

Visualizing reconciliations in co-phylogeny (Extended Abstract)*

Tiziana Calamoneri¹, Valentino Di Donato²,
Diego Mariottini², and Maurizio Patrignani²

¹ Computer Science Department, University of Rome “Sapienza”, Rome, Italy

² Engineering Department, Roma Tre University, Rome, Italy

1 Introduction

An urgent need in the biological research field consists of unambiguously and effectively representing co-phylogenetic trees, that are pairs of phylogenetic trees that have a mapping among their nodes. The typical application is the investigation of the co-evolution of families of organisms such as hosts and parasites.

A *phylogenetic tree* is a full binary rooted tree representing the inferred evolutionary relationships among relative species. In order to be meaningful from a biological point of view, a phylogenetic tree is required to be a full binary tree (i.e., each node has zero or two children). The biologists who study the co-evolution of hosts and parasites start with a host phylogenetic tree H , a parasite phylogenetic tree P , and a mapping function φ from the leaves of P to the leaves of H and represent them as a tanglegram (i.e. a pair of plane trees whose leaves are connected by straight-line edges [1,4]). Triple (H, P, φ) is called *co-phylogenetic tree* and only represents the input of a more complicated analysis process that aims at computing a mapping γ , called *reconciliation*, that extends φ and maps all the parasite nodes onto the host nodes with constraints. Intuitively, a reconciliation is a reasonable reconstruction of the co-evolution events that produced the co-phylogenetic tree [2,6].

Given a co-phylogenetic tree, a great number of different reconciliations are possible. Some of them can be easily discarded, since they are not consistent with time (i.e. they induce contradictory constraints on the periods of existence of the species associated to internal nodes). The remaining reconciliations are generally ranked based on some quality measure. Nevertheless, the number of plausible reconciliations is still so high and the difference between two equally-ranked reconciliations may be so big that biologists have to perform a painstaking manual inspection to select the reconciliations that are more compatible with the two phylogenetic trees and with their understanding of the evolution phenomena.

In this context, the contributions of this paper are the following:

* This research was partially supported by MIUR project “MODE – MORphing graph Drawings Efficiently”, prot. 20157EFM5C_001 and by *Sapienza* University of Rome projects “Graph Algorithms for Phylogeny: a promising approach” and “Combinatorial structures and algorithms for problems in co-phylogeny”.

1. We propose a new and unambiguous metaphor to represent reconciliations of co-phylogenetic trees. The main idea is that of representing H in a suitable space-filling style and of using a traditional node-link style to represent P .
2. In order to pursue readability we study the number of crossings that are introduced in the drawing of tree P in order to respect the represented reconciliation; on the one hand, we characterize some “planar instances”, i.e., triples $\langle H, P, \varphi \rangle$, in which every reconciliation can be represented without crossings; on the other hand, we show that, in the general case and for all kinds of representations that comply with a restricted set of aesthetic criteria, reducing the number of crossings in the representation of the reconciliations is NP-hard.

2 A new model for the visualization of reconciliations

While a decade-old literature is devoted to the visualization of co-phylogenetic trees as tanglegrams (see, for example, [1,4]), little can be said about the representation of the reconciliations of co-phylogenetic trees, except some visualization strategies adopted by the available tools for their computation.

In particular, few strategies can be highlighted.

The simplest one, schematically represented in Fig. 1(a), is that of representing the two trees by adopting the traditional node-link metaphor, where the nodes of P are drawn close to the nodes of H they are associated to. The advantage of this strategy lays in its simplicity. However, when several parasite nodes are associated to the same host node, the drawing becomes cluttered and the attribution of parasite nodes to host nodes becomes unclear. Further, even if tree P was drawn without crossings (tree H always is), the overlapping of the two trees produces a high number of crossings.

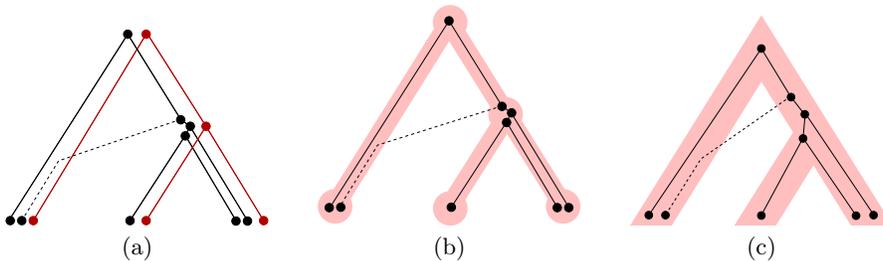


Fig. 1: Three visualization strategies for representing co-phylogenetic trees.

An alternative strategy (Fig. 1(b)) consists in representing H as a background shape, such that its nodes are light-colored circles and its edges are thick pipes, while P is contained into H and drawn in the traditional node-link style. This strategy is particularly effective, as it is unambiguous and crossings between the

two trees are strongly reduced, but it is still cluttered when a parasite subtree has to be squeezed inside the reduced area of a host node.

To solve this cluttering problem some visualization tools adopt the strategy of keeping the containment metaphor while only drawing thick edges of H and omitting host nodes (Fig. 1(c)). This produces a node-link drawing of the parasite tree drawn inside the pipes representing the host tree. Also this strategy is sometimes ambiguous, since it is unclear how to attribute parasite nodes to host nodes.

With the aim of overcoming the limitations of existing visualization strategies, we propose a new metaphor for the representation of reconciliations. Our strategy is complementary to the third strategy just described: we omit altogether the arcs of H and represent its nodes as regions of the plane. Arcs are implicitly represented by the contact of the regions corresponding to the nodes. Tree P , instead, maintains the traditional node-link representation, where parasites are placed into the regions associated with the hosts they are mapped to. This representation is not ambiguous, since the containment relationship between host and parasites unambiguously represents the mapping γ .

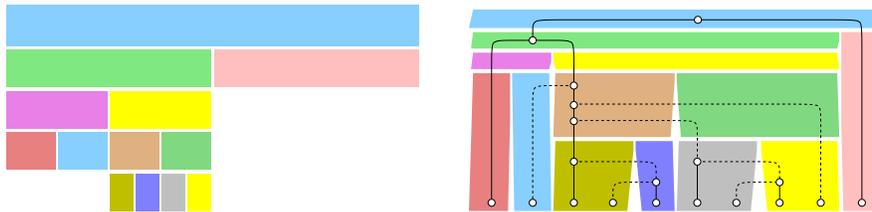


Fig. 2: (a) An icicle. (b) The representation adopted for trees H and P , where dotted lines represent host-transfers.

Indeed, the representation of tree H is a variant of a representation known in the literature with the name of *icicle* [5]. An *icicle* (Fig. 2(a)) is a space-filling representation of hierarchical information in which nodes are represented by rectangles and arcs are represented by the contact of rectangles, such that the bottom side of the rectangle of a node touches the top sides of its children.

With respect to the standard icicle representation, we draw H with the following variants (Fig. 2(b)):

- we impose that the bottom sides of the rectangles representing the leaves of H touch the bottom border of the drawing area. This forces all contemporaneous hosts to share the same bottom line that represents current time,
- for each generation of the host tree, we modify the height of the rectangles in such a way that all the parasite subtrees are contained into their hosts rectangles,

- we specialize the icicle to the representation of binary tree, rounding the corners of the rectangles in order to ease the visual recognition of the two children of each node.

3 Planar instances and reconciliations with the minimum number of crossings

In this section we characterize the reconciliations that can be planarily drawn, showing that there exist some “planar instances” not depending on the represented reconciliation. Then we naturally wonder what happens when the given instance is not planar and show an NP-hardness result. We omit the proof for the sake of brevity.

Let be given a co-phylogenetic tree (H, P, φ) and a reconciliation γ .

Definition 1. *A drawing $\mathcal{D}(\gamma)$ of γ is a visual representation of γ according to any metaphor such that:*

- the leaves of H are drawn along a horizontal line (i.e. they have the same y -coordinate, w.l.o.g. let it be equal to 0),
- tree H is embedded (i.e. drawn in a plane way) in the half-plane with non-negative y -coordinates,
- each node $p \in P$ is drawn inside the representation of node $h \in H$ such that $\gamma(p) = h$,
- each edge (s, t) of P (except the ones performing a host-transfer) is drawn inside the representation of the nodes and the edges of H belonging to path from $\gamma(s)$ and $\gamma(t)$,
- the sub-forest induced by all nodes of P mapped on the same node of H follows a strictly downward convention.

We call $V(H)$, $A(H)$, $V(P)$ and $A(P)$ the set of nodes and of arcs of H and P , respectively. $V_L(H)$ and $V_L(P)$ are the set of leaves of the two trees.

Let $G(H, P, \varphi)$ be the graph whose arc set is $A(H) \cup A(P)$ and node set is $V(H) \cup (V(P) \setminus V_L(P))$ and every leaf l of P is identified with leaf $\varphi(l)$ of H . $G(H, P, \varphi)$ is cyclic and, in general, not planar. It can be represented on the plane so that:

- all the leaves of H (and hence also those of P) lie on a horizontal line l ,
- internal nodes and edges of H are embedded (i.e. drawn in a plane way) in the half-plane above l in a strictly downward fashion,
- internal nodes and edges of P are represented (not necessarily without crossings) in the half-plane under l in a strictly upward fashion.

In the following, we will speak about *drawing of $G(H, P, \varphi)$* without other specifications to mean a representation on the plane satisfying the above constraints.

Theorem 1. *The following claims are equivalent:*

1. $G(H, P, \varphi)$ is planar and a plane drawing of it is known,

2. for each time-consistent reconciliation γ , there exists a plane drawing $\mathcal{D}(\gamma)$.

Theorem 1 states that we will have a plane drawing of a reconciliation γ if and only if graph $G(H, P, \varphi)$ is planar. It is hence natural to wonder what happens otherwise. We focus on the problem of the crossing minimization in the drawing of a given reconciliation and prove that computing a drawing of a reconciliation with the minimum number of crossings is NP-hard.

Given γ and a constant k , we consider the problem RECONCILIATION LAYOUT (RL) as the problem asking whether there exist a permutation of the leaves of H and a mapping of the nodes of H and of P on the plane such the drawing of γ has at most k crossings.

Theorem 2. *Problem RL is NP-complete.*

We prove it by reducing to RL the NP-complete problem Two-Trees Crossing Minimization (TTCM), whose input consists of two binary trees T_1 and T_2 , whose leaf sets are in one-to-one correspondence, and a constant k ; the question is whether T_1 and T_2 admit a tanglegram drawing with at most k crossings among the tangles [3].

We observe that, in the reduction, a primary role is played by host-transfer edges, and this is reasonable, since they connect nodes in completely different parts of the tree, possibly very far. Nevertheless, host-transfers are very special edges and we believe that biologists have no problems in understanding the drawing even if these edges introduce some crossings, so we wonder whether crossings involve only host-transfers or not. We answer in a negative way to this question by showing that some crossings can occur, even if host-transfer edges are eliminated from the representation.

References

1. Buchin, K., Buchin, M., Byrka, J., Nöllenburg, M., Okamoto, Y., Silveira, R.I., Wolff, A.: Drawing (complete) binary tanglegrams. *Algorithmica* 62(1-2), 309–332 (2012), <http://dx.doi.org/10.1007/s00453-010-9456-3>
2. Charleston, M.A.: Jungles: a new solution to the host/parasite phylogeny reconciliation problem. *Math Biosci* 149(2), 191–223 (1998)
3. Fernau, H., Kaufmann, M., Poths, M.: Comparing trees via crossing minimization. *J. Comput. Syst. Sci.* 76(7), 593–608 (2010)
4. Fernau, H., Kaufmann, M., Poths, M.: Comparing trees via crossing minimization. *J. Comput. Syst. Sci.* 76(7), 593–608 (2010)
5. Kruskal, J.B., Landwehr, J.M.: Icicle plots: Better displays for hierarchical clustering. *The American Statistician* 37(2), 162–168 (1983), <http://www.jstor.org/stable/2685881>
6. Page, R.D., Charleston, M.A.: Trees within trees: phylogeny and historical associations. *Trends Ecol Evol* 13(9), 356–359 (1998)