



OTHER KINDS OF NETWORKS: PROBLEMS IN (CO-)PHYLOGENY & MEDICINE

(+SOME ONGOING PROBLEMS)

1

Prof. Tiziana Calamoneri
Network Algorithms
A.y. 2024/25

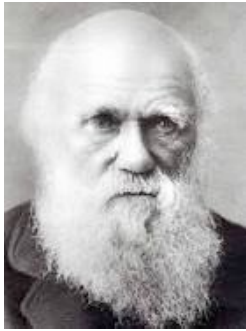


TREE RECONCILIATION

2

Prof. Tiziana Calamoneri
Network Algorithms
A.y. 2024/25

COEVOLUTION



"I can understand how a flower and a bee might slowly become, either simultaneously or one after the other, modified and adapted to each other in the most perfect manner, by the continued preservation of all the individuals which presented slight deviations of structure mutually favourable to each other."

Charles Darwin, *The Origin of Species*

3

MOTIVATION

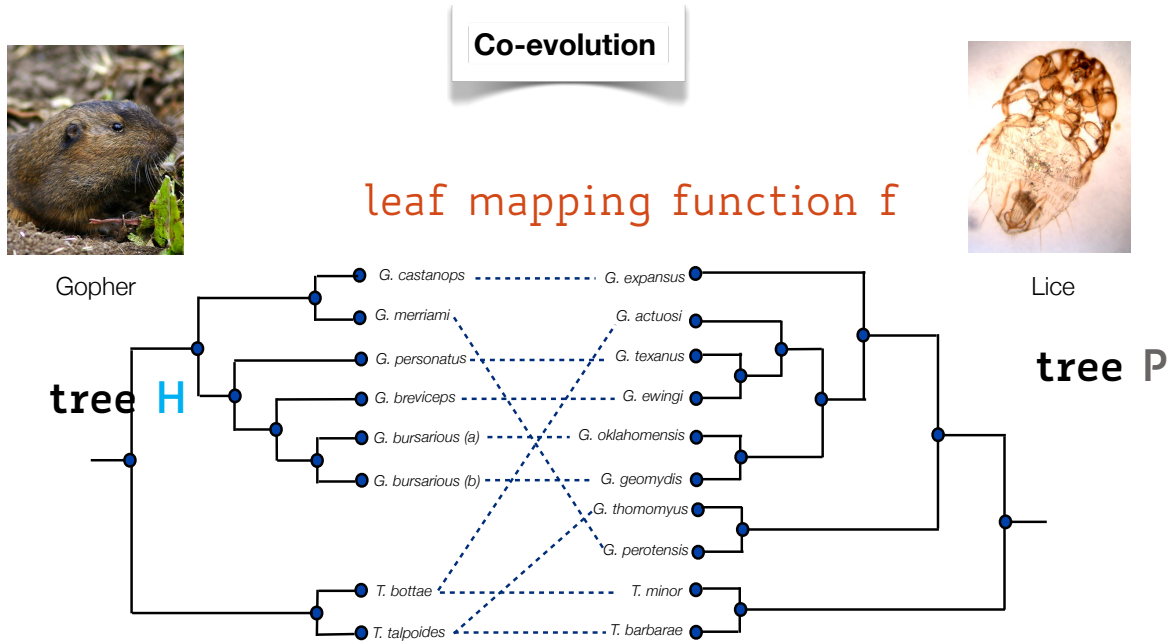
Different systems "coevolve"

- hosts and their parasites or pathogens
- whole organisms and their genes
- geographical areas and the species which inhabit them
- cultural traditions and populations

↑
same
model

4

THE INGREDIENTS

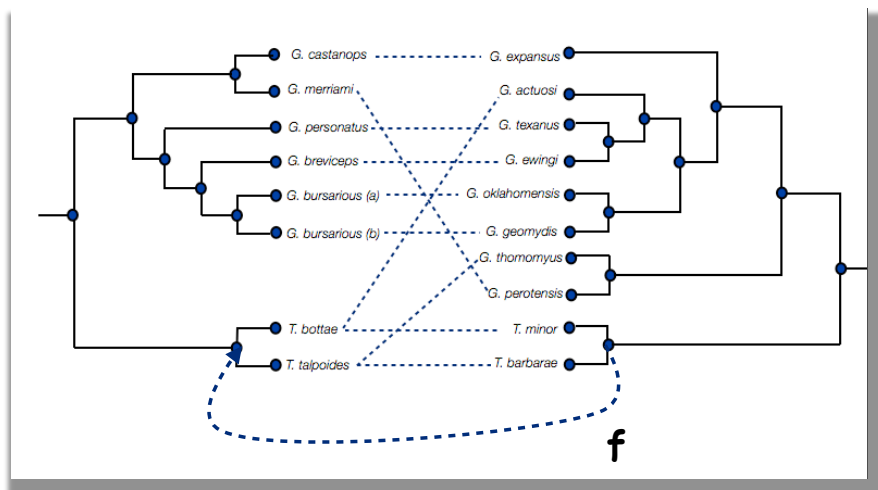


5

*adapted from Penn, Dustin J (Apr 2001) Coevolution: Host-Parasite. In: eLS. John Wiley & Sons Ltd, Chichester. <http://www.els.net>

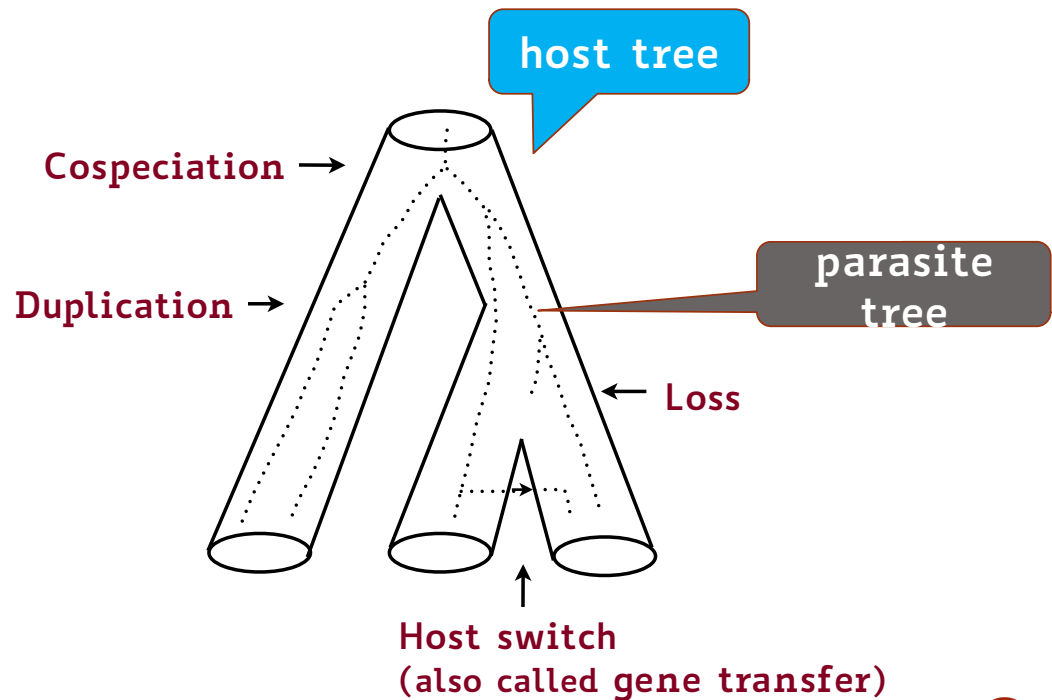
RECONCILIATIONS (1)

Very informally, a **reconciliation** is a mapping from the nodes of the parasite tree P to the nodes of the host tree H such that the leaf mapping function f is respected.



6

RECONCILIATIONS (2)



7

CO-PHYLOGENY RECONSTRUCTION PROBLEM (1)

- Determine reconciliations, given H , P and f
- Optimality of the solution: assigns a cost to each of the four types of events and then minimizes the total cost (Parsimony principle).

8

OUR PROBLEMS (1)

Biologists want to see all possible reconciliations, in order to understand which ones are biologically feasible and which ones are not.



1. **Enumerating all the optimal reconciliations**

OUR PROBLEMS (2)

- The number of optimal reconciliations increases rapidly even for small trees (exponential in the size of the trees).
- The size of the trees can be large.



2. **Reduce the cardinality of the set of optimal reconciliations**
3. **Visualize a given reconciliation in a "nice and clear" way**

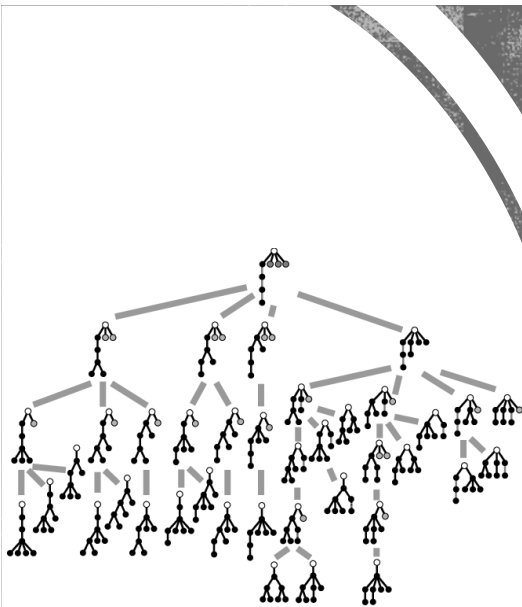
OUR PROBLEMS (3)

Modern methods of tree reconstruction may produce unrooted trees.



4. Exploit reconciliations to transform an unrooted tree into a rooted one

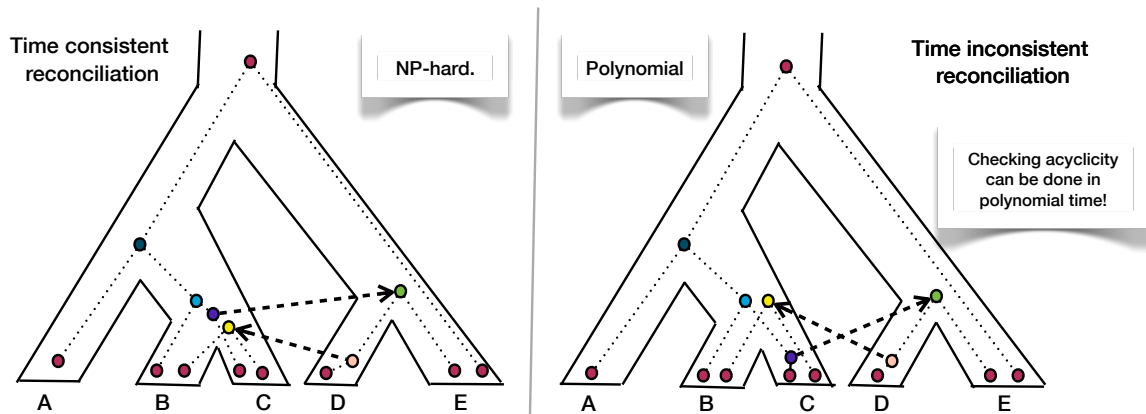
11



12

1.
ENUMERATING
ALL THE OPTIMAL
RECONCILIATIONS

TIME CONSISTENCY



Time consistency

13

ENUMERATION ALGORITHMS (1)

If the time-consistency constraint is dropped, the problem can be solved efficiently in polynomial time using a dynamic programming algorithm.

Several algorithms and tools, all based on the same approach...

14

ENUMERATION ALGORITHMS (2)

General idea: each internal vertex of P contributes separately to the total cost by means of its associated events, which depend only on the mappings of itself and of the two children



Dynamic programming technique

15

ENUMERATION ALGORITHMS (3)

Computing a single optimal reconciliation:

- Size of the dynamic programming matrix: $O(|P| |H|)$;
- each cell labeled by a parasite/host association ($p : h$) contains the information needed to construct all min reconciliations between the subtree of P rooted at p and H , such that p is mapped to h , i.e., a list of pairs of pointers to previously-filled cells of the matrix, representing the combinations of the mappings of the two children of p .
Size of a cell: $O(|H|)$.
- Once the matrix has been filled, the optimal solution is in correspondence of the mapping of the root of P .

16

ENUMERATION ALGORITHMS (4)

Enumerating all optimal reconciliations:

- Using the pointers, traverse in a depth-first search fashion the matrix and get all the optimal solutions.
- This is a polynomial delay enumeration algorithm:
it takes $O(|H|^2 |P|)$ time to fill the matrix the first time and then only $O(|P|)$ time to output each subsequent optimal reconciliation.

17

OPEN PROBLEMS (1)

PROBLEM 1. Consider more realistic models:

- deal with errors in phylogenetic trees (the phylogenetic trees are assumed to be correct, which may be not the case...) [Urbini, Sinimeri, Matias, Sagot '19]
- ϕ is not a function: multiple hosts – multiple parasites (a single parasite can infect more than one host...)

18

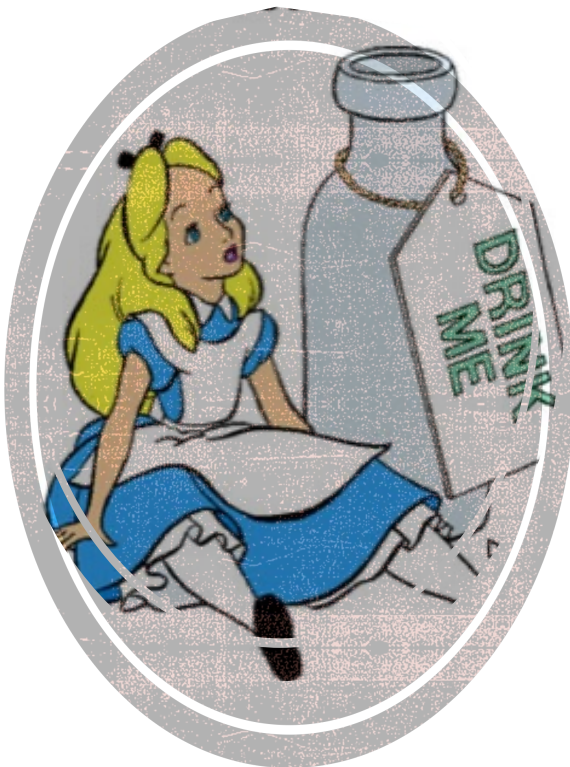
OPEN PROBLEMS (2)

PROBLEM 2. Consider special cases:

- Is it possible to compute an optimal reconciliation where the distance between the extremes of the host-switches is bounded by k in polynomial time? [C., Tavernelli, Vocca '19]
- Is it possible to compute in polynomial time an optimal time-consistent reconciliation for some particular topologies of trees?

[C., Monti, Sinimeri '19]

19



2. REDUCING THE CARDINALITY OF THE SET OF OPTIMAL RECONCILIATIONS

"SIMILAR"

RECONCILIATIONS (1)

Two main approaches:

1. Based on the definition of a similarity measure defined on the set of optimal reconciliations.

Idea: find a subset S of reconciliations that are representative of the whole set, i.e., such that each of the optimal reconciliations is at distance at most d from at least one of the reconciliations in S .

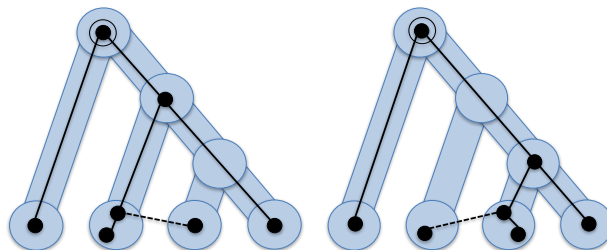
21

"SIMILAR"

RECONCILIATIONS (2)

Example:

- similarity measure as the smallest number of operations needed to change one reconciliation into another



CON: very similar reconciliations may need a large number of operations to be changed one into the other!

22

EQUIVALENT RECONCILIATIONS (1)

Two main approaches (cntd)

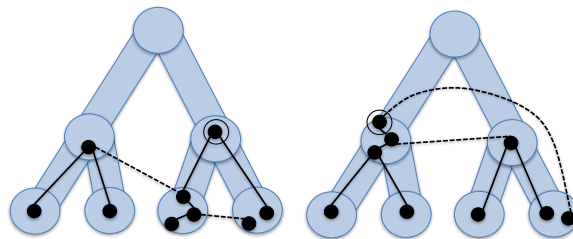
2. Based on the definition of equivalence classes to group the reconciliations that may be considered biologically equivalent and output a single solution for each class.

23

EQUIVALENT RECONCILIATIONS (2)

Example:

- two reconciliations are equivalent if they have the same event vector



PRO: while the number of optimal solutions can be exponential, the number of event vectors is polynomial

CON: reconciliations with the same event vector may be even very different!

24

EQUIVALENT RECONCILIATIONS (3)

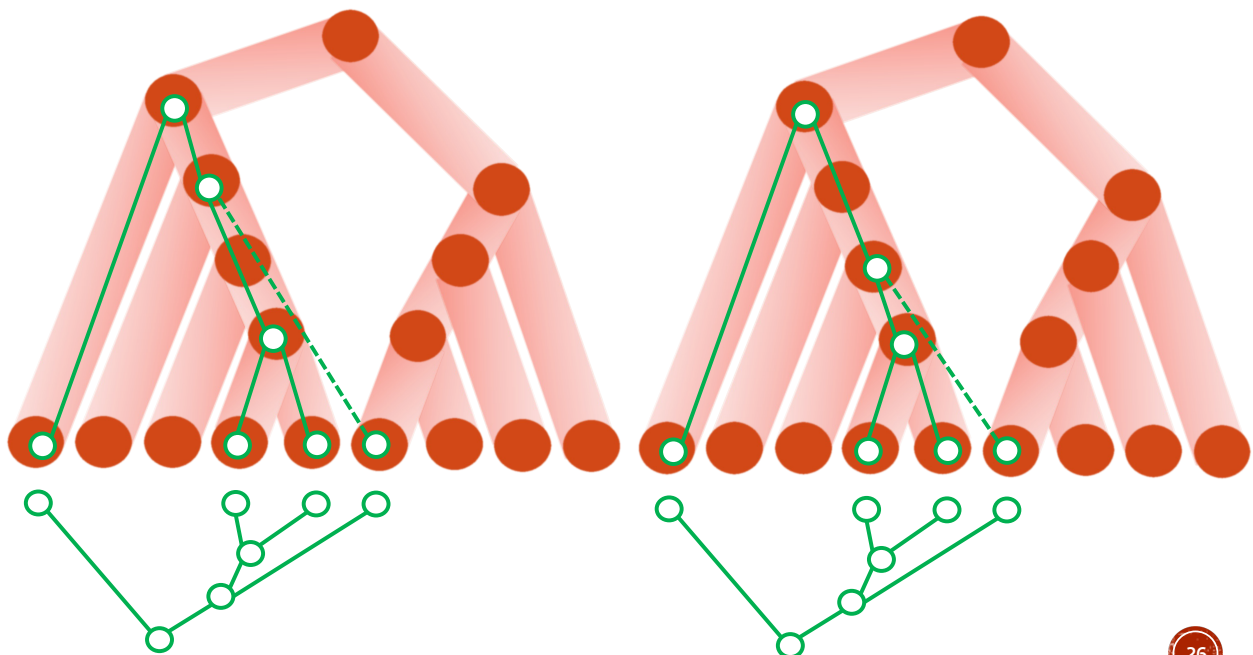
Idea inspired by the following theoretical result:
once the set of vertices of P that are associated to host-switches is fixed, an optimal reconciliation can be easily identified using the least common ancestor mapping



R' and R'' are identical iff they have the same host-switches [Gastaldello, C., Sagot '17]

25

EQUIVALENT RECONCILIATIONS (4)



26

EQUIVALENT RECONCILIATIONS (5)

CON: This method requires first the enumeration of all the optimal solutions and then to cluster them according to the equivalence notion. When the number of reconciliations is too large, for example, $> 10^{42}$ (Wolbachia dataset), listing all the solutions is not feasible.

Question: is it possible to enumerate only one representative for equivalence class without considering all the elements?

Done for three (artificial) notions of equivalence

[Wang, Mary, Sagot, Sinaimeri '21].

27

OPEN PROBLEMS

- **PROBLEM 3.** Explore the connections between the equivalences defined in this paper and determine whether there exist polynomial delay algorithms enumerating the representative reconciliations of the (biological meaningful) equivalence classes defined in [Gastaldello, C., Sagot '17].
- **PROBLEM 4.** Define other (biologically meaningful) notions of equivalence for which such algorithms exist.

28



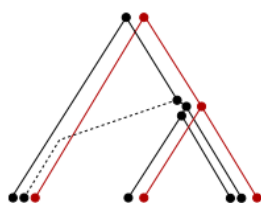
VISUALIZING RECONCILIATIONS

29

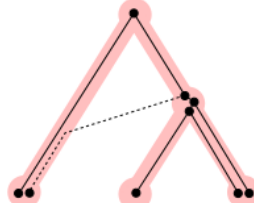
VISUALIZING RECONCILIATIONS (1)

[C., Di Donato, Mariottini, Patrignani '20]

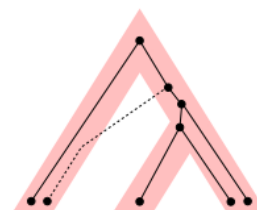
- Given H , P , ϕ and a reconciliation R , we have to draw H and P (on H) to highlight ϕ and R in a nice and clear way
- Three main strategies:
 1. representing two paired trees
 2. parasites are drawn inside their hosts
 3. host tree is made of pipes and parasites are drawn into the pipes



Strategy 1



Strategy 2



Strategy 3

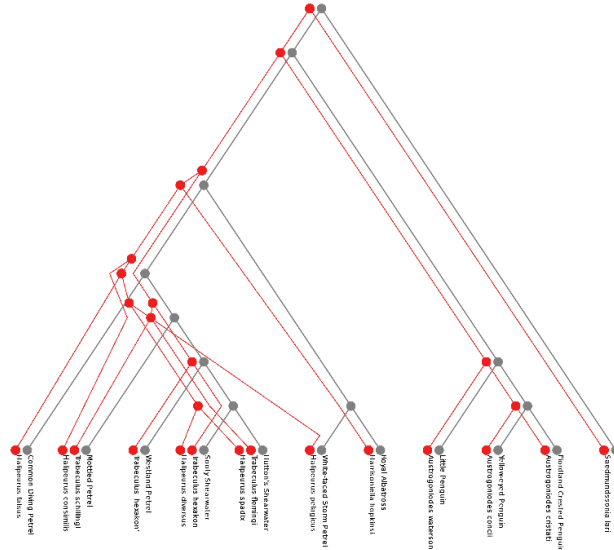
30

VISUALIZING RECONCILIATIONS (2)

Example of the 1st strategy:

- CoRe-PA

[Wieseke, Hartmann, Bernt, Middendorff '15]



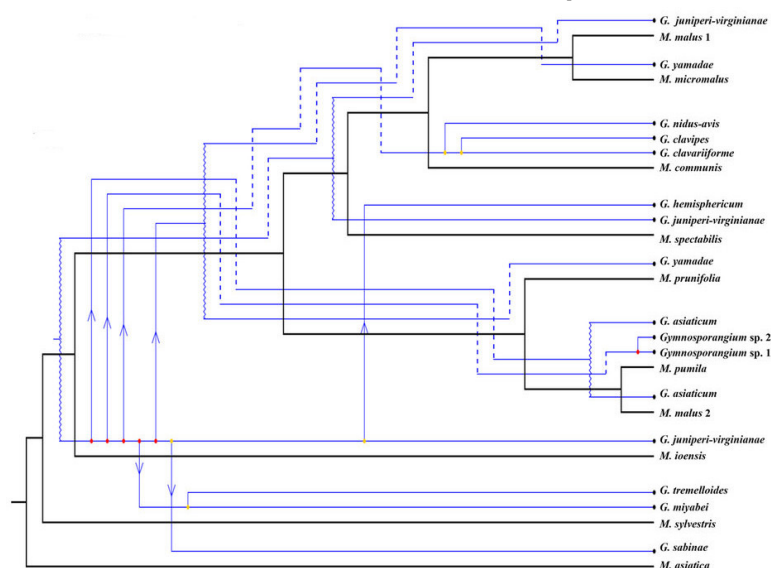
31

VISUALIZING RECONCILIATIONS (3)

Example of the 1st strategy:

- Jane 4

[Conow, Fielder, Ovadia, Libeskind-Hadas '10]



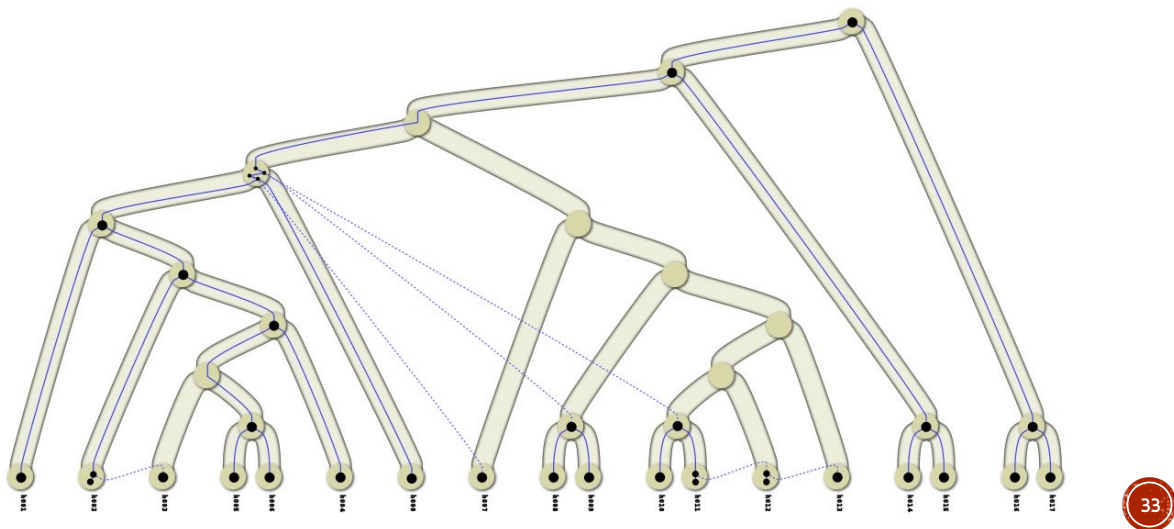
32

VISUALIZING RECONCILIATIONS (4)

Example of the 2nd strategy:

- CophyTrees

[Donati, Baudet, Sinaimeri, Crescenzi, Sagot '15]

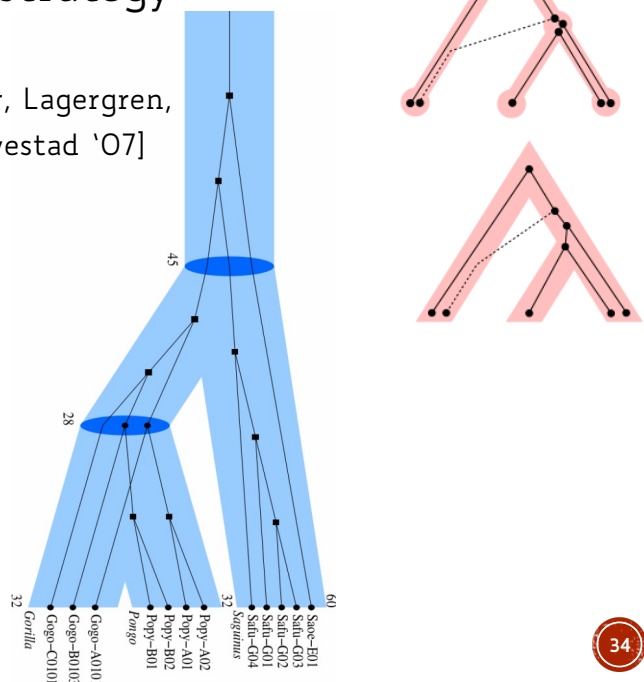


VISUALIZING RECONCILIATIONS (5)

Example of the 2nd/3rd strategy:

- Primetv

[Sennblad, Schreil, Sonnhhammer, Lagergren,
Arvestad '07]

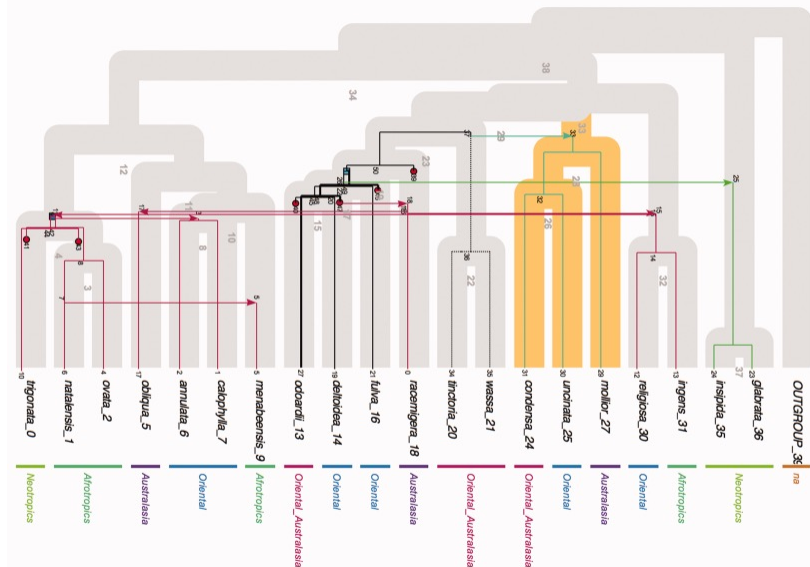
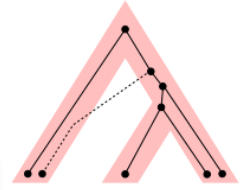


VISUALIZING RECONCILIATIONS (6)

Example of the 3rd strategy:

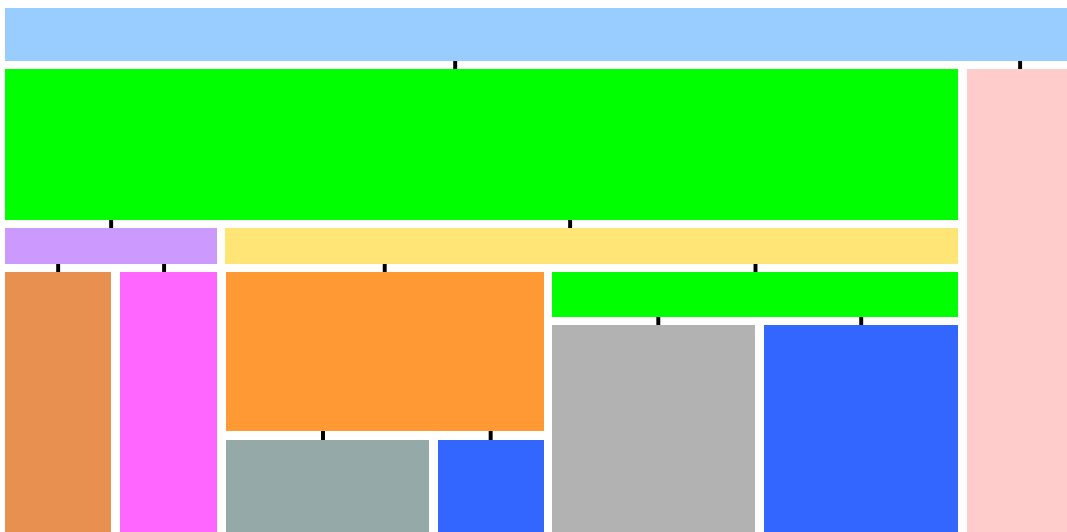
• SylvX

[Chevenet, Doyon, Scornavacca, Jacox, Jousset, Berry '16]



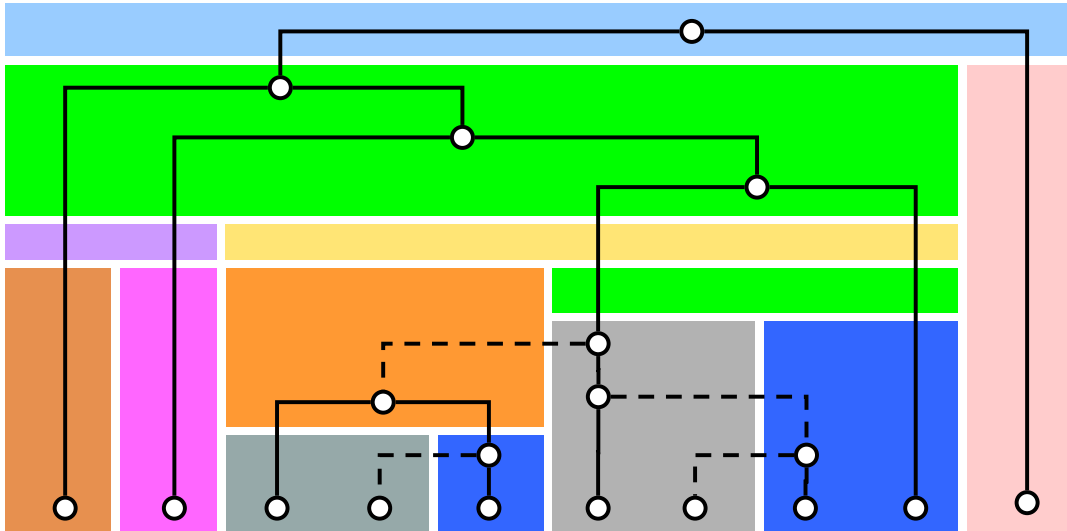
35

NEW METAPHOR: HP-DRAWINGS (1)



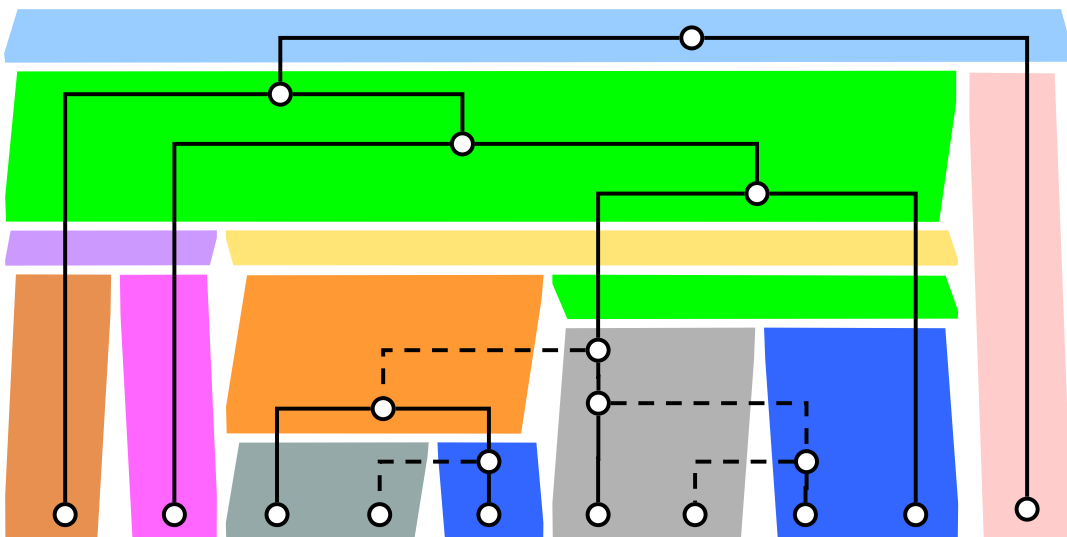
36

NEW METAPHOR: HP-DRAWINGS (2)



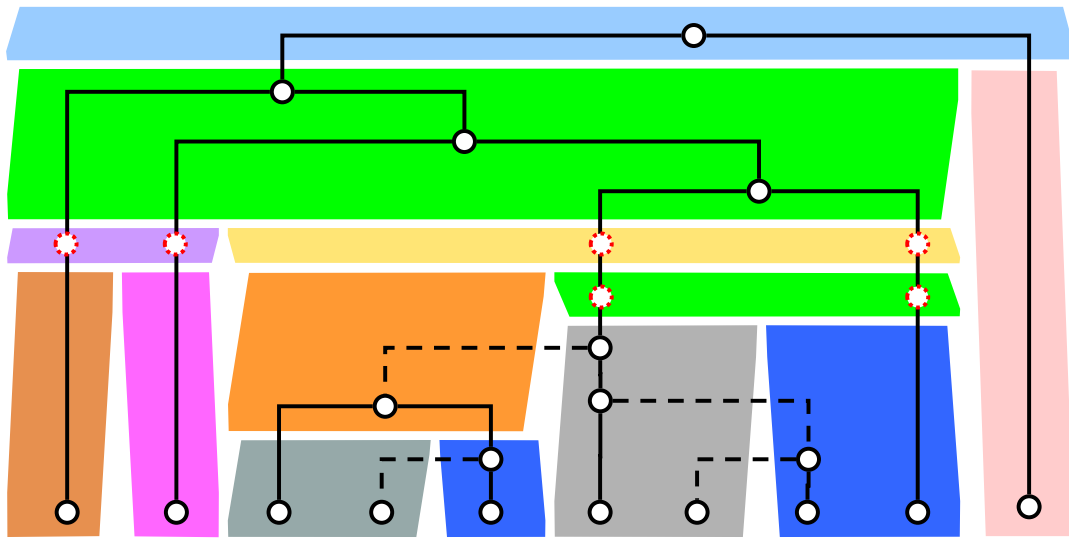
37

NEW METAPHOR: HP-DRAWINGS (3)



38

EVOLUTIONARY PHENOMENA (1)

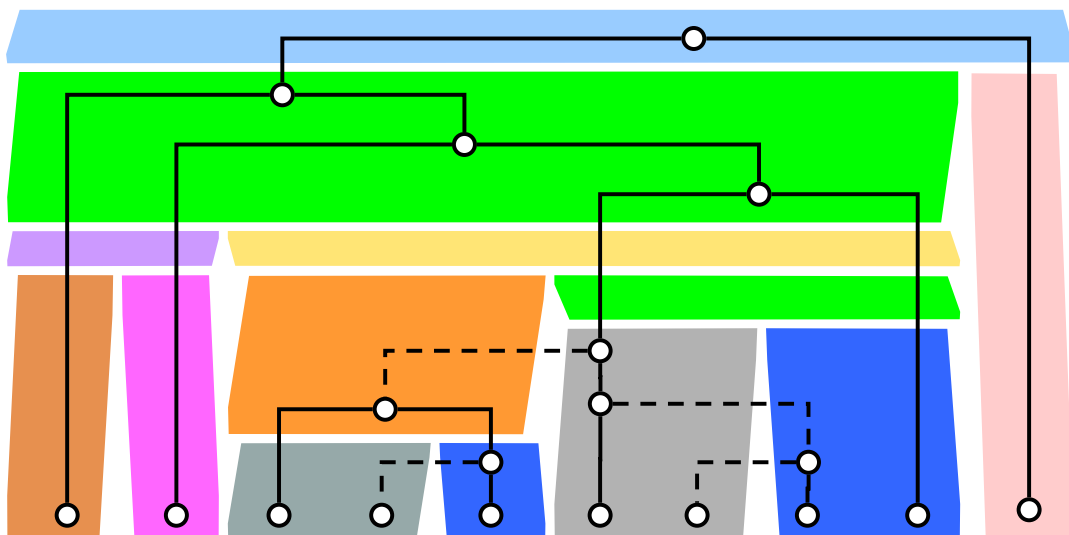


Loss:

- a parasite is transmitted to one child but not to the other child

39

EVOLUTIONARY PHENOMENA (2)

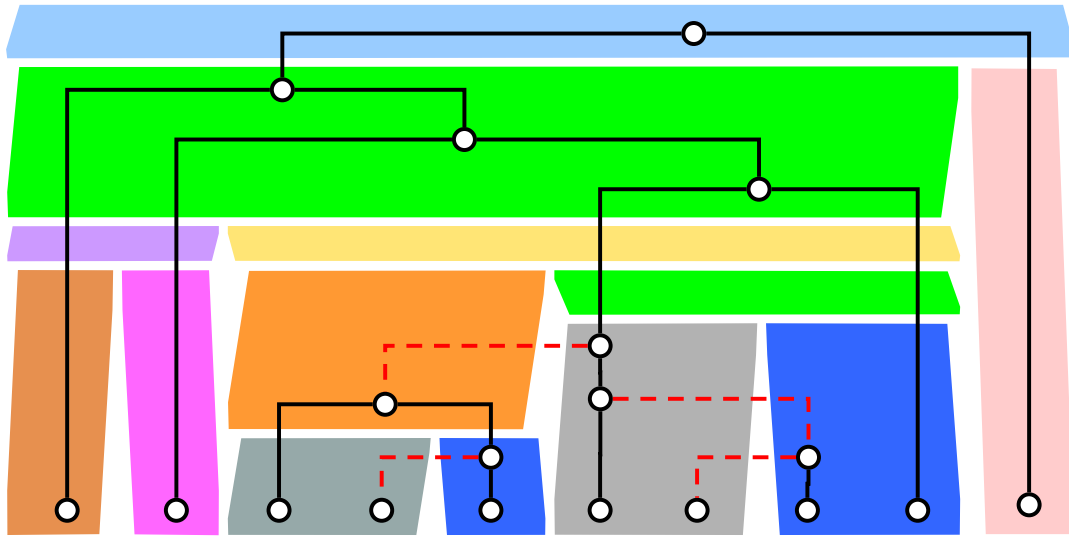


Duplication:

- both the children of a node go down in the same "direction"

40

EVOLUTIONARY PHENOMENA (3)



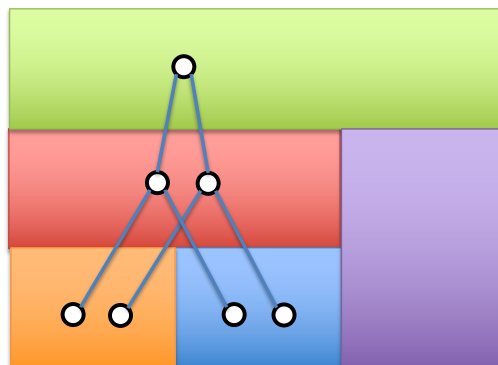
○ Host switch:

- a parasite is transmitted to a host that is not a descendant of the current one

41

OPTIMIZATIONS (1)

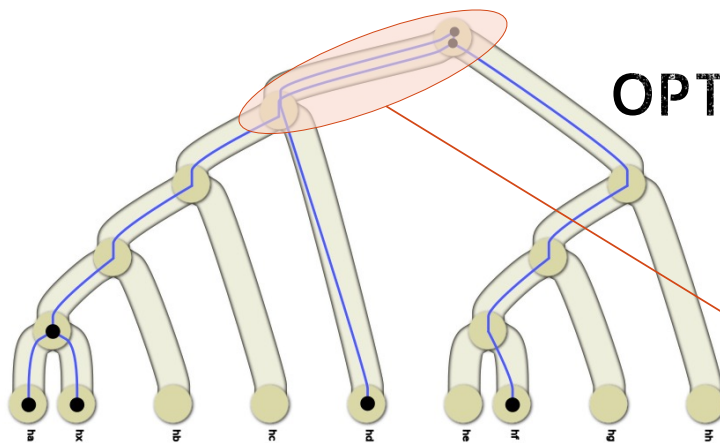
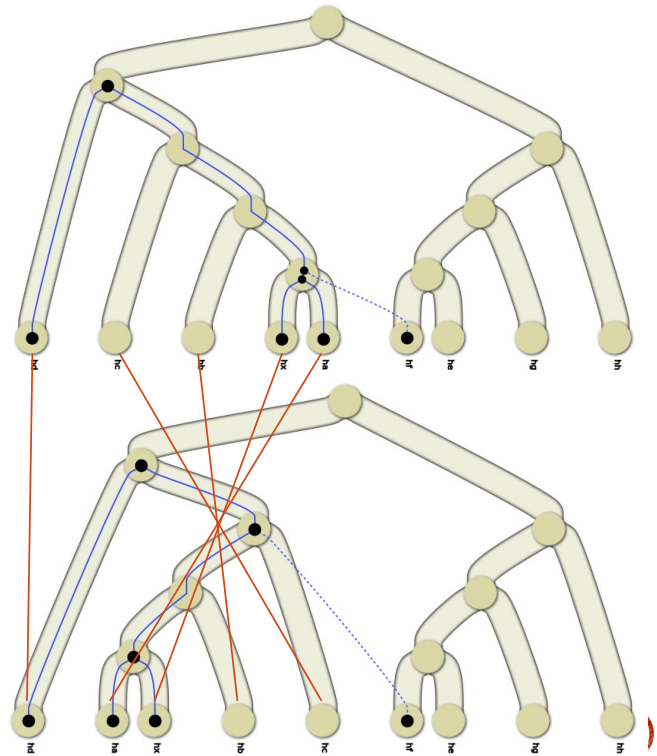
- Given H , P , f and a reconciliation R , we would like to:
 1. minimize the crossing number (not always possible to avoid crossings)



42

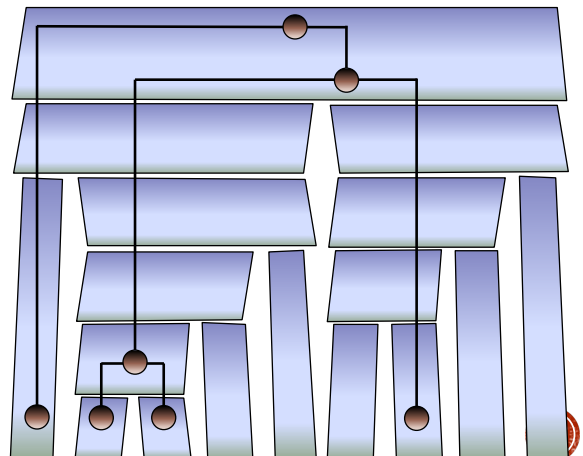
OPTIMIZATIONS (2)

- Given H , P , f and a reconciliation R , we would like to:
 - keep the **mental map** passing from a reconciliation to another one (not so in CophyTrees)

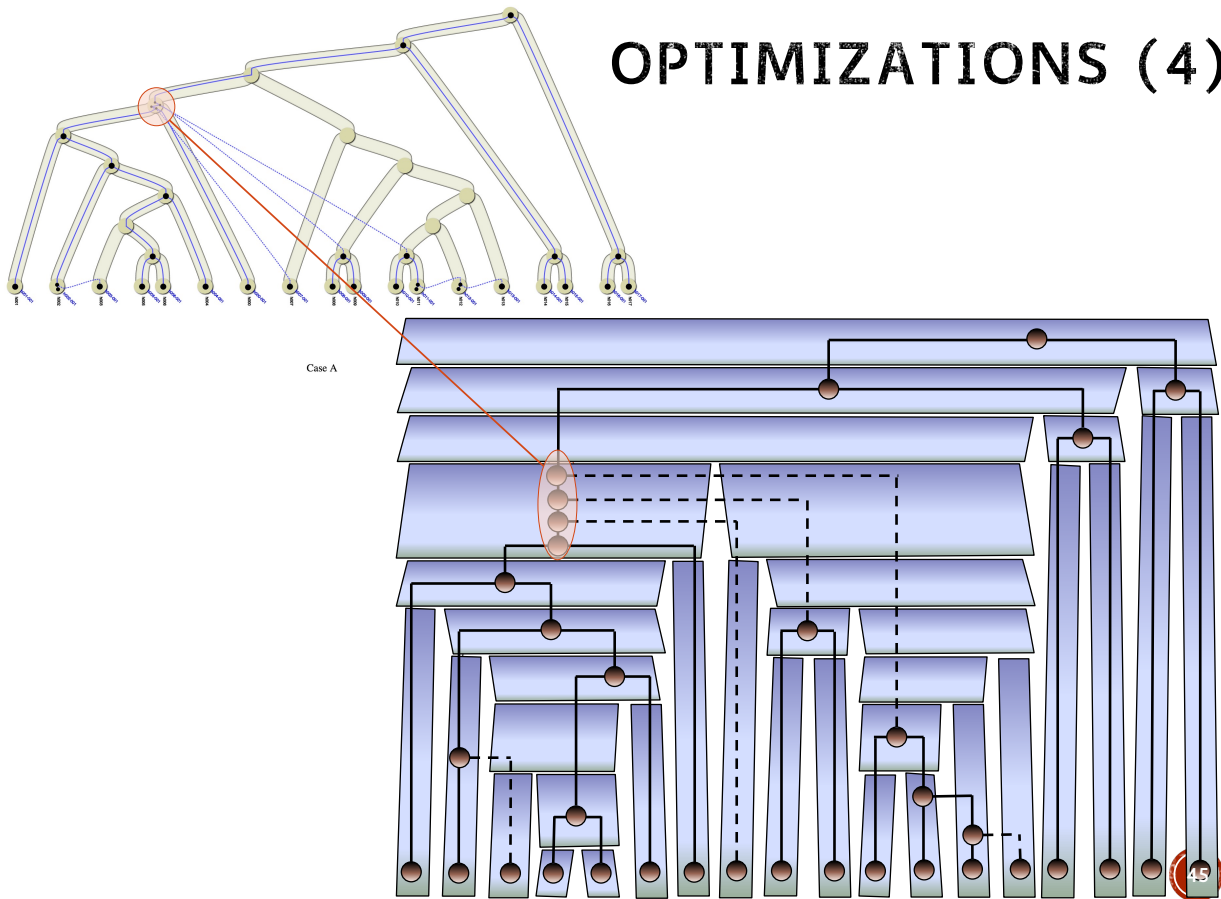


OPTIMIZATIONS (3)

Note: Our model makes easier to understand what happens and keep the mental map, while trying to minimize the crossing number.



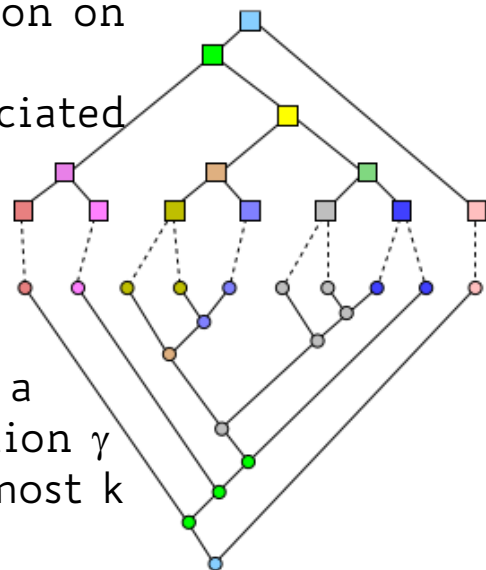
OPTIMIZATIONS (4)



PLANAR INSTANCES

Given H , P , f , construct the associated tanglegram:

- **Theorem:** Every reconciliation on H , P , f admits a planar representation iff the associated tanglegram is planar.
- So, we speak about **planar** and **not planar instances**.
- **Theorem:** deciding whether a time-consistent reconciliation γ admits a drawing with at most k crossings is NP-complete.



OPEN PROBLEMS

PROBLEM 5.

- Test the tool:
 - Do biologists like this metaphor?
- Models with more information:
 - handle additional information (e.g. geography)
[Berry, Chevenet, Doyon, Jousset '18] using colors?

47

A NEW GRAPH BASED REPRESENTATION FOR MEDICAL DATA



THE CONTEXT

[Belluomo, C., Paesani, Salvo '24]

- It is a fact that, when curing cancer, less than half of the patients positively react to drugs.
- Two patients apparently very similar (from the disease point of view) react in opposite ways to the same drugs.
- we have the genomic profiling (in terms of mutations of genes) derived from NGS (Next Generation Sequencing) of many patients

49

THE IDEA (1)

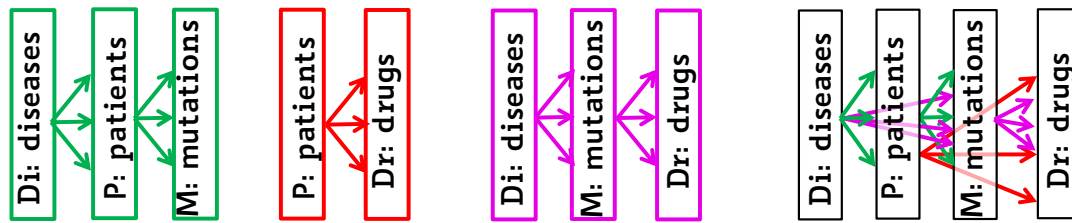
Works involving graphs for medical issues exploit available data bases and extract from them information modeled as graphs; the same can be done with a clever query.

Our idea is:

- to model several kinds of information (genetic information, general medical knowledge, medical records) in the same graph
- to exploit graph algorithms to infer some information that would not be available looking at only one of the data sets.

50

THE IDEA (2)



Green graph **G**: Genetic information

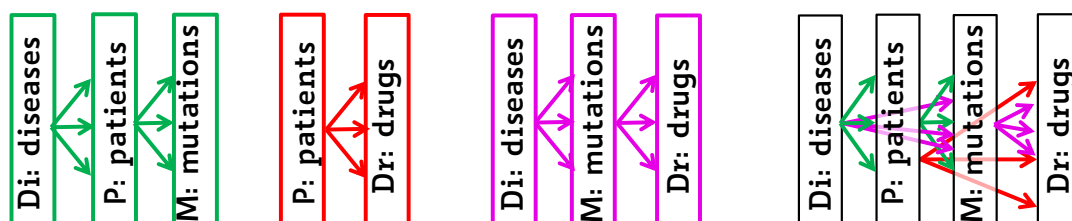
Red graph **R**: medical Record

Magenta graph **M**: Medical knowledge

$H = G \cup R \cup M$

51

SOME EXAMPLES (1)

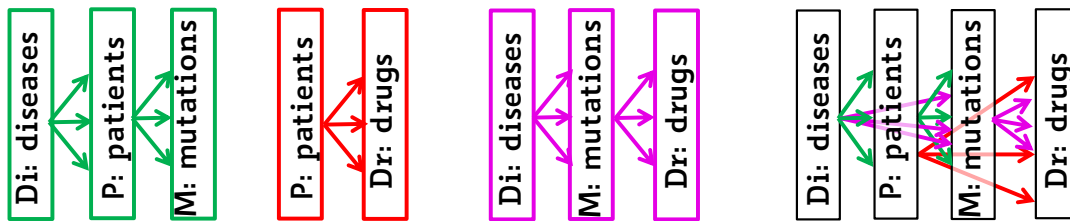


Example 1. Consider a projection on patients w.r.t. gene mutations, i.e. a new edge-labeled multi-graph $\text{Pr}(P|M)$ whose node set coincides with P and an edge (p_1, p_2) is in $E(\text{Pr}(P|M))$ if and only if there exists a node $m \in M$ such that $(p_1, m) \in E(G)$ and $(p_2, m) \in E(G)$; such an edge is labeled with m .

Of course, it is possible that p_1 and p_2 are connected by more edges with different labels.

52

SOME EXAMPLES (2)

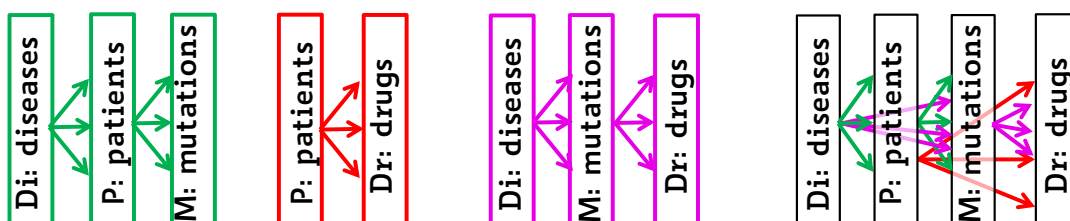


Keeping in mind that an edge labeled m in $\Pr(P|M)$ connects two patients sharing gene mutation m , our aim is to cluster the nodes of $\Pr(P|M)$ that are characterized by the same identical set of gene mutations.

It is easy to see that each such cluster appears in $\Pr(P|M)$ as a clique whose connections are multiple edges labeled with the same gene mutations.

53

SOME EXAMPLES (3)



Example 2. Consider subgraph $S(p, M, Dr)$. The subset of nodes from Dr in this graph is the set of all drugs known to act to one of the gene mutations of patient p ; so, hypothetically administering all these drugs to p would guarantee to treat all their possible gene mutations. But we need to give less drugs, anyway treating patient p at best; hence, from the drugs in this subgraph, we select only those such that all the gene mutations of patient p are treated: Min Dominating Set problem...

54

ONGOING PROBLEMS

PROBLEM 6.

- Validate the model using data

PROBLEM 7.

- Implement some features of the tool

PROBLEMS 8., 9., 10., ...

...