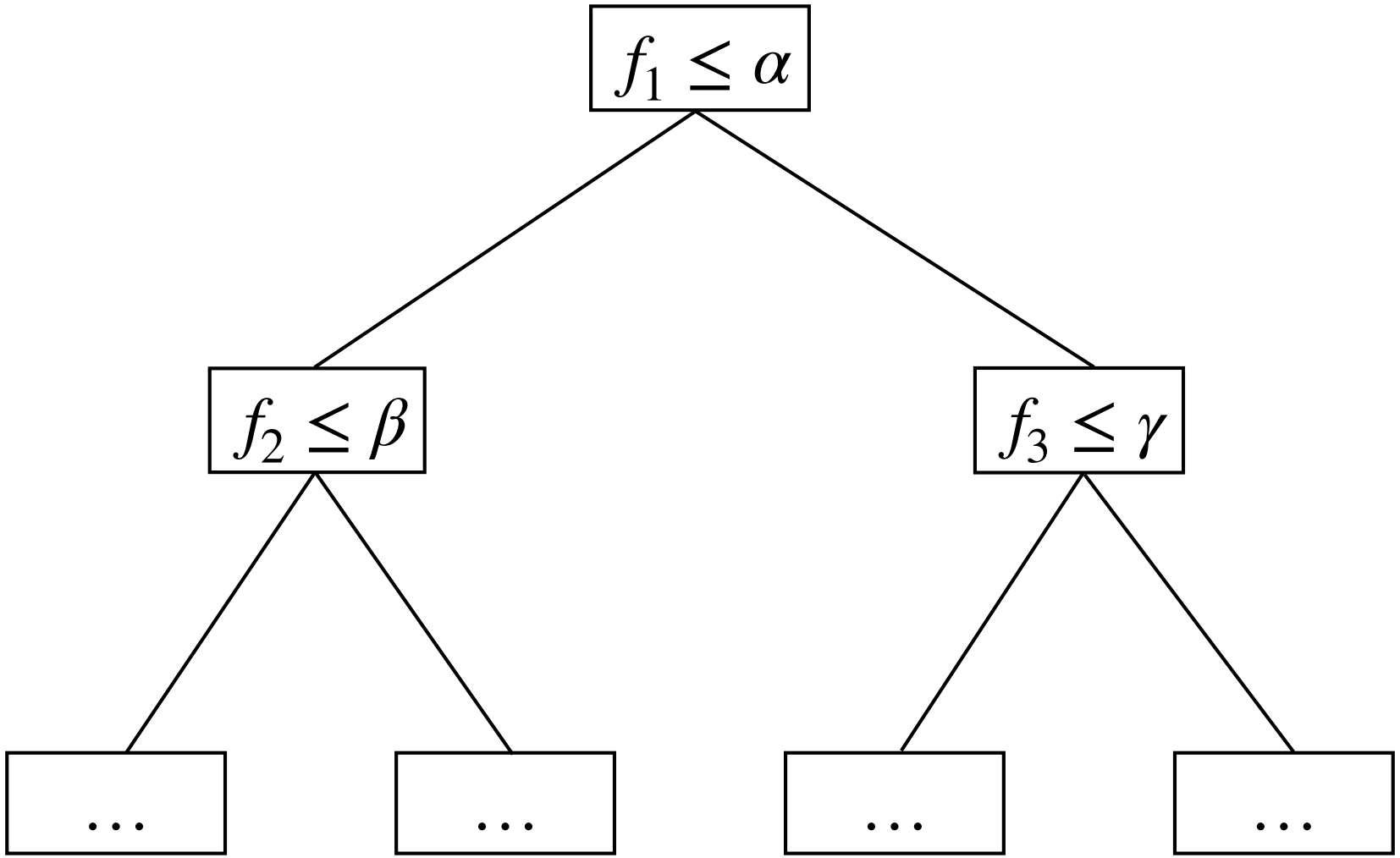
Minimize the length of $S \Leftrightarrow$ pick a (reticulated) cherry of N

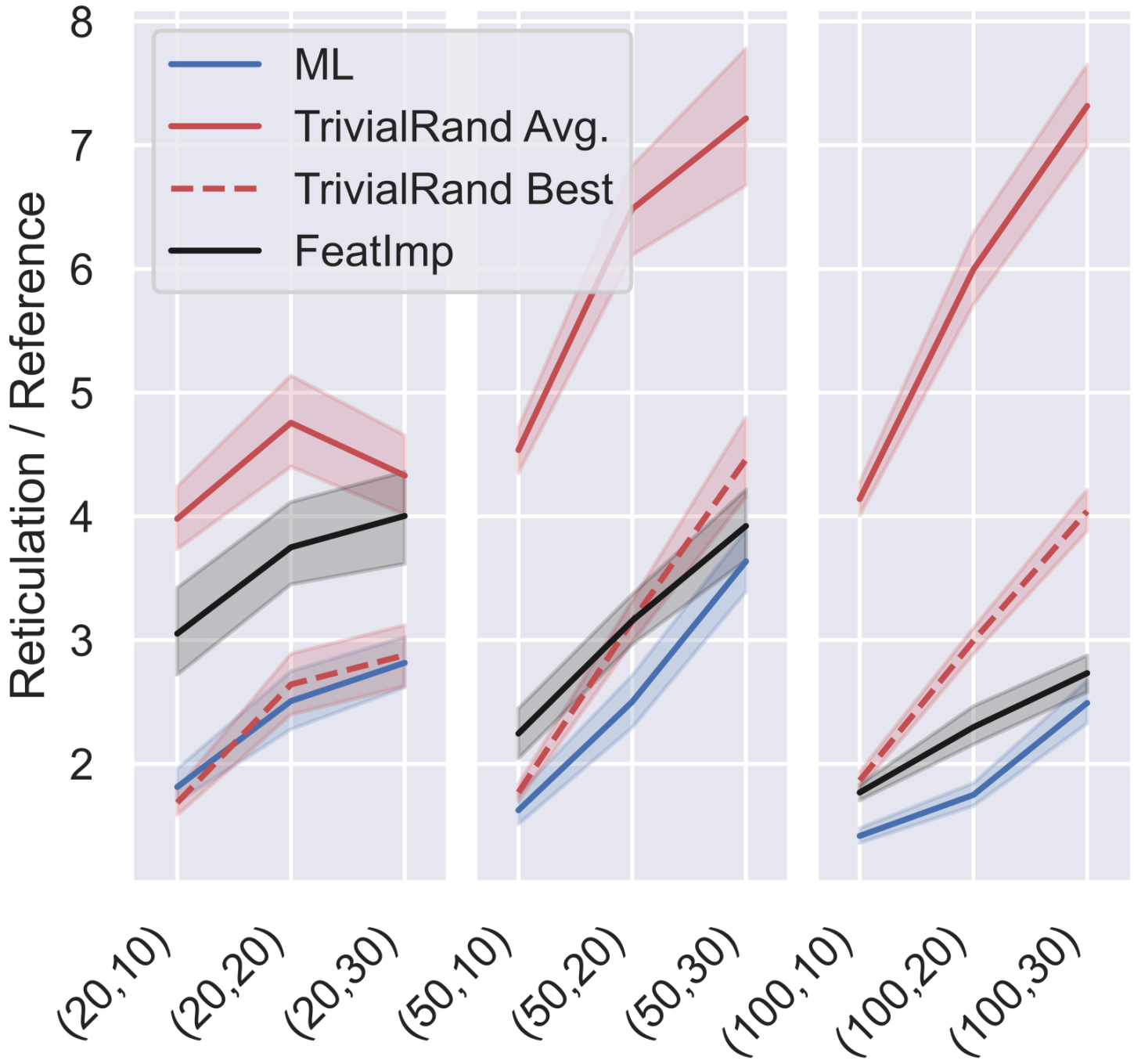

N
a b c d
T_1
a b c d
T_2
a b c d


c d
a b
c b
ML
good
bad
good

A DATAPOINT

DATA FOR CHERRY (x, y)	
INPUT: 19 features	**OUTPUT**
- Avg leaf distance of x and y - Number of trees with (x, y) - Avg depth of (x, y) - Avg depth of x - Avg depth of y $\vdots$	**Good (1) or bad (0)**


$f_1 \leq \alpha$
$f_2 \leq \beta$
$f_3 \leq \gamma$
...
...
...
...


Reticulation / Reference
8
7
6
5
4
3
2
ML
TrivialRand Avg.
TrivialRand Best
FeatImp
(20,10)
(20,20)
(20,30)
(50,10)
(50,20)
(50,30)
(100,10)
(100,20)
(100,30)

https://www.youtube.com/watch?v=Unzc731iCUY&ab_channel=MITOpenCourseWare