Five equivalent ways to represent a phylogenetic tree

Jiayue Qi

DK9 progress report, joint work with Professor Schicho

2020.09.24. DK Statusseminar

Doctoral Program Computational Mathematics
Numerical Analysis and Symbolic Computation

RISC RESEARCH INSTITUTE FOR SYMBOLIC COMPUTATION JKU JOHANNES KEPLER UNIVERSITY LINZ

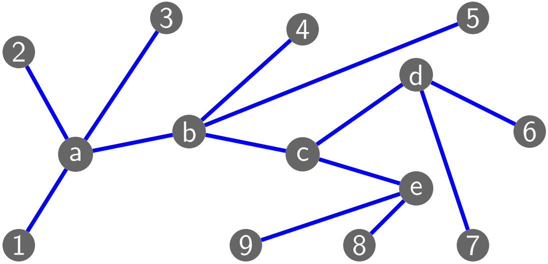
JOHANNES KEPLER UNIVERSITY LINZ

FWF Der Wissenschaftsfonds.

Five equivalent representations

- phylogenetic tree
- phylogenetic set of partitions
- phylogenetic set of cuts
- phylogenetic crossing relations
- phylogenetic equivalences of triples

Example: phylogenetic tree

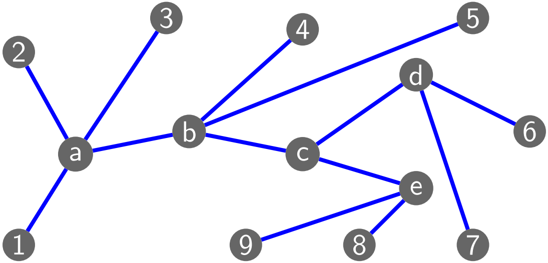


This is a phylogenetic tree with leaf set $N = \{1, 2, \cdots, 9\}$. The set of inner vertices is $V \setminus N = \{a, b, c, d, e\}$.

Definition: phylogenetic tree

- A **phylogenetic tree** with leaf set N is a tree (V, E) with no vertex of degree 2 such that $N \subset V$ is the set of leaves. We call the elements in N **labels** of the tree.
- Two phylogenetic trees are **isomorphic** iff there is a graph isomorphism between them which is the identity when restricted to the leaf set.
- In the last example, we can permute the inner vertices and obtain a phylogenetic tree that is isomorphic to the given one.

set of partitions



- For every inner vertex, we can collect the labels on each branch, respectively, forming a partion of set N.
- Then we obtain a set of partitions from the given phylogenetic tree.
- For the given tree, what partitions for N can we obtain?

running example: set of partitions

- $P = \{p_a, p_b, p_c, p_d, p_e\}$
- $p_a = \{\{1\}, \{2\}, \{3\}, \{4, 5, 6, 7, 8, 9\}\}$
- $p_b = \{\{1, 2, 3\}, \{4\}, \{5\}, \{6, 7, 8, 9\}\}$
- $p_c = \{\{1, 2, 3, 4, 5\}, \{6, 7\}, \{8, 9\}\}$
- $p_d = \{\{1, 2, 3, 4, 5, 8, 9\}, \{6\}, \{7\}\}$
- $p_e = \{\{1, 2, 3, 4, 5, 6, 7\}, \{8\}, \{9\}\}$

phylogenetic set of partitions

A set of partitions of N is **phylogenetic** iff it fulfills the following axioms:

- Each partition has at least 3 parts.
- Any cardinality-one subset of N belongs to a unique partition.
- Any subset of N belongs to at most one partition.
- For any subset $A \subset N$ that belongs to some partition, its complement $N \setminus A$ also belongs to some partition.

- Is P phylogenetic?
- Given a phylogenetic set of partitions, how to convert it back to a tree?

From partitions to tree

- VERTICES: Each partition is a vertex, each single-element set $\{x\}$ in the partition contributes to a leaf x attached to the vertex.
- EDGES: Draw an edge between vertex v_1 and v_2 iff $I \in v_1$ and $N \setminus I \in v_2$.
- From P, we can also try this method, see if we obtain a pylogenetic tree?
- Let's try it on the whiteboard.

From partitions to tree

- $P = \{p_a, p_b, p_c, p_d, p_e\}$
- $p_a = \{\{1\}, \{2\}, \{3\}, \{4, 5, 6, 7, 8, 9\}\}$
- $p_b = \{\{1, 2, 3\}, \{4\}, \{5\}, \{6, 7, 8, 9\}\}$
- $p_c = \{\{1, 2, 3, 4, 5\}, \{6, 7\}, \{8, 9\}\}$
- $p_d = \{\{1, 2, 3, 4, 5, 8, 9\}, \{6\}, \{7\}\}$
- $p_e = \{\{1, 2, 3, 4, 5, 6, 7\}, \{8\}, \{9\}\}$

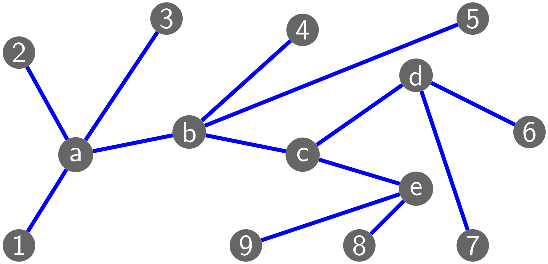
From partitions to tree

- The above mentioned two algorithms transfer between phylogenetic tree and phylogenetic set of partitions. Both compositions are the identity.
- These two representations are equivalent.

Five equivalent representations

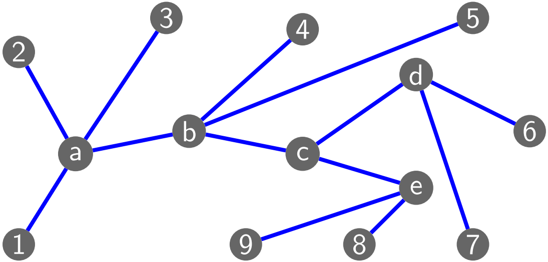
- phylogenetic tree
- phylogenetic set of partitions
- phylogenetic set of cuts
- phylogenetic crossing relations
- phylogenetic equivalences of triples

running example: set of cuts



- A **cut of** N is a partition of N into two subsets I, J such that cardinality of both I and J are bigger than one.
- Starting from a phylogenetic tree with leaf set N, for each inner edge, we can collect the labels on two sides of the edge respectively, forming a cut of N.
- Then we obtain a set of cuts.

running example: set of cuts



Let's see what we will obtain from this given tree!

running example: set of cuts

- C_{ab}: $\{\{1,2,3\},\{4,5,6,7,8,9\}\}$
- C_{bc}: $\{\{1,2,3,4,5\},\{6,7,8,9\}\}$
- C_{cd}: $\{\{1,2,3,4,5,8,9\},\{6,7\}\}$
- C_{ce}: $\{\{1,2,3,4,5,6,7\},\{8,9\}\}$
- We obtain the set $C = \{C_{ab}, C_{bc}, C_{cd}, C_{ce}\}$.
- We say that a set of cut C is **phylogenetic** iff for any two cuts $\{I_1, J_1\}$, $\{I_2, J_2\}$ in C, (at least) one of these four sets is empty: $I_1 \cap I_2$, $I_1 \cap J_2$, $J_1 \cap I_2$, $J_1 \cap J_2$.
- Is C in our running example phylogenetic?
- We have an algorithm converting from a phylogenetic set of cuts to a phylogenetic tree. Both compositions are the identity.
- These two representations are equivalent.

Five equivalent representations

- phylogenetic tree
- phylogenetic set of partitions
- phylogenetic set of cuts
- phylogenetic crossing relations
- phylogenetic equivalences of triples

Crossing relations

- A **crossing relation** is a set X of a pair of cardinality-two subsets of N. We denote as $(i, j \mid k, l)$ the element $\{\{i, j\}, \{k, l\}\}$ and we call it a **cross** of X.
- Starting from a set of cuts C, we can construct a crossing relation X_C as follows: $(i, j \mid k, l) \in X_C$ iff $i, j \in I$ and $k, l \in J$ for some cut $\{I, J\} \in C$.
- We say X is **phylogenetic** iff it fulfills the following axioms:

X1 If $(i, j \mid k, l) \in X$, $(i, k \mid j, l) \notin X$.

X2 If $(i, j \mid k, l), (i, j \mid k, m) \in X$ and $l \neq m$, then $(i, j \mid l, m) \in X$.

X3 If $(i, j \mid k, l) \in X$ and m is distinct from i, j, k, l, then $(i, j \mid k, m) \in X$ or $(i, m \mid k, l) \in X$. (Note that here "or" means at least one incident should happen.)

recall: phylogenetic set of cuts

- $C = \{C_{ab}, C_{bc}, C_{cd}, C_{ce}\}$.
- C_{ab}: $\{\{1, 2, 3\}, \{4, 5, 6, 7, 8, 9\}\}$
- C_{bc}: $\{\{1, 2, 3, 4, 5\}, \{6, 7, 8, 9\}\}$
- C_{cd}: $\{\{1, 2, 3, 4, 5, 8, 9\}, \{6, 7\}\}$
- C_{ce}: $\{\{1, 2, 3, 4, 5, 6, 7\}, \{8, 9\}\}$

running example: crossing relations

- Starting from the set of cuts C in our running example, we obtain a crossing relation X_C containing the following elements.
- $i, j \in \{1, 2, 3\}$ and $k, l \in \{4, 5, 6, 7, 8, 9\}$ (45 crosses);
- $i, j \in \{1, 2, 3, 4, 5\}$ and $k, l \in \{6, 7, 8, 9\}$ (60 crosses);
- $i, j \in \{1, 2, 3, 4, 5, 8, 9\}$ and $\{k, l\} = \{6, 7\}$ (21 crosses);
- $i, j \in \{1, 2, 3, 4, 5, 6, 7\}$ and $\{k, l\} = \{8, 9\}$ (21 crosses).
- We can check that X_C is a phylogenetic crossing relation.
- We also have an algorithm converting from phylogenetic crossing relation to a phylogenetic set of cuts. The two compositions are both the identity.
- These two representations are equivalent.

Five equivalent representations

- phylogenetic tree
- phylogenetic set of partitions
- phylogenetic set of cuts
- phylogenetic crossing relations
- phylogenetic equivalences of triples

phylogenetic equivalent classes of triples

- A **triple** in N is a cardinality-three subset of N. We denote the set of all triples in N by $\binom{N}{3}$.
- We say an equivalence relation $\sim$ on $\binom{N}{3}$ is **phylogenetic** if the following two axioms are fulfilled.

E1 For any subset $Q \subseteq N$ of cardinality 4, either all 4 triples in $\binom{Q}{3}$ are equivalent, or there are two equivalence classes, each containing two triples of $\binom{Q}{3}$.

E2 If i, j, k, l, m are pairwise distinct, and $\{i, j, k\} \sim \{i, j, l\}$, then $\{i, j, k\} \sim \{i, j, m\}$ or $\{i, k, l\} \sim \{m, k, l\}$.

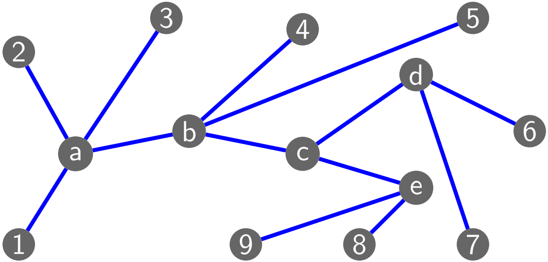
example: equivalences of triples

- Let $N = \{1, 2, 3, 4, 5\}$. We define an equivalence relation E with three distinct classes as follows:
- $\{1, 2, 3\} \sim \{1, 2, 4\} \sim \{1, 2, 5\}$
- $\{1, 4, 5\} \sim \{2, 4, 5\} \sim \{3, 4, 5\}$
- $\{1, 3, 4\} \sim \{1, 3, 5\} \sim \{2, 3, 4\} \sim \{2, 3, 5\}$
- We can check that it is phylogenetic.
- Starting from a phylogenetic equivalence relation E of triples in $\binom{N}{3}$, we can construct a crossing relation on N X_E as follows.
- Cross $(i, j \mid k, l) \in X_E$ iff $\{i, j, k\} \sim \{i, j, l\} \not\sim \{i, k, l\} \sim \{j, k, l\}$.
- Continue with the above example, we obtain a crossing relation containing the following elements:

from crossing relation to equivalences of triples

- $(1,2 \mid 3,4), (1,2 \mid 3,5), (1,2 \mid 4,5), (1,3 \mid 4,5), (2,3 \mid 4,5)$.
- We can check that it is phylogenetic.
- Starting from a phylogenetic crossing relation X in N, we can define equivalences of triples E_X on $\binom{N}{3}$ as follows.
- First, define a relation R_X on $\binom{N}{3}$ as follows: $(t_1, t_2) \in R_X$ iff $t_1 = t_2$ or $\mid t_1 \cap t_2 \mid = 2$ and – say $t_1 = \{i,j,k\}$, $t_2 = \{i,j,l\}$ – neither $ik \mid jl$ nor $il \mid jk$ is in X.
- E_X is defined to be the transitive closure of R_X.
- Starting from a phylogenetic crossing relation, via the above construction, we actually will obtain a phylogenetic equivalences of triples.
- Composition of the two transformations are both the identity.
- These two representations are equivalent.

more comments?



From set of partitions to equivalences of triples.

Five equivalent representations ?

- phylogenetic tree
- phylogenetic set of partitions
- phylogenetic set of cuts
- phylogenetic crossing relations
- phylogenetic equivalences of triples

meta-level overview

- Edges? Vertices?
- Macro level? Micro level?
- macro level + focusing on edges $\Longrightarrow$ set of cuts. (Each cut corresponds to an edge.)
- micro level + focusing on edges $\Longrightarrow$ crossing relation.
- macro level + focusing on vertices $\Longrightarrow$ set of partitions. Each partition corresponds to a vertex.
- micro level + focusing on vertices $\Longrightarrow$ equivalent classes of triples. Each class corresponds to a vertex.

Thank You