Five equivalent ways to represent a phylogenetic tree

Jiayue Qi

joint work with Herwig Hauser (university of Wien)

and Josef Schicho (University of Linz)
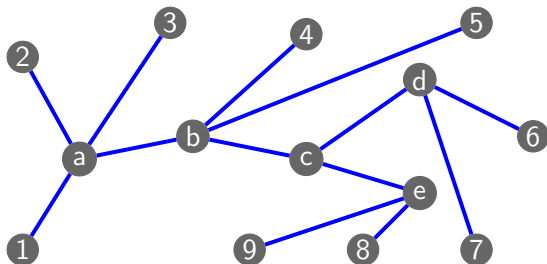
2021.11.12

Five equivalent representations

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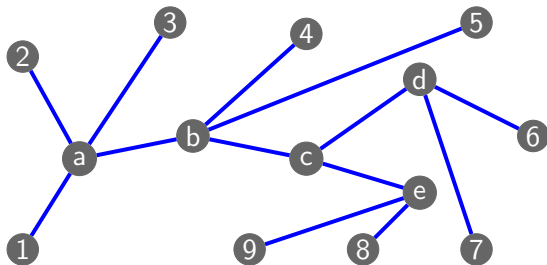
Example: phylogenetic tree

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 the path from this vertex to one of its incident edges, 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

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

running example: set of partitions

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

What kind of partition sets can be obtained from a phylogenetic tree?

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 $A \subset N$ with cardinality bigger than one 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 blackboard.

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 in between a phylogenetic tree and a phylogenetic set of partitions. Both compositions are the identities.
- 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
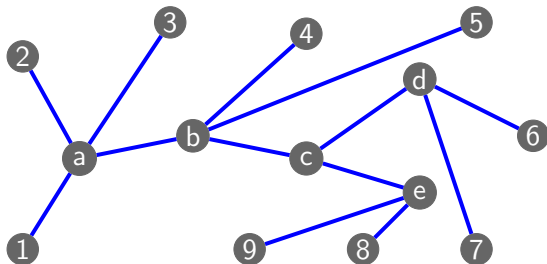
- A **cut of** N is a partition of N into two subsets I, J such that the cardinalities of both I and J are bigger than one; then I, J are called *parts* of this cut.
- Starting from a phylogenetic tree with leaf set N, removing any inner edge and collecting the labels on two connected components gives us a cut of N.
- Collect the cuts obtained in the above described way, we obtain a set of cuts.

running example: set of cuts
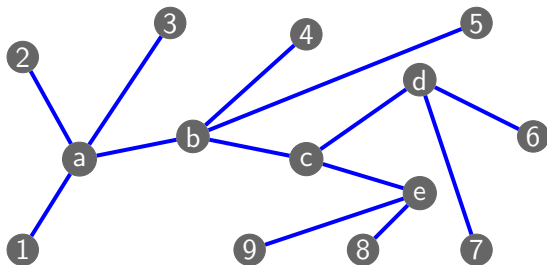
Let's see what we will obtain from this given tree!

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 (which we call *tree algorithm*) converting from a phylogenetic set of cuts to a phylogenetic tree. Both compositions are the identities.
- These two representations are equivalent.

tree algorithm: from cuts to tree

- Input: a phylogenetic set of cuts C
- Output: its corresponding phylogenetic tree
- Step 1: collect all parts of those cuts in C in set P.
- Step 2: pick any cut from set C, say $c = (I, J) \in C$.
- Step 3: go through all elements in P, find those that is either a subset of I or a subset of J, collect them together in set P_1.
- Step 4: create a Hasse diagram H of elements in P_1 w.r.t. set containment order.
- Step 5: consider H as a graph (V, E). Each element in P_1 has a corresponding vertex in H. We denote the vertex v_I for $I \in P_1$.

tree algorithm

- Step 6: For each vertex v of H, define the leaf set $h(v)$ attached to v as its corresponding element in P_1.

- Step 8: Go through the vertices again, update the leaf sets: $h(v) := h(v) \setminus h(v_1)$ if v_1 is less than v in H (in the Hasse diagram relation).

- Step 9: $E = E \cup \{v_I, v_J\}$. This edge corresponds to the cut we pick in Step 3.

Let us now apply this algorithm to our running example, on the blackboard!

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).
- One 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 identities.
- 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 3-element subset of N. We denote the set of triples in N by $\binom{N}{3}$. A set $S \subset \binom{N}{3}$ of triples is called **diverse** if it is non-empty and it fulfills the following two axioms:

> (D1) If $\{i, j, k\} \in S$, and $l \in N$, then S also contains one of the triples $\{i, j, l\}$, $\{i, k, l\}$, or $\{j, k, l\}$.

> (D2) Let $a, b, c, x, y, z \in N$. If S contains the triples $\{a, x, y\}$, $\{b, y, z\}$, and $\{c, x, z\}$, then it also contains $\{x, y, z\}$.

We say that an equivalence relation on $\binom{N}{3}$ is *phylogenetic* if and only if the following axiom is fulfilled:

> (E0) Each class of the equivalence relation is diverse.

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 set of partitions P_E on N as follows.
- For any set S of triples, let G_S be the graph with vertex set N, and an edge between $i, j \in N$ if and only if no triples of S contain both i and j. Let p_S be the partition of N defined by the connected components of G_S.
- P_E is defined as the collection of those p_S for each equivalence class S in E.
- What partition set do we get from this example above?

conversion: partitions and triples

- Starting from a phylogenetic set of partition P, for each partition p, let S_p be the set of triples that are separated by p.
- Collection of S_p for each $p \in P$ gives us the equivalence classes of triples.
- What equivalence class of triples do we get from the set of partitions above?
- Composition of the two transformations are both the identities.
- 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

meta-level overview

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

Reference

📄 Qi Jiayue, Schicho Josef.
Five Equivalent Ways to Describe Phylogenetic Trees. arXiv preprint arXiv:2011.11774

Thank You