The complexity of computing the RNNI distance between phylogenetic trees

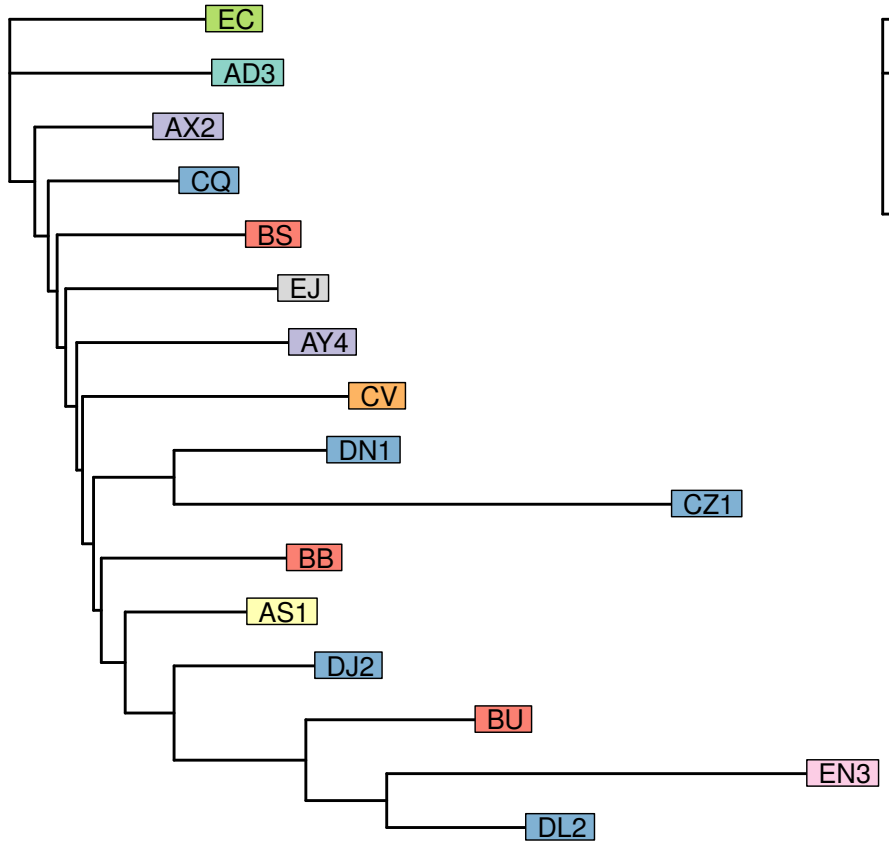
Lena Collienne

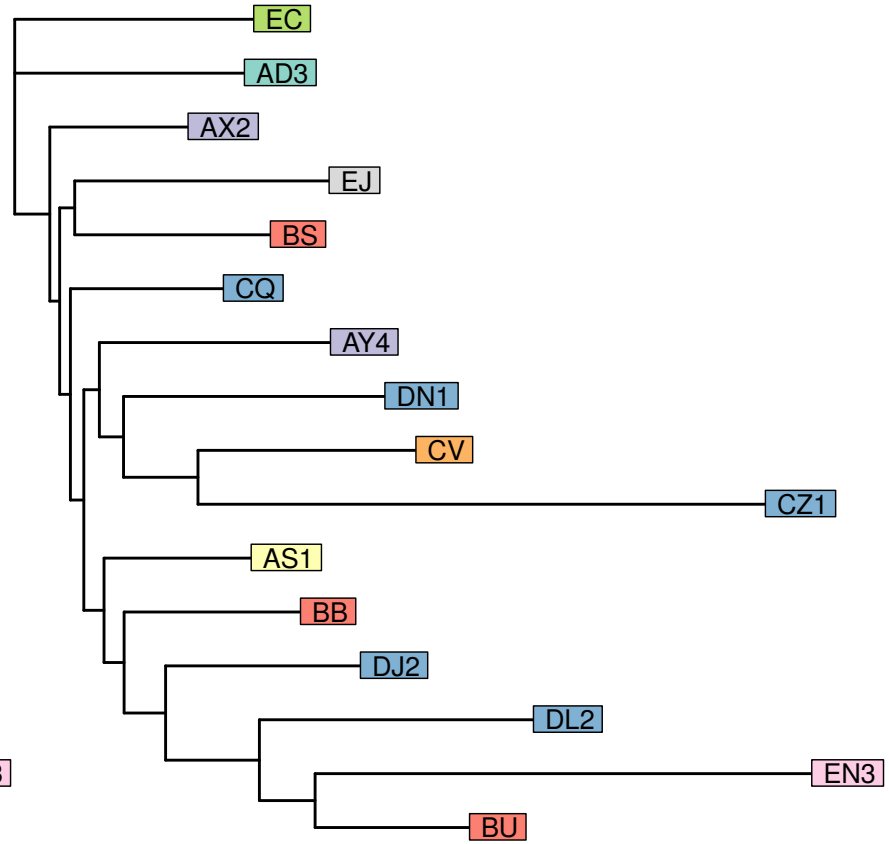
Biological Data Science Lab
Department of Computer Science
University of Otago

25/08/2020

Phylogenetic trees

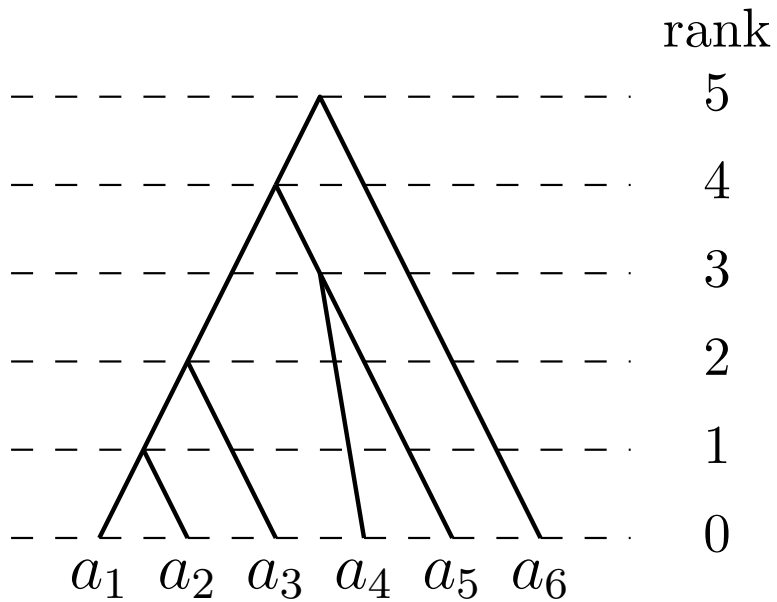
We want distances!



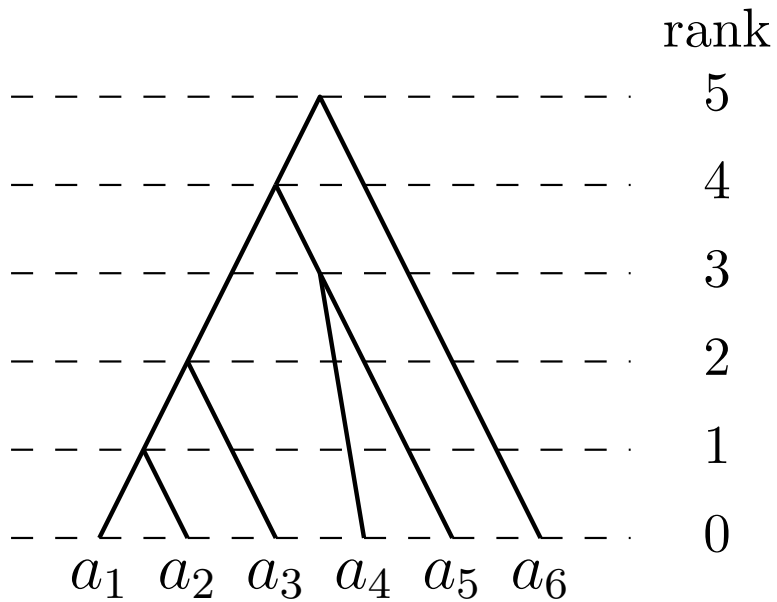


Breast **Pancreas** **Lung lesion**
Thyroid **Vertebra C2** **Eye**
Lung (main tumour) **Kidney** **Brain**

Ranked Trees



Ranked Trees



There are $\frac{(n-1)!n!}{2^{n-1}}$ such ranked trees with n leaves

A Tree Space

T_5

T_1 T_8

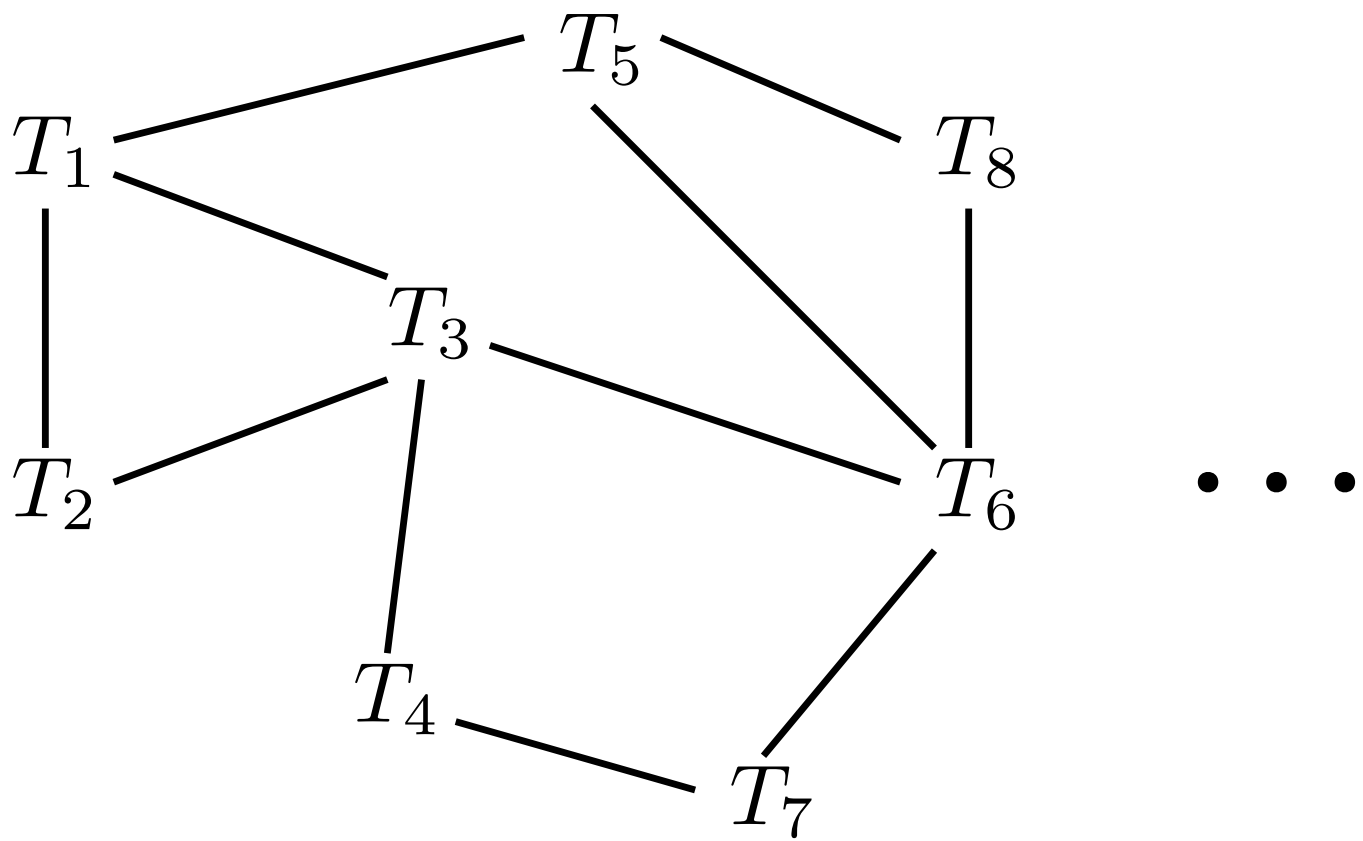
T_3

T_2 T_6 $\bullet \bullet \bullet$

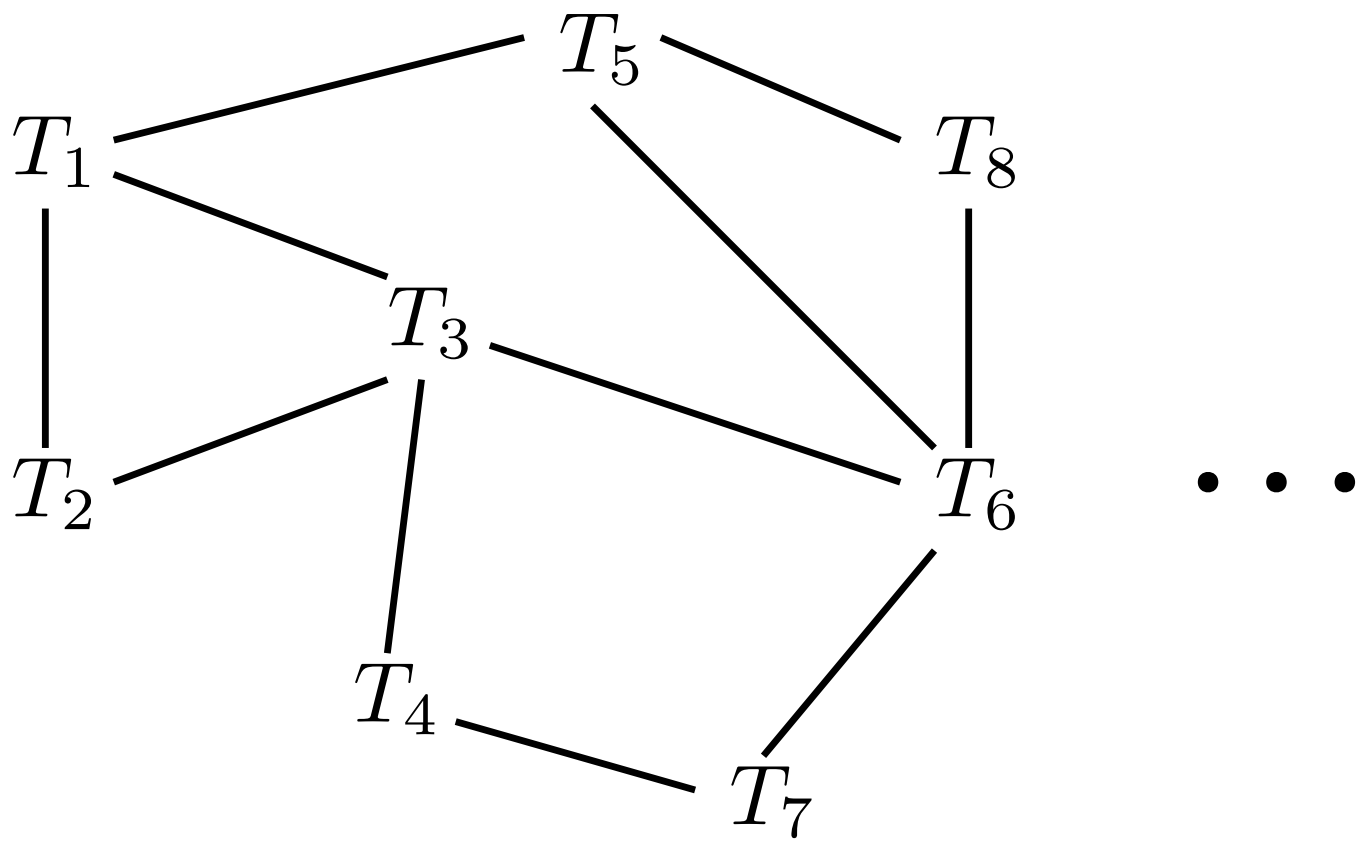
T_4

T_7

A Tree Space



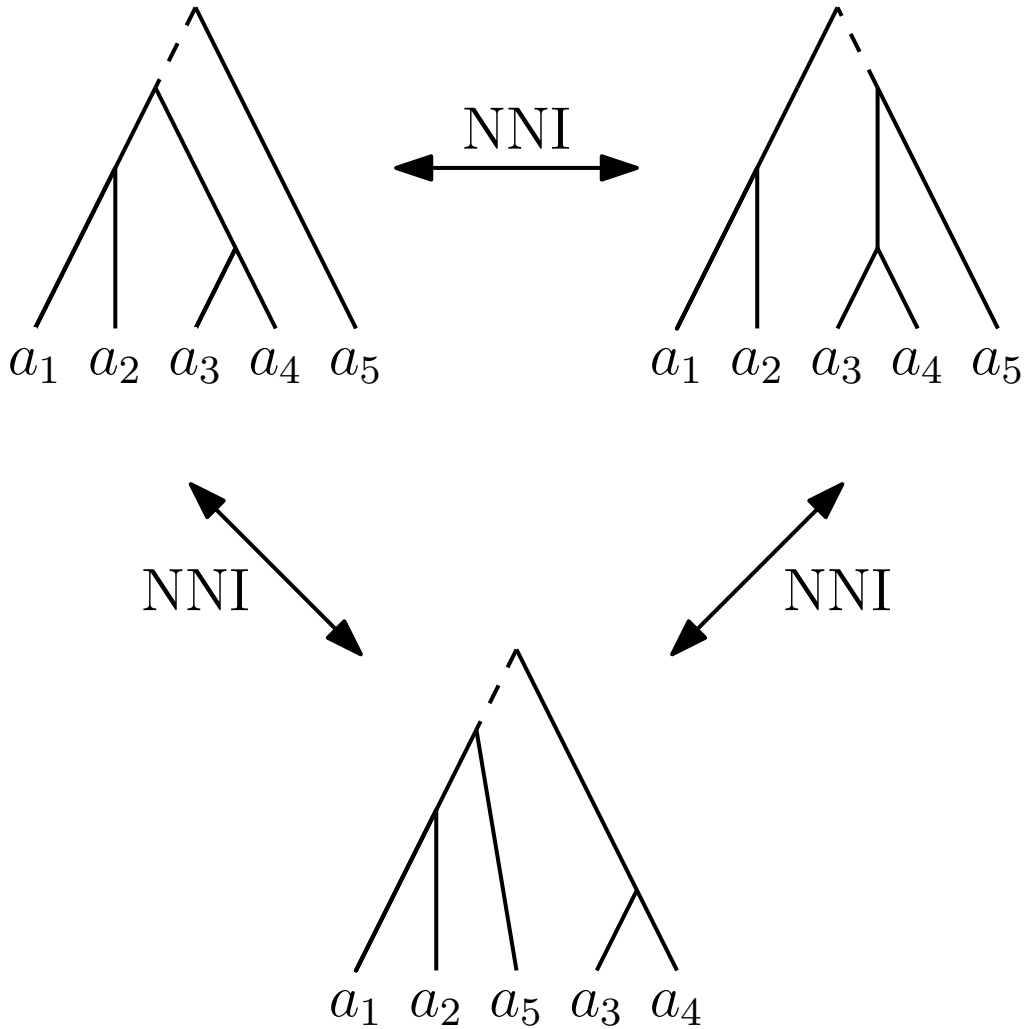
A Tree Space



Problem: Distances in all known tree-rearrangement based spaces are $\mathcal{NP}$-hard to compute

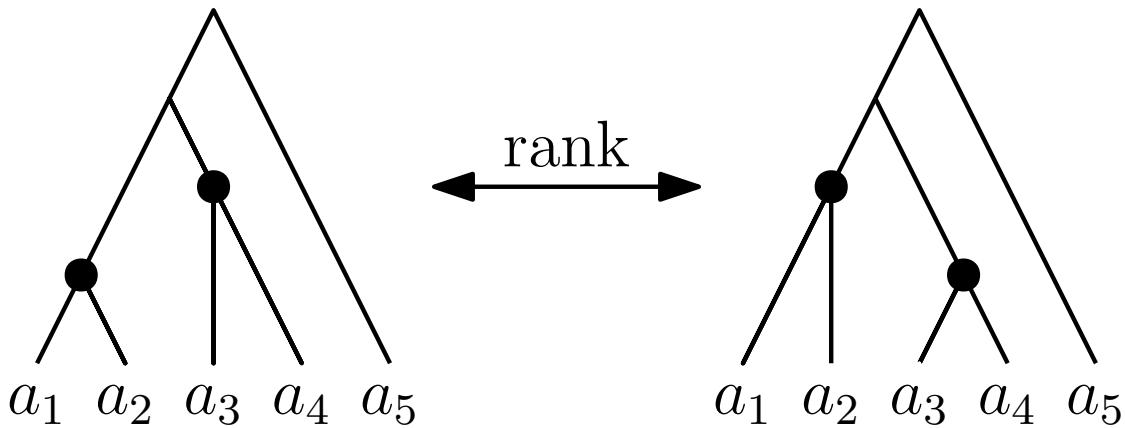
The RNNI space

NNI neighbours



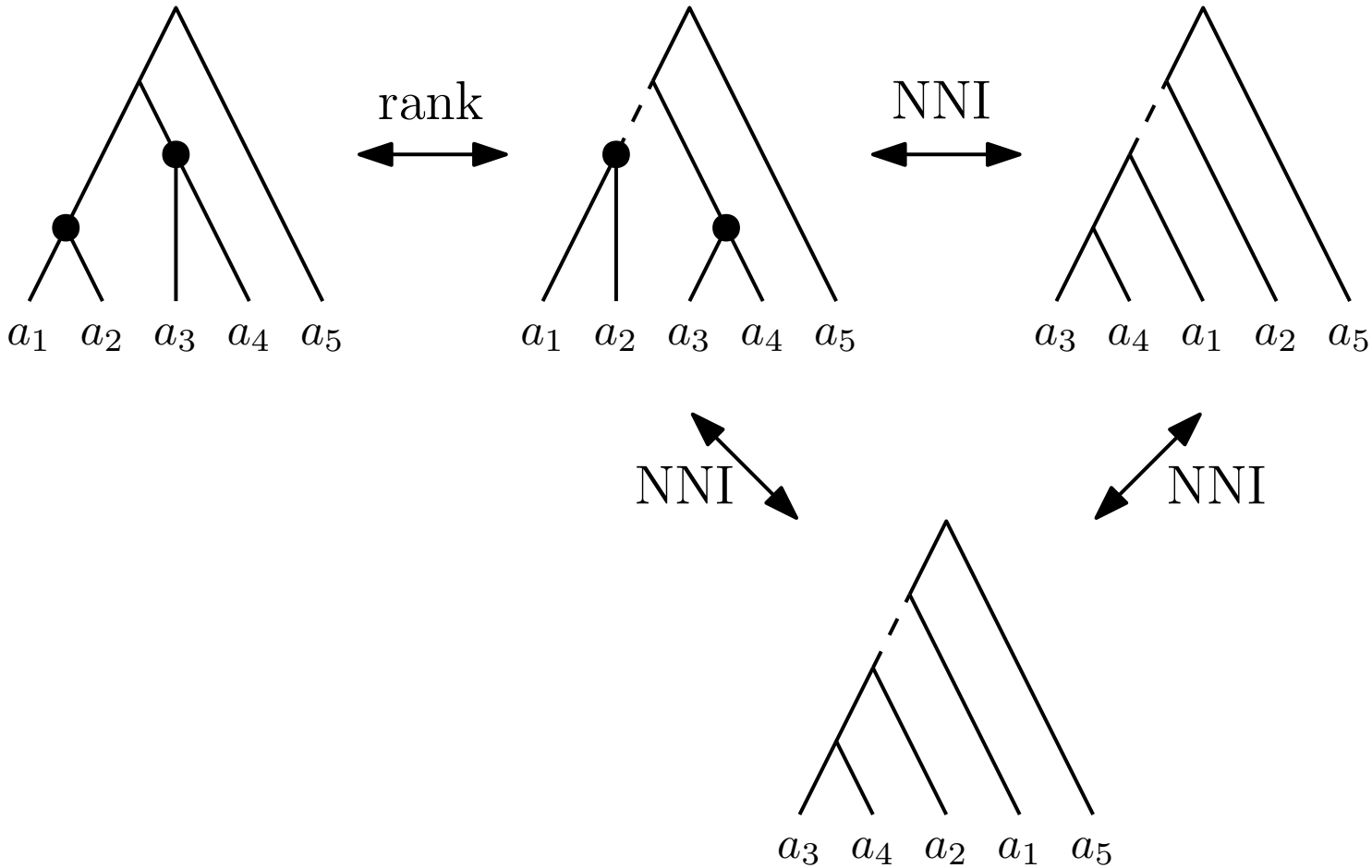
The RNNI space

Rank move

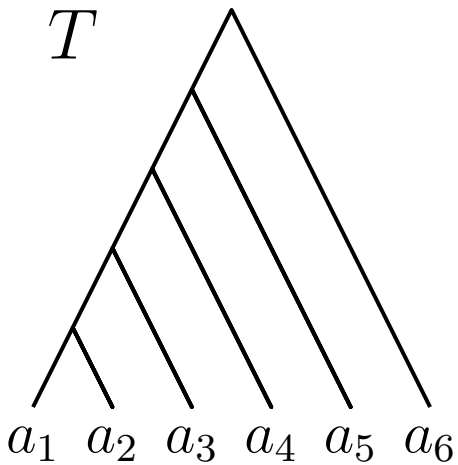


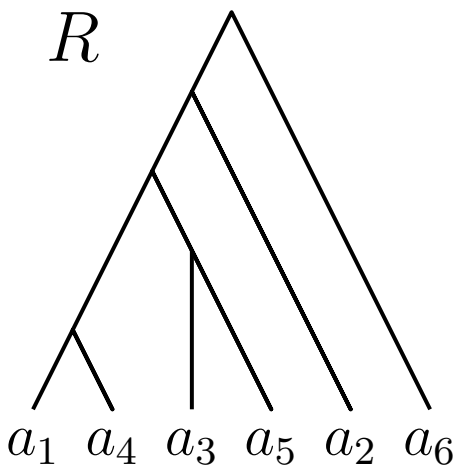
The RNNI space

Rank move

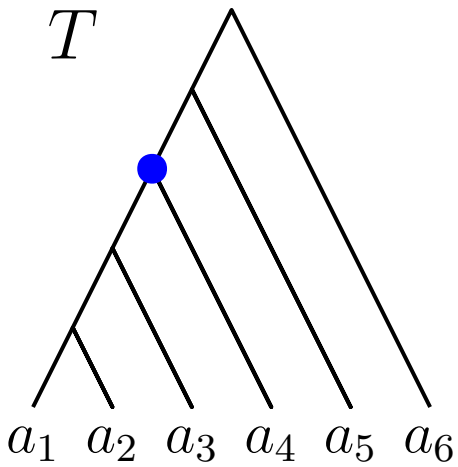


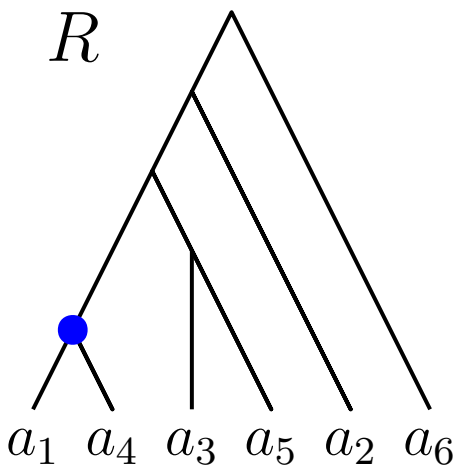
FindPath



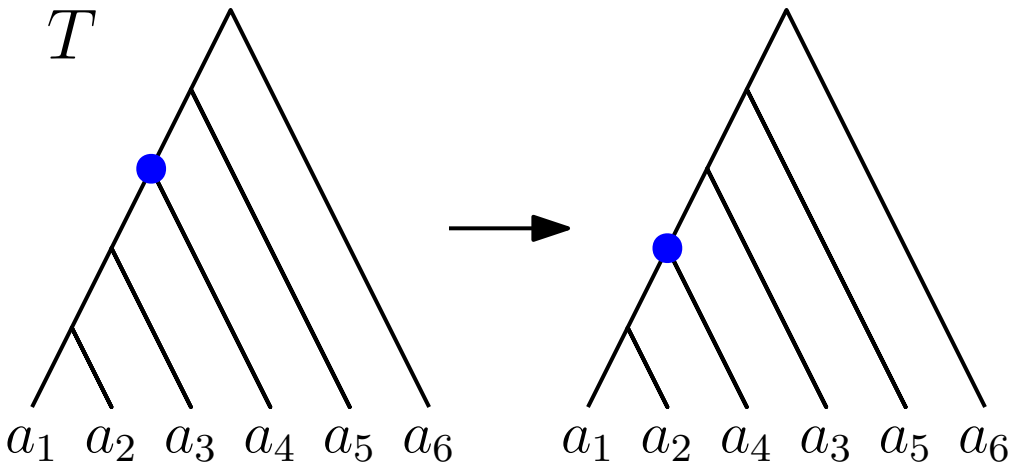


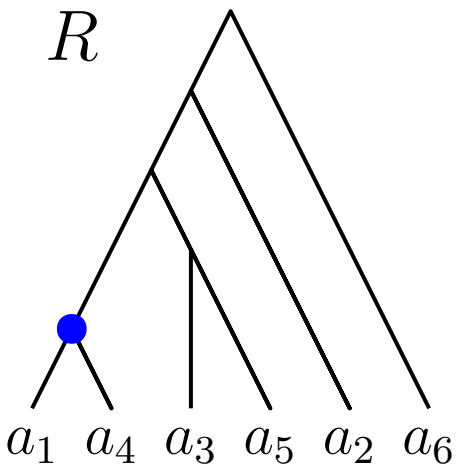
FindPath



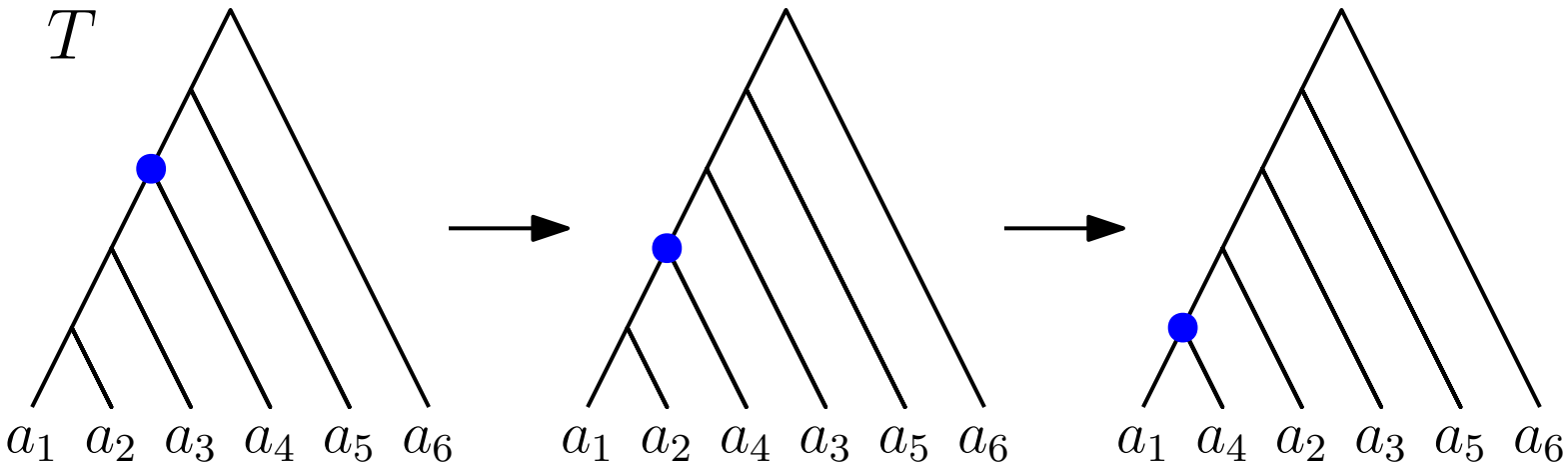


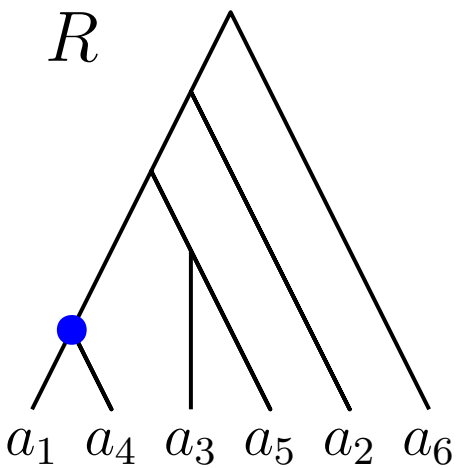
FindPath



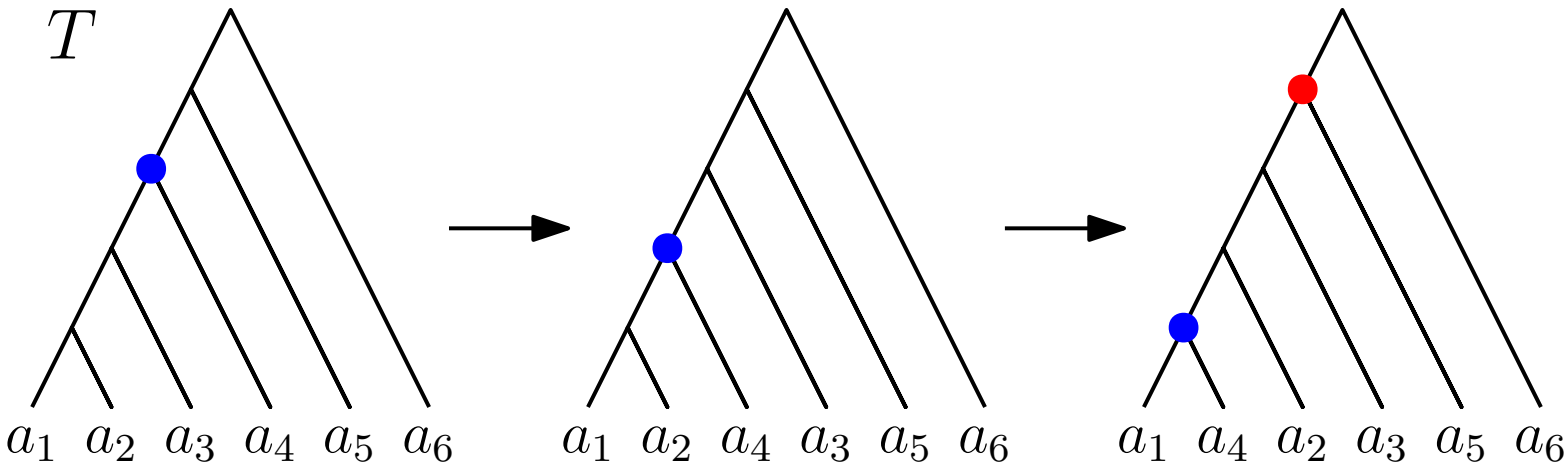


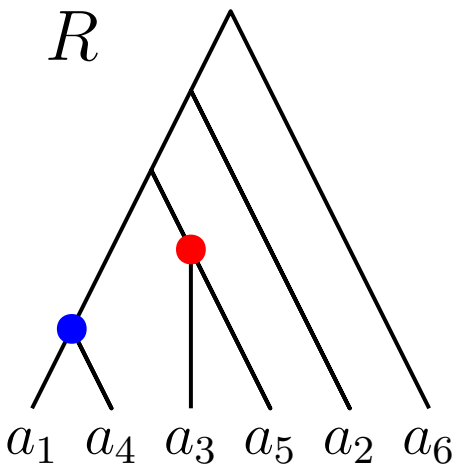
FindPath


T
a1 a2 a3 a4 a5 a6
a1 a2 a4 a3 a5 a6
a1 a4 a2 a3 a5 a6

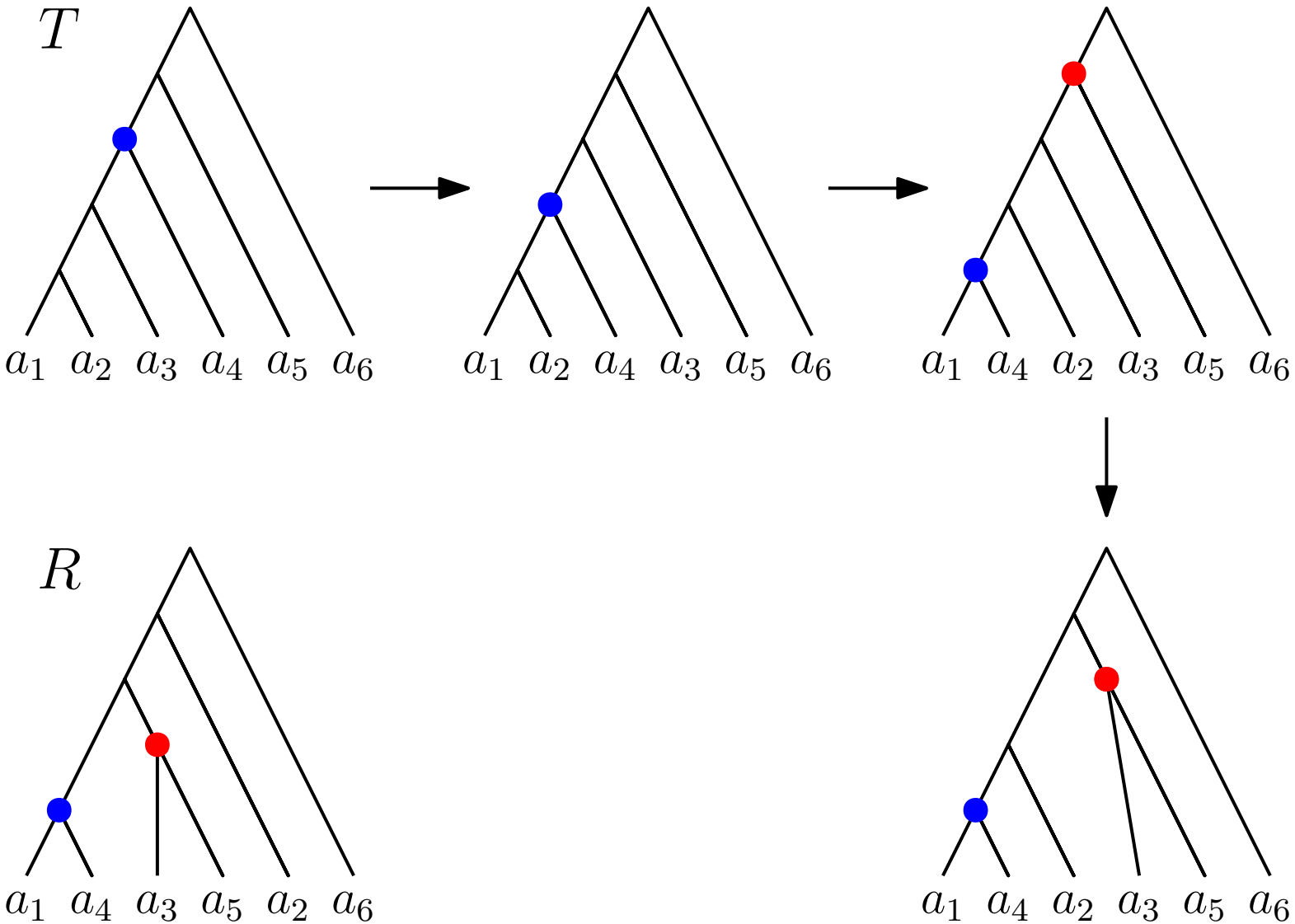

R
a1 a4 a3 a5 a2 a6

FindPath

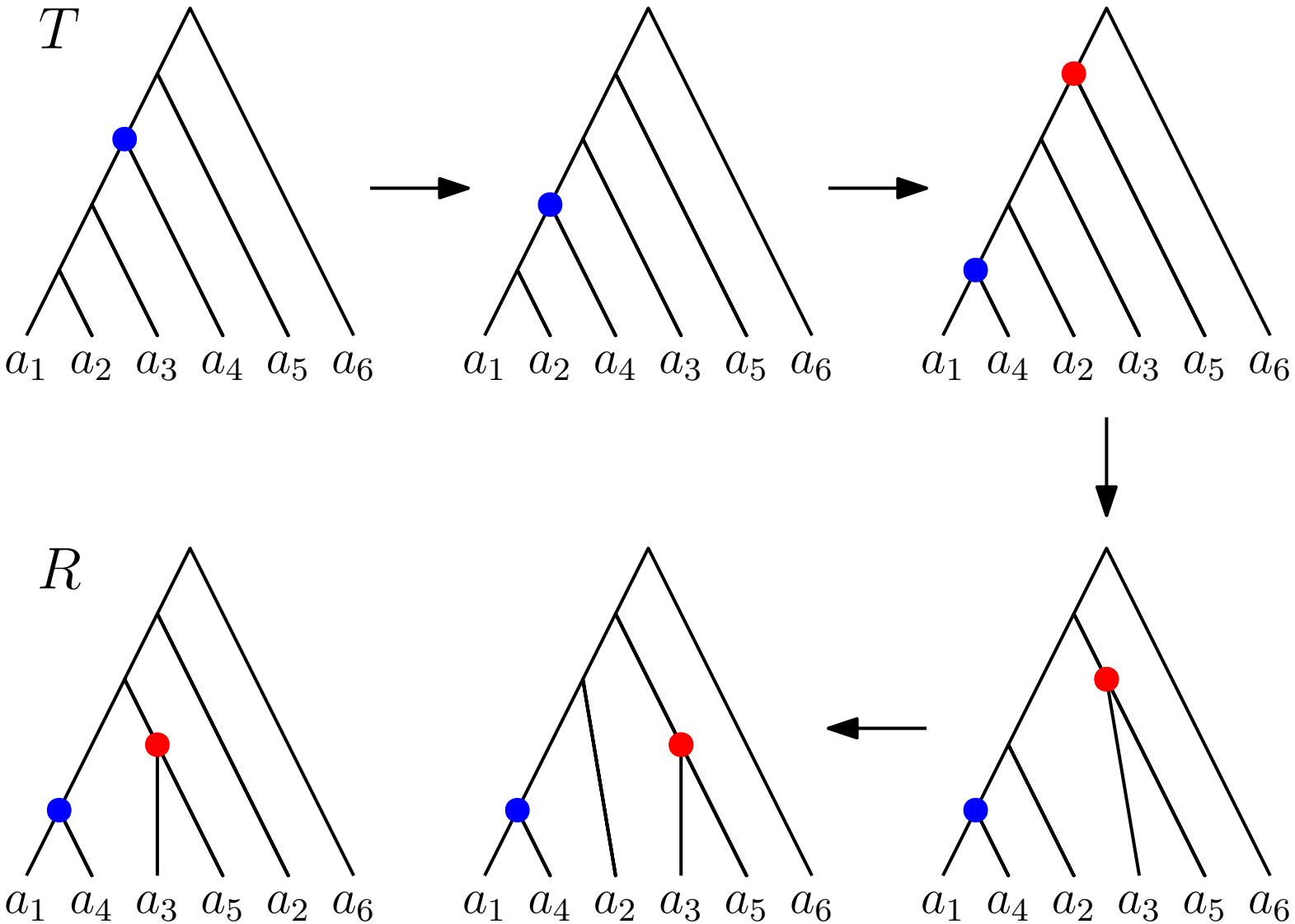

T
a1 a2 a3 a4 a5 a6
a1 a2 a4 a3 a5 a6
a1 a4 a2 a3 a5 a6


R
a1 a4 a3 a5 a2 a6

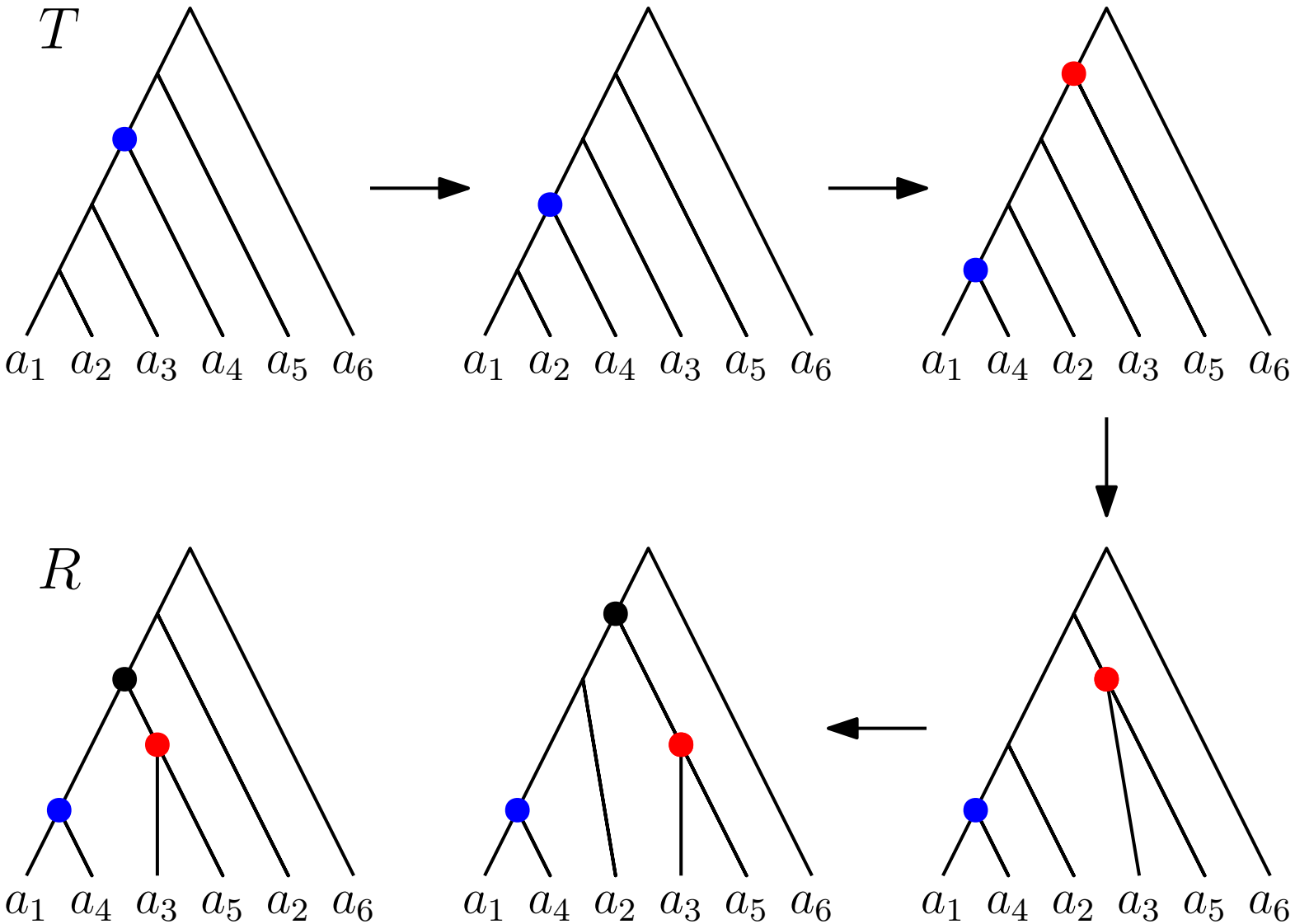
FindPath


T
a_1 a_2 a_3 a_4 a_5 a_6
a_1 a_2 a_4 a_3 a_5 a_6
a_1 a_4 a_2 a_3 a_5 a_6
R
a_1 a_4 a_3 a_5 a_2 a_6
a_1 a_4 a_2 a_3 a_5 a_6

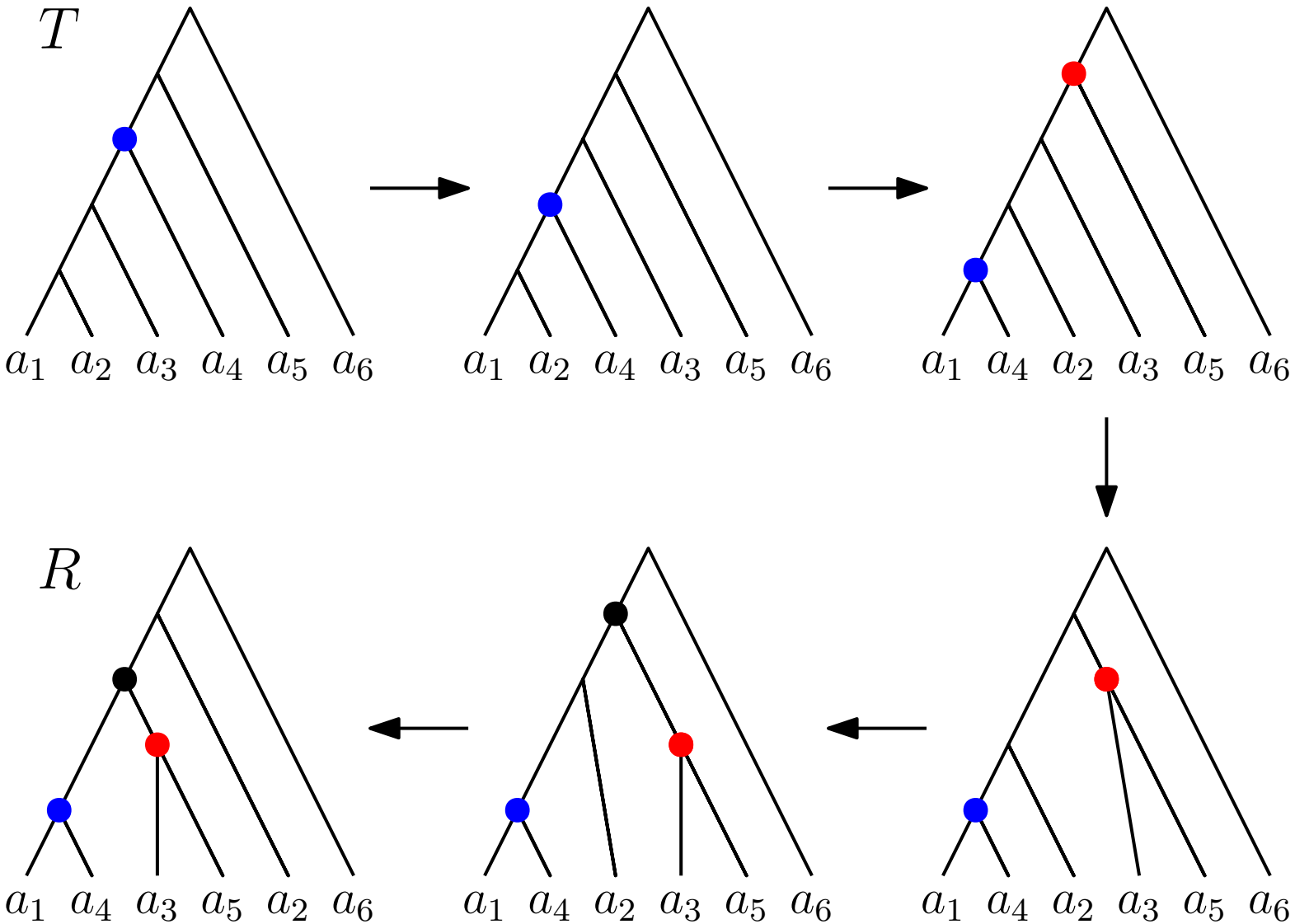
FindPath


T
a1 a2 a3 a4 a5 a6
a1 a2 a4 a3 a5 a6
a1 a4 a2 a3 a5 a6
a1 a4 a2 a3 a5 a6
a1 a4 a3 a5 a2 a6
R
a1 a4 a3 a5 a2 a6

FindPath


T
a1 a2 a3 a4 a5 a6
a1 a2 a4 a3 a5 a6
a1 a4 a2 a3 a5 a6
a1 a4 a2 a3 a5 a6
a1 a4 a3 a5 a2 a6
R
a1 a4 a3 a5 a2 a6

FindPath


T
a1 a2 a3 a4 a5 a6
a1 a2 a4 a3 a5 a6
a1 a4 a2 a3 a5 a6
a1 a4 a2 a3 a5 a6
a1 a4 a3 a5 a2 a6
a1 a4 a3 a5 a2 a6
R

FindPath

Theorem

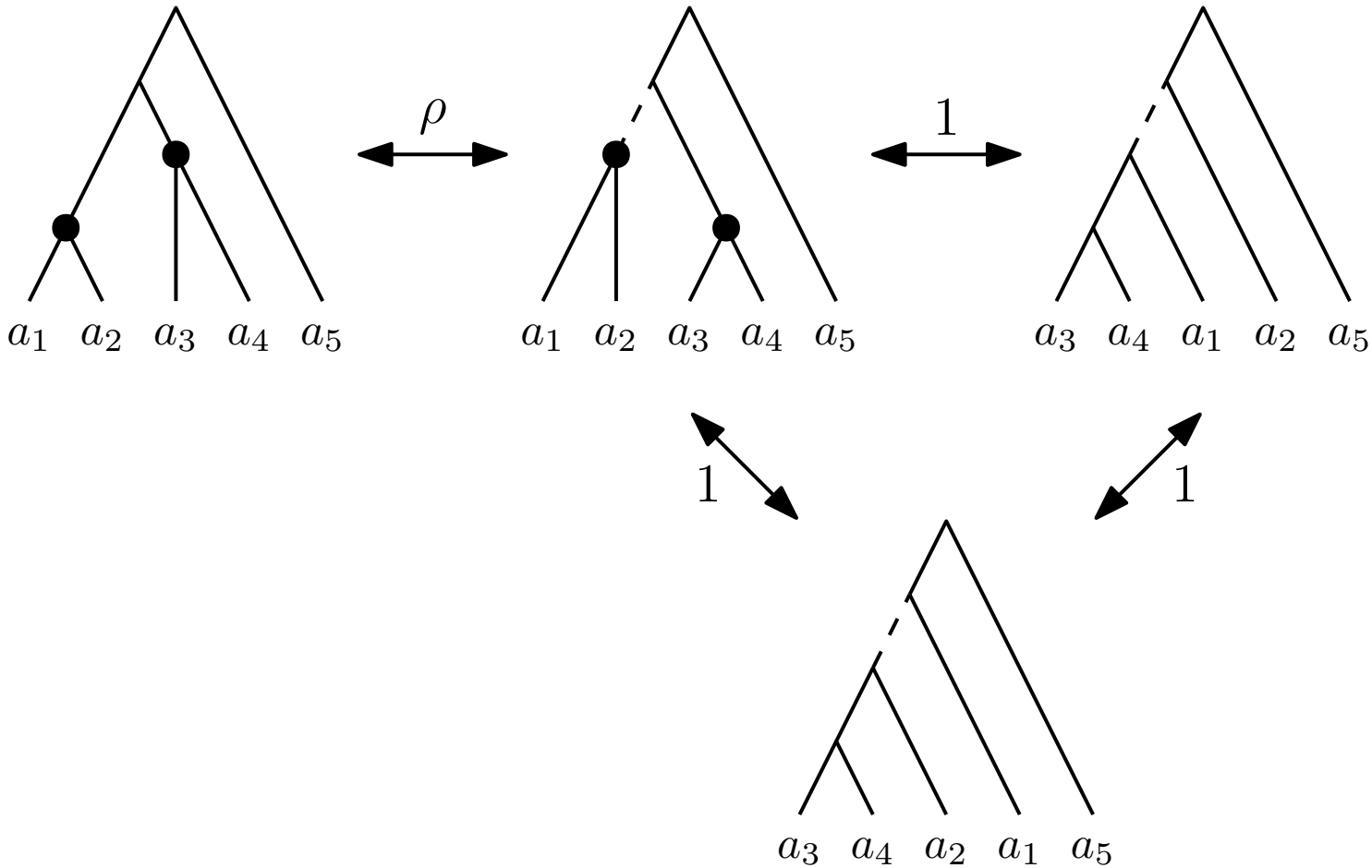
FindPath *computes shortest paths in* RNNI *in* $\mathcal{O}(n^2)$.

FindPath

Theorem

FindPath *computes shortest paths in* RNNI *in* $\mathcal{O}(n^2)$.

RNNI(ρ)



$\mathrm{RNNI}(\rho)$

$\mathrm{RNNI}(\rho)$-SP:

INSTANCE: A pair of trees T and R
FIND: A path of minimal weight between T and R in $\mathrm{RNNI}(\rho)$

$\mathrm{RNNI}(\rho)$

$\mathrm{RNNI}(\rho)$-SP:

INSTANCE: A pair of trees T and R
FIND: A path of minimal weight between T and R in $\mathrm{RNNI}(\rho)$

- $\mathrm{RNNI}(0)$-SP is $\mathcal{NP}$-hard (NNI)

RNNI(ρ)

RNNI(ρ)-SP:

INSTANCE: A pair of trees T and R
FIND: A path of minimal weight between T and R in RNNI(ρ)

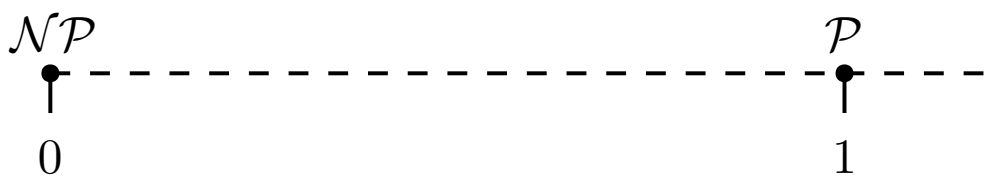
- RNNI(0)-SP is $\mathcal{NP}$-hard (NNI)
- RNNI(1)-SP is polynomial

RNNI(ρ)

RNNI(ρ)-SP:

INSTANCE: A pair of trees T and R
FIND: A path of minimal weight between T and R in RNNI(ρ)

- RNNI(0)-SP is $\mathcal{NP}$-hard (NNI)
- RNNI(1)-SP is polynomial

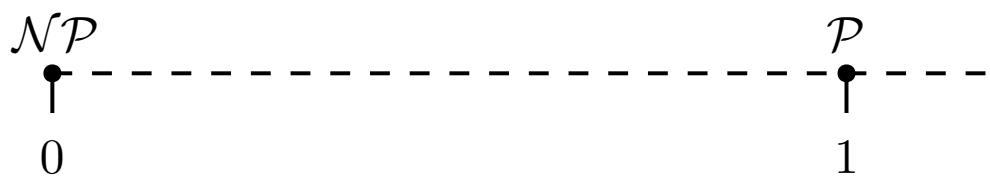


RNNI(ρ)

RNNI(ρ)-SP:

INSTANCE: A pair of trees T and R
FIND: A path of minimal weight between T and R in RNNI(ρ)

- RNNI(0)-SP is $\mathcal{NP}$-hard (NNI)
- RNNI(1)-SP is polynomial



- What about $0 < \rho < 1$?

Thank you