Discrete time-trees

Lena Collienne

Ernst-Moritz-Arndt-Universität Greifswald

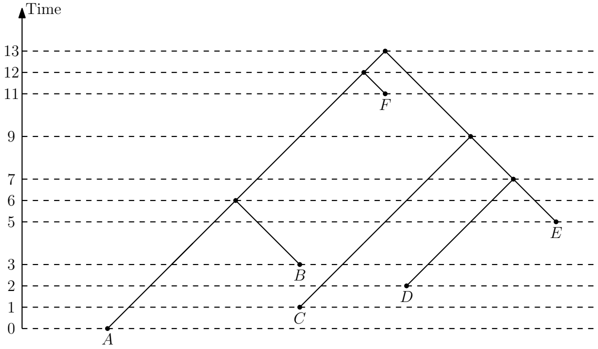
09.08.2017

Table of contents

Motivation

ERNST MORITZ ARNDT
UNIVERSITÄT GREIFSWALD
Wissen lockt. Seit 1456

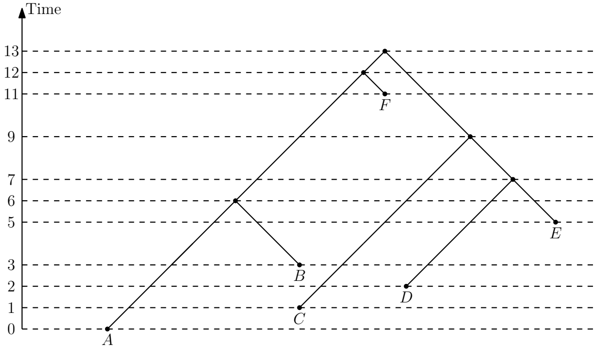
time-trees: branch length $\widehat{=}$ time
discrete time-trees: branch lengths are non-negative integers



Motivation

time-trees: branch length $\widehat{=}$ time

discrete time-trees: branch lengths are non-negative integers



Why this discretization?

$\rightarrow$ Generalization of NNI $\Rightarrow$ approximation algorithms?

$\rightarrow$ Understanding behaviour of MCMC on time-trees

Phylogenetic trees

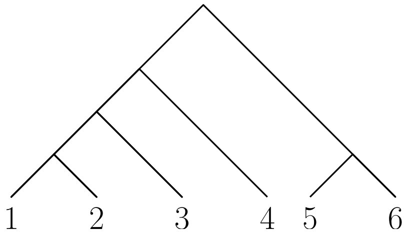


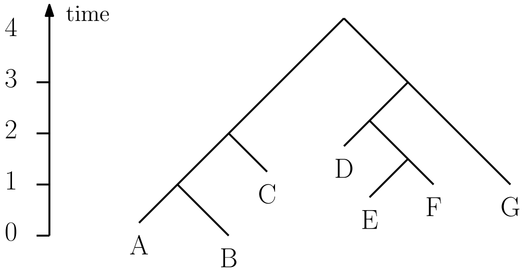
Figure: Phylogenetic tree on set $S = \{1, \ldots, 6\}$

Definition

A *rooted phylogenetic tree* $\mathcal{T}$ on a set S is a pair (T, Φ), where T is a rooted binary tree and Φ is a bijective map from S to the leaf set of T.

Phylogenetic trees

Time-trees

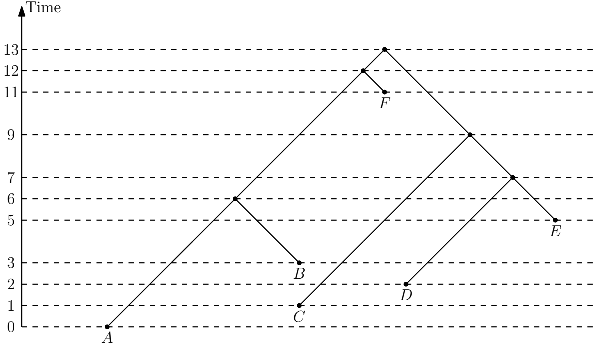


Definition

A *time-tree* is a rooted phylogenetic tree where each node is assigned a non-negative real number (time/divergence date). Each node has a smaller time than its parent.

Phylogenetic trees

Discrete time-trees



Definition

A *discrete time-tree* is a time-tree where all assigned times are distinct and from the set of non-negative integers.

Definition

An *ultrametric* discrete time tree is a discrete time tree where all leaves have the same time.

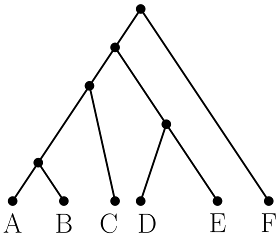


Figure: Ultrametric discrete time tree

Phylogenetic trees

Discrete time-trees

Definition

The *rank* of a node is the number of nodes in the tree with strictly smaller time.

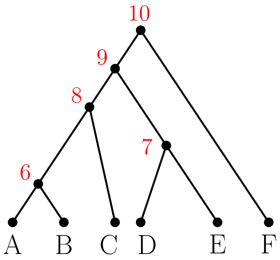


Figure: Ultrametric discrete time tree with ranks

Phylogenetic trees

Discrete time-trees

Definition

A pair of nodes v_1, v_2 in a tree is called *event interval* if $|rank(v_1) - rank(v_2)| = 1$

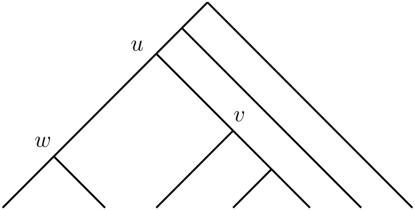


Figure: u, v and v, w are event intervals, u, w not

Hierarchy of discrete time-trees

Nearest Neighbour Interchange

The NNI graph builds the bottom level of our hierarchy:

Definition

Two discrete time-trees $\mathcal{T}$ and $\mathcal{R}$ are NNI neighbours, if there exist edges e in $\mathcal{T}$ and f in $\mathcal{R}$ such that:

- both edges are not adjacent to a leaf and
- the graph obtained from $\mathcal{T}$ by shrinking e to a vertex is isomorphic to the graph obtained from $\mathcal{R}$ by shrinking f.

Hierarchy of discrete time-trees

ERNST MORITZ ARNDT UNIVERSITÄT GREIFSWALD
Wissen lockt. Seit 1456

Nearest Neighbour Interchange

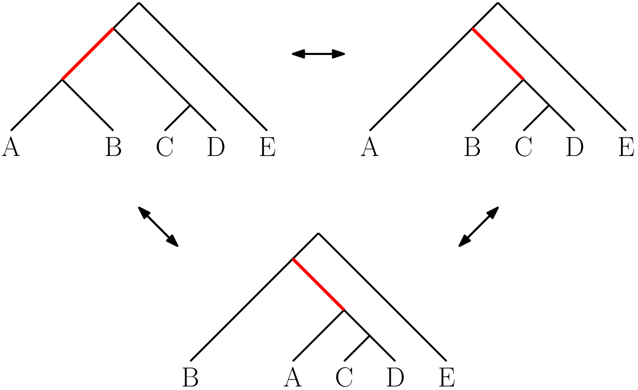


Figure: All possible NNI moves on the interval I

Nearest Neighbour Interchange

Definition

The graph $G = (V, E)$ with:

- $V :=$ set of discrete time-trees on n leaves
- $E := \{(u, v) |\ u$ and v are NNI neighbours$\}$

is called NNI graph. We denote this graph by DtT_0.

Hierarchy of discrete time-trees

$\mathrm{DtT_m}$

The level $m > 0$ of our hierarchy:

Definition

The graph $\mathrm{DtT_m} = (V, E)$ is defined as follows:
V := set of discrete time-tree $\mathcal{T}$ on n leaves with event intervals of length $\leq m$ There is an edge between two trees $\mathcal{T}, \mathcal{R}$, if:

- they are NNI neighbours
- there are two vertices u, v which swap their ranks
- there is a length move between $\mathcal{T}$ and $\mathcal{R}$

Hierarchy of discrete time-trees

Event intervals

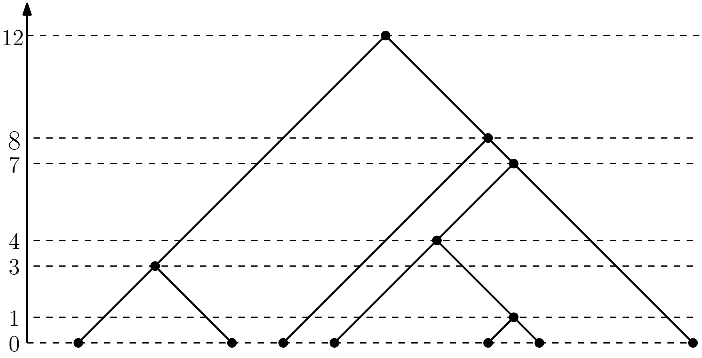


Figure: All event intervals have lengths less or equal to $m = 4$

Hierarchy of discrete time-trees

Rank swap

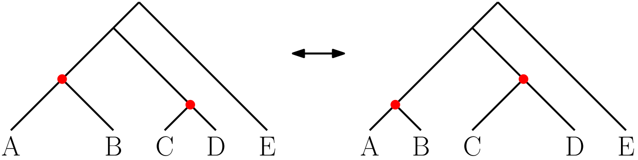


Figure: A rank swap of the two nodes

Hierarchy of discrete time-trees

Length move

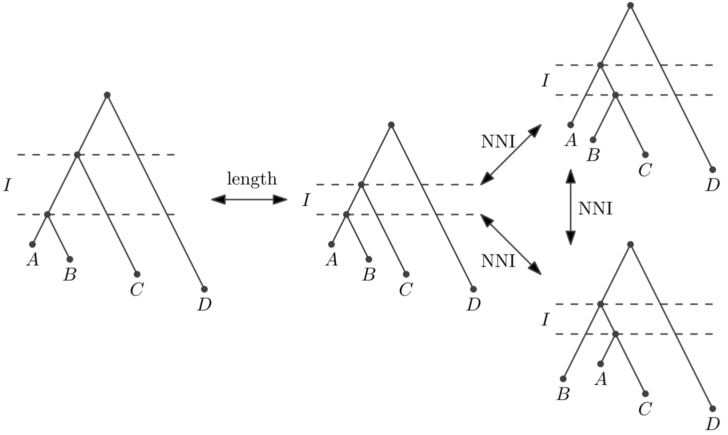
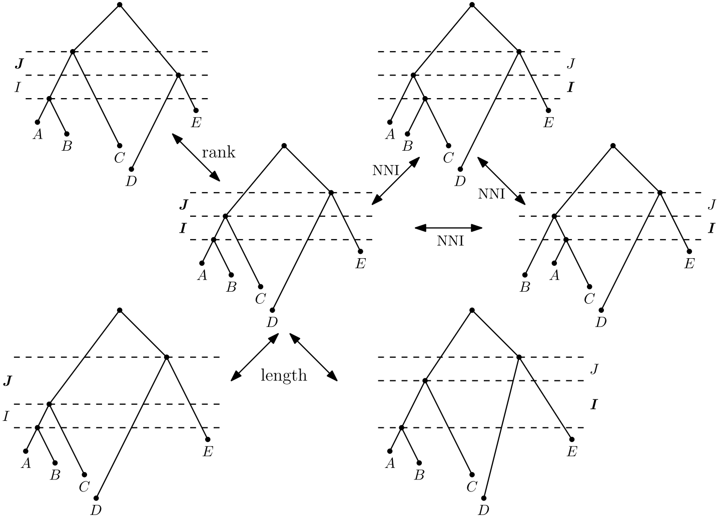


Figure: Length and NNI moves on the event interval I

Hierarchy of discrete time-trees

Summary



RNNI

Definition

ERNST MORITZ ARNDT UNIVERSITÄT GREIFSWALD

Wissen lockt. Seit 1456

Definition

The RNNI graph is the graph DtT_1

RNNI

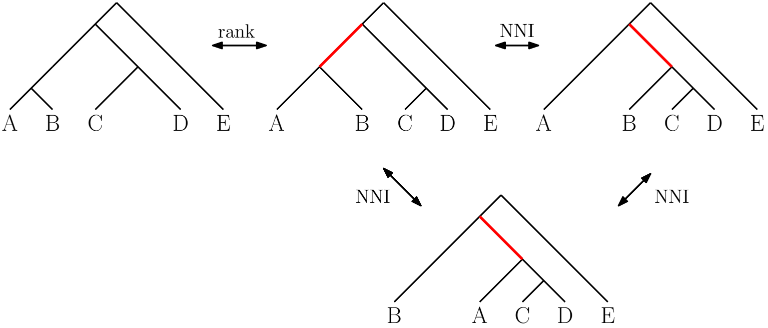
Definition

Definition

The RNNI graph is the graph DtT_1

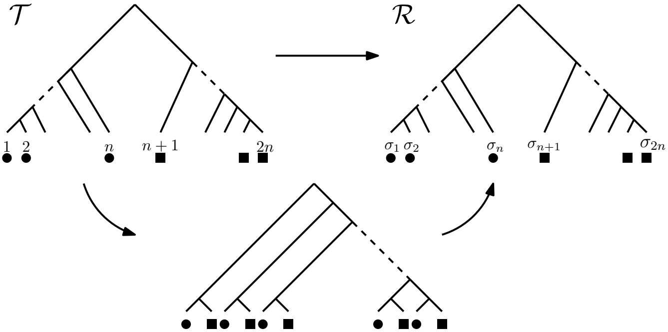
$\Rightarrow$ There are only rank and NNI moves in the RNNI graph

Now: RNNI on ultrametric trees

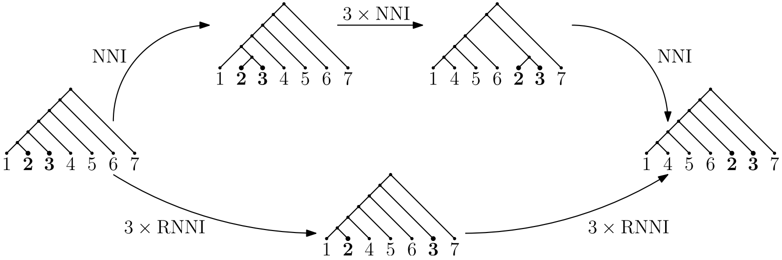


Theorem

It is NP*-hard to compute* NNI *distances.*



Is computing RNNI distances NP-hard, too?



RNNI

Complexity

In general:

Theorem

The set of caterpillar trees is convex in the RNNI *space.*

RNNI

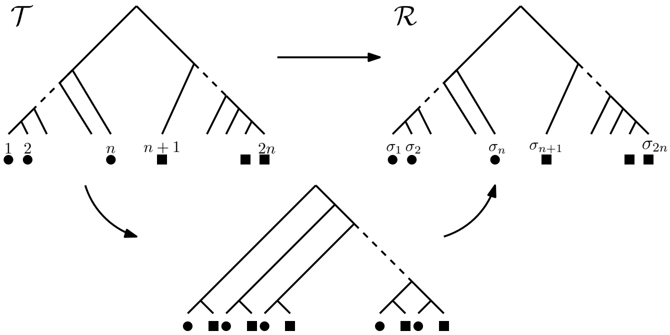
Complexity

In general:

Theorem

The set of caterpillar trees is convex in the RNNI space.

$\Rightarrow$ The proof of the $\mathrm{NP}-$hardness of NNI can't be used for RNNI



RNNI

Split Theorem

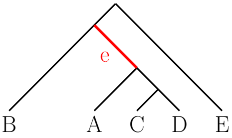


Figure: The edge e leads to the split $BE|ACD$

Conjecture (Split Theorem)

In RNNI *it holds:*
If a split given by an edge is present in two trees $\mathcal{T}$ *and* $\mathcal{R}$ *then this split is present in every tree on every shortest path from* $\mathcal{T}$ *to* $\mathcal{R}$.

Split Theorem

The Split Theorem is not true for NNI:

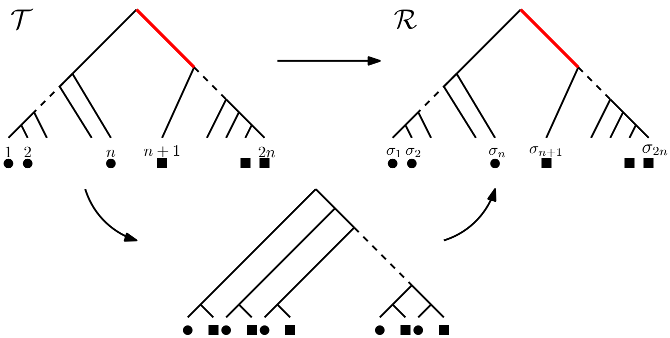
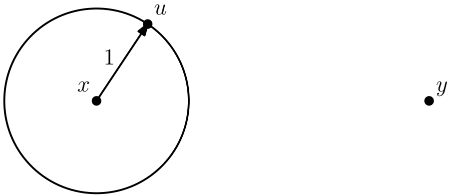


Figure: Split present in $\mathcal{T}$ and $\mathcal{R}$: $12...n|n+1...2n$

Theorem

Let x and y be trees and $N(x)$ the 1-RNNI-neighbourhood of x. Then the following statement holds for the RNNI space:
$|\{u \in N(x) : d(u,y) \leq d(x,y)\}| \leq 3 * d(x,y)$



Theorem

Let x be a caterpillar tree, y a caterpillar depth p tree and $N(x)$ the 1-RNNI-neighbourhood of x. Then the following statement holds for the RNNI space:
$|\{u \in N(x) : d(u, y) \leq d(x, y)\}| \leq 3 * d(x, y)$

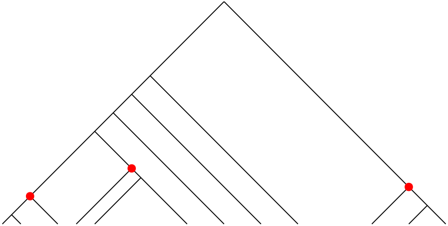


Figure: A caterpillar depth 3 tree

RNNI

Main idea for the proof: differ between topologies and distances to y

Topologies:

- Same topology as x (caterpillar)
- One additional cherry

Distances to y:

- $d(u, y) = d(x, y) - 1$
- $d(u, y) = d(x, y)$

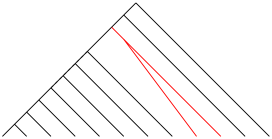


Figure: Caterpillar tree with additional cherry

RNNI

Lemma

The number of caterpillar trees u in $N(x)$ with $d(u, y) = d(x, y)$ equals the number of caterpillar trees with an additional cherry in $N(x)$ and distance $d(x, y) - 1$ to y.

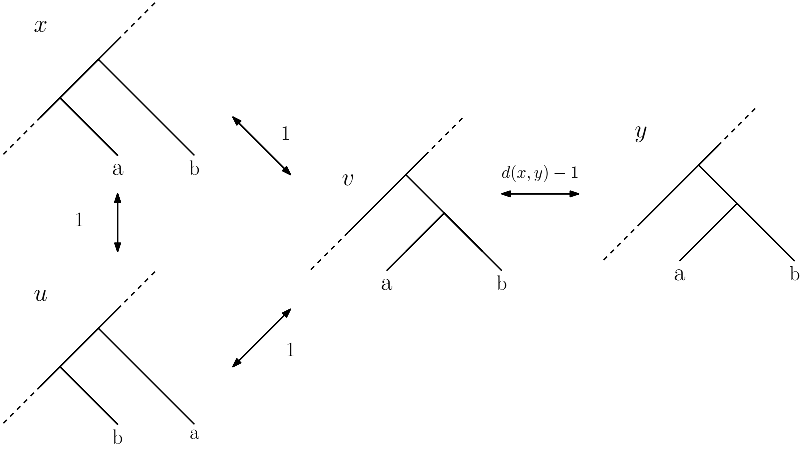


Figure: $d(x, y) = d(u, y) \Rightarrow d(v, y) = d(x, y) - 1$

RNNI

Lemma

The number of caterpillar trees with an additional cherry u *in* $N(x)$ *with* $d(u,y) = d(x,y)$ *equals the number of caterpillar trees in* $N(x)$ *with distance* $d(x,y) - 1$ *to* y*.*

RNNI

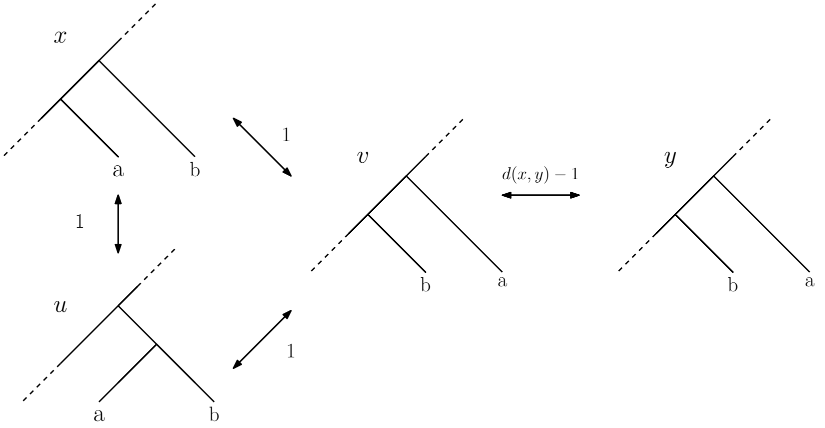
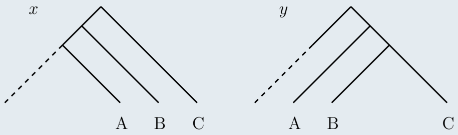


Figure: $d(x, y) = d(u, y) \Rightarrow d(v, y) = d(x, y) - 1$

Proof.

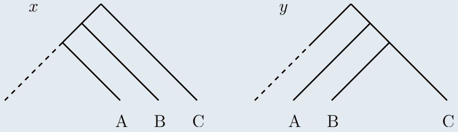
We want: $|\{u \in N(x) : d(u,y) \leq d(x,y)\}| \leq 3 * d(x,y)$

1. Trees $u \in N(x)$ with additional cherry and $d(u,y) = d(x,y) - 1$
   $\rightarrow$ For each of the p caterpillar subtrees of size s_p there might be $s_p - 1$ cherries built new
   $\Rightarrow \sum(s_p - 1)$



Proof.

2. Caterpillar trees $u \in N(x)$ with $d(u, y) = d(x, y) - 1$
   $\rightarrow$ For each caterpillar subtree there are $s_p - 1$ RNNI$-$moves necessary to build that subtree
   $\Rightarrow d(x, y) - \sum(s_p - 1)$



$\Rightarrow$ In total: $2 * \sum(s_p - 1) + 2 * (d(x, y) - \sum(s_p - 1)) = 2d(x, y)$?
No: The running index of both sums may differ.
$\Rightarrow$ case-by-case analysis depending on the positions of cherry taxa of x in y. □

Summary

Conluding we have:

- A hierarchy on discrete time-trees
- The RNNI graph as one level:
  $\rightarrow$ Set of caterpillars is convex
  $\rightarrow$ Some knowledge about neighbourhoods

Summary

Conluding we have:

- A hierarchy on discrete time-trees
- The RNNI graph as one level:
  $\rightarrow$ Set of caterpillars is convex
  $\rightarrow$ Some knowledge about neighbourhoods

Next steps:

- Prove the Split Theorem
- Curvature of RNNI graph $\rightarrow$ MCMC