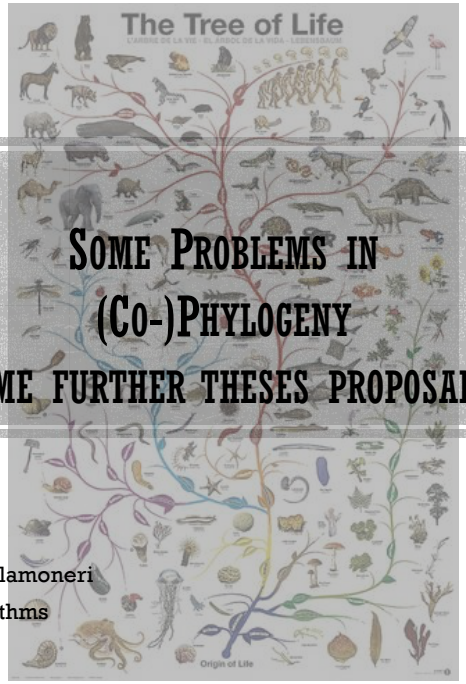
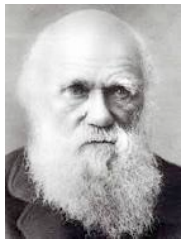


Some Problems in (Co-)Phylogeny (some further theses proposals)

1

Prof. Tiziana Calamoneri
Network Algorithms
A.y. 2019/20

Coevolution



"Thus 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

2

Motivation

- **Different systems "coevolve"**
  - hosts and their parasites or pathogens ⟵ same model
  - whole organisms and their genes
  - geographical areas and the species which inhabit them
  - cultural traditions and populations

3

The ingredients

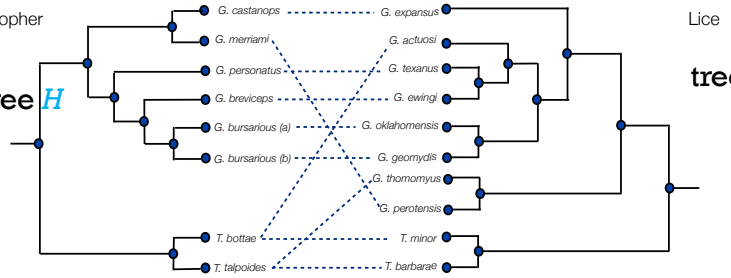
Gopher

Co-evolution

Lice

leaf mapping function *f*

tree *H*



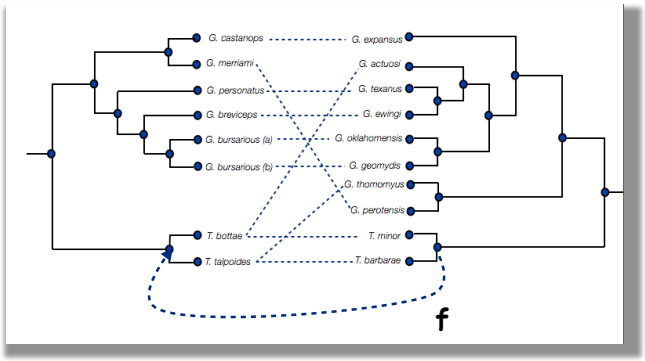
tree *P*

4

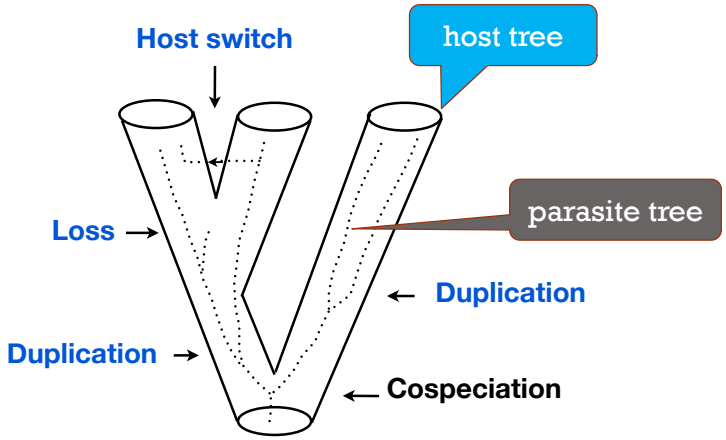
*adapted from Penn, Dustin J(Apr 2001) Coevolution: Host–Parasite. In: eLS. John Wiley & Sons Ltd, Chichester. http://www.els.net

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.



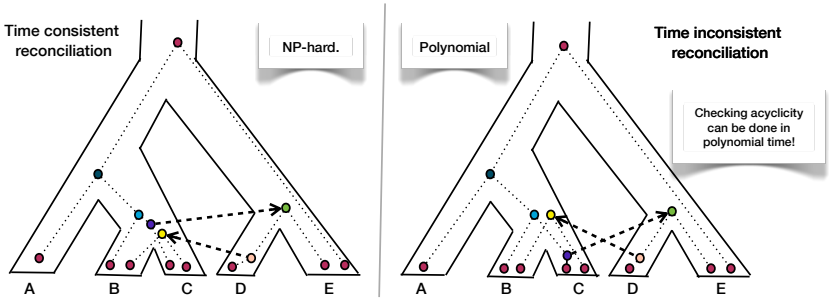
RECONCILIATIONS (2)



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).
- Aim: generating all the optimal reconciliations

CO-PHYLOGENY RECONSTRUCTION PROBLEM (2)



Time consistency

CO-PHYLOGENY RECONSTRUCTION PROBLEM (3)

- 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.

- Many papers and tools trying to solve this issue ✔

OUR PROBLEMS

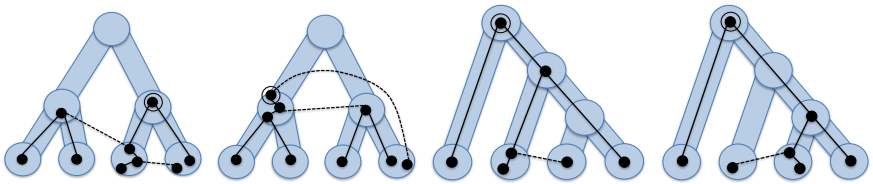
Biologists want to see all possible reconciliations, in order to understand which ones are biologically feasible and which ones are not.

Either reduce the cardinality of or cluster the set of optimal reconciliations

Visualize a given reconciliation in a "nice and clear" way

"SIMILAR" RECONCILIATIONS (1)

- Almost nothing in the literature:
  - similarity by event vectors
  - Similarity as the smallest number of operations needed to change one reconciliation into an- other

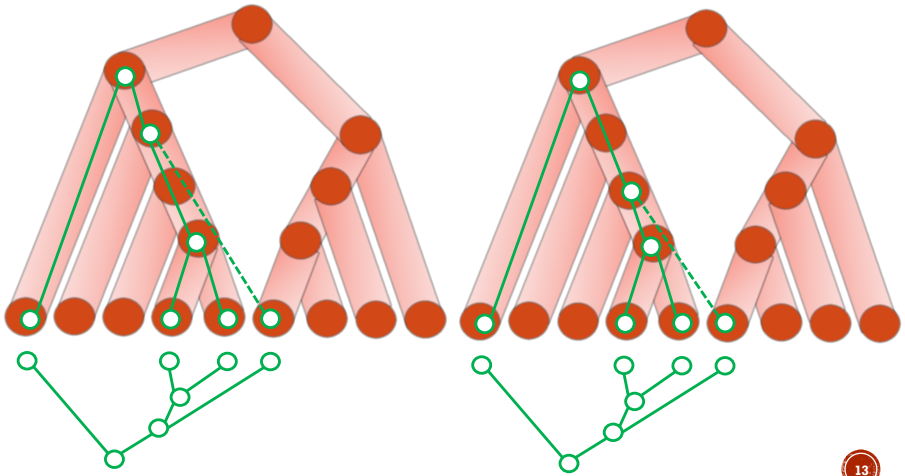


"SIMILAR" RECONCILIATIONS (2)

[Gastaldello, C., Sagot '19]

- We introduced a new notion of metric to measure similarity between reconciliations:
  - take into account host-switches
  - *R'* and *R"* are the same reconciliation iff they have the same host-switches
  - Characteristic vector of each reconciliation (an item for each possible host-switch): relation with the Hamming distance of sufficiently long binary arrays (-> Hypercube)

"SIMILAR" RECONCILIATIONS (3)



"SIMILAR" RECONCILIATIONS (4)

Dataset	# solutions	# $\sim_1$
GL	2	2
RH	2208	368
FD	408	180
COG2085	37568	3200
COG3715	9	7
COG4964	36	4
COG4965	640	576
PP	72	72
SFC	40	16
EC	18	18
PMP	2	2
PML	2	2
Wolbachia	$\sim 1.01 \cdot 10^{47}$	$\sim 3.77 \cdot 10^{44}$

With a cost vector promoting co-speciation and discouraging host-switch: (0,1,2,1)

CLUSTERING RECONCILIATIONS

- More open questions than solutions:
  - We would like to exploit the similarity with the Hypercube, but the set of all the reconciliations represent only a subset of the nodes of the hypercube
  - Do the connected components have any biological meaning? apparently no…
  - How to choose the cluster head?
  - Is the cluster head really representative?

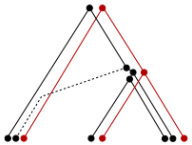
FURTHER PROBLEMS

- PROBLEM 1. Reconciliation comparison:
  - new metrics
  - exact and approximate algorithms
- More realistic models:
  - PROBLEM 2. deal with errors in phylogenetic trees (here the phylogenetic trees have been assumed to be correct, which may be not the case…)
  - PROBLEM 3. f is not a function: multiple hosts – multiple parasites (a single parasite can infect more than one host…)
  - PROBLEM 4. handle unrooted trees (many phylogenetic tree reconstruction algorithms produce unrooted trees; the outgroup method has the problem of the availability of a proper outgroup)

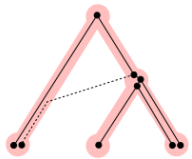
VISUALIZING RECONCILIATIONS (1)

[C. Di Donato, Mariottini, Patrignani '19]

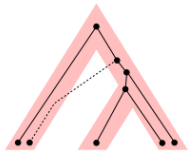
- Given H, P, φ and a reconciliation R, we have to draw H and P (on H) to highlight φ and R in a nice and clear way
- Three main strategies:
  1. representing two paired trees
  2. parasites are drawn inside their hosts
  3. host tree is made of pipes and parasites are drawn into the pipes



Strategy 1



Strategy 2

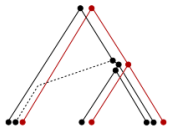


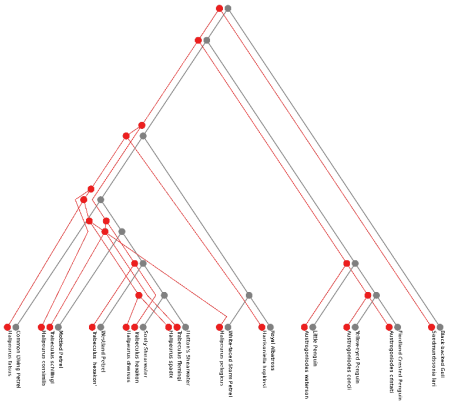
Strategy 3

VISUALIZING RECONCILIATIONS (2)

Example of the 1st strategy:

- CoRe-PA

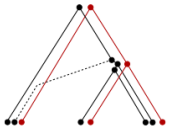


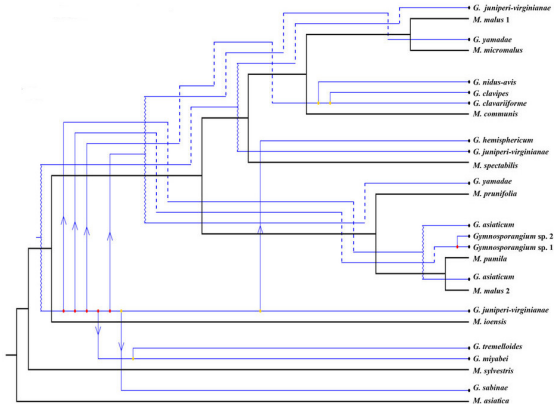


VISUALIZING RECONCILIATIONS (3)

Example of the 1st strategy:

- Jane 4

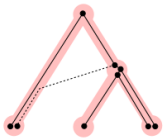


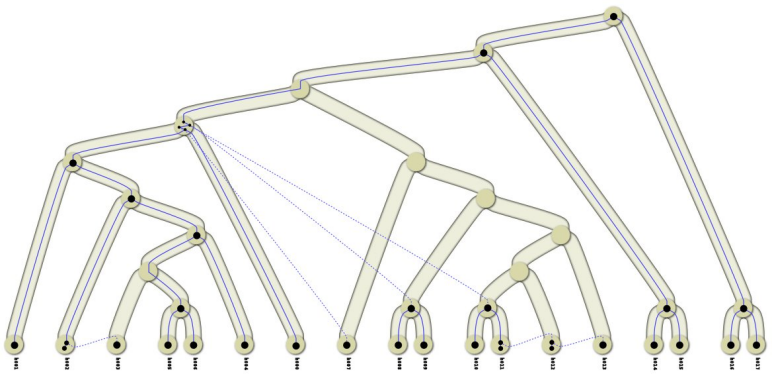


VISUALIZING RECONCILIATIONS (4)

Example of the 2nd strategy:

- CophyTrees

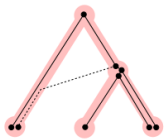


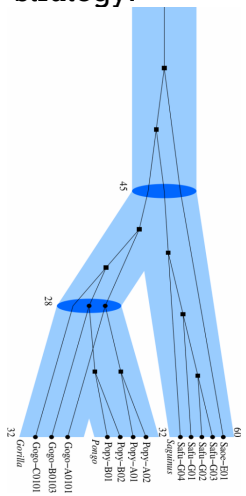


VISUALIZING RECONCILIATIONS (5)

Example of the 2nd/3rd strategy:

- Primetv

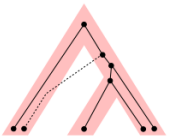


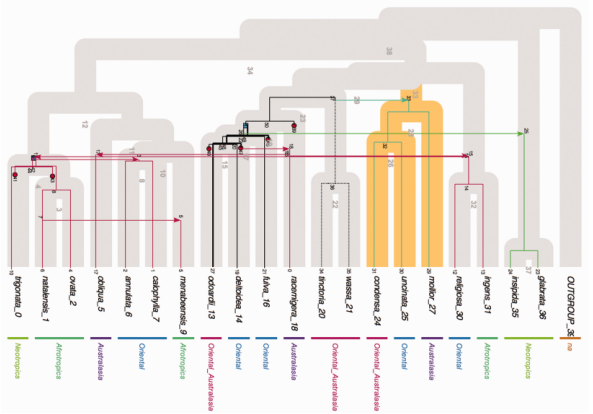


VISUALIZING RECONCILIATIONS (6)

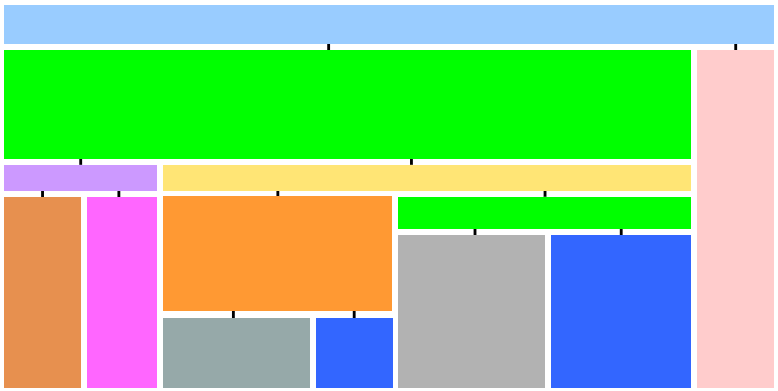
Example of the 3rd strategy:

- SylvX

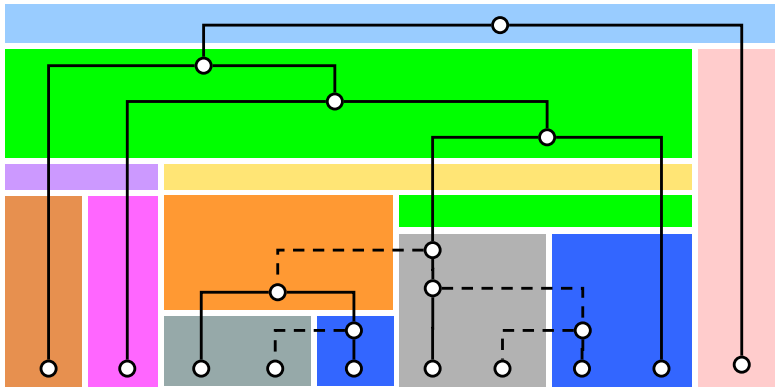




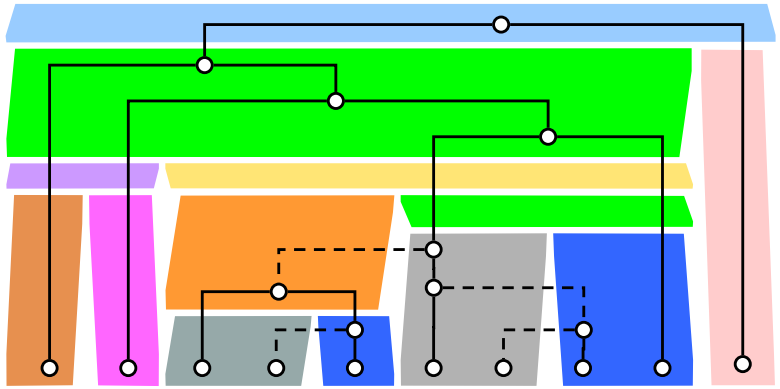
NEW METAPHOR: HP-DRAWINGS (1)



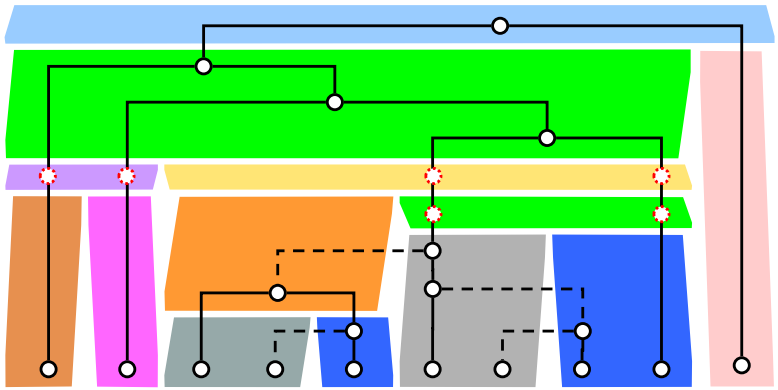
NEW METAPHOR: HP-DRAWINGS (2)



NEW METAPHOR: HP-DRAWINGS (3)

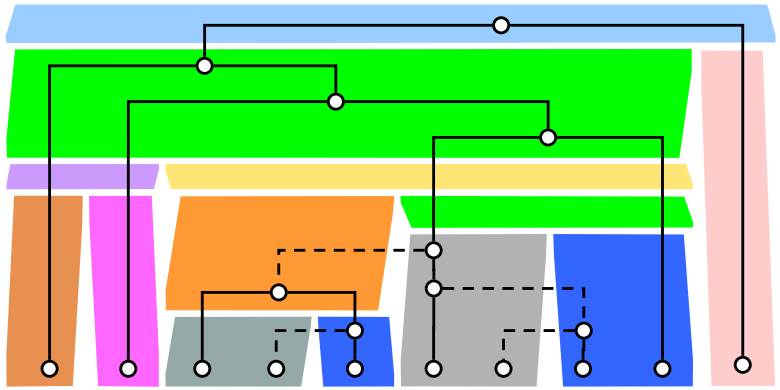


EVOLUTIONARY PHENOMENA (1)



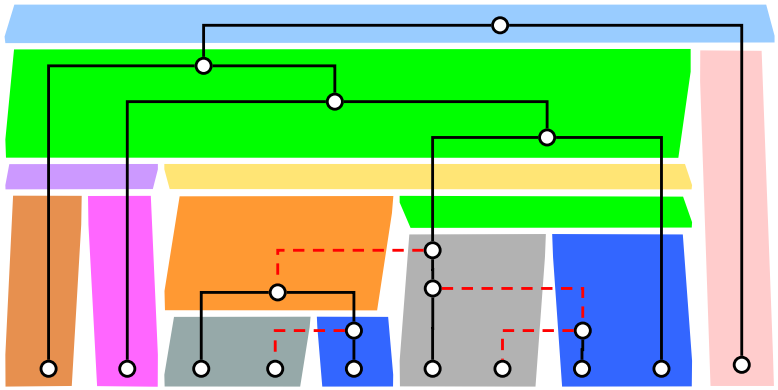
- Loss:
  - a parasite is transmitted to one child but not to the other child

EVOLUTIONARY PHENOMENA (2)



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

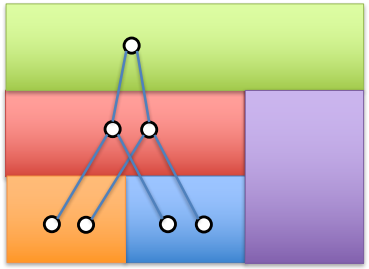
EVOLUTIONARY PHENOMENA (3)



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

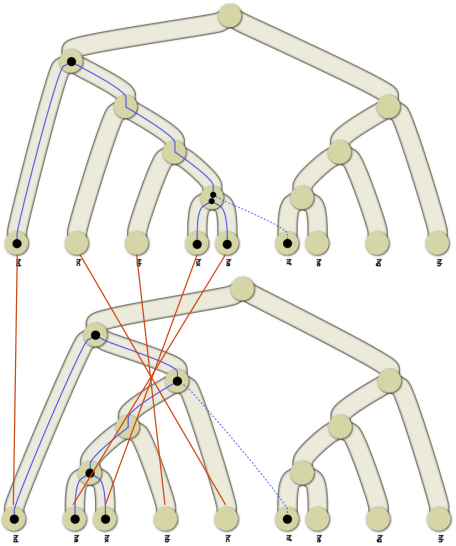
OPTIMIZATIONS (1)

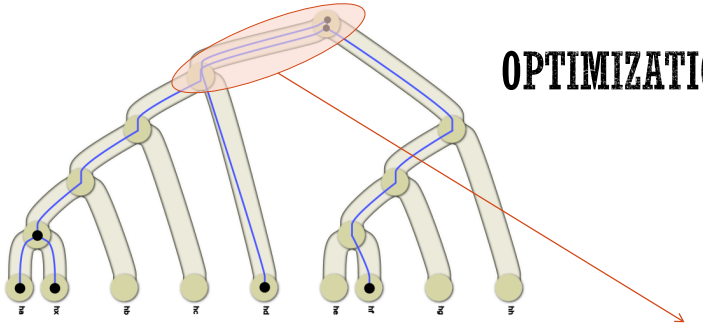
- Given *H*, *P*, *f* and a reconciliation *R*, we would like to:
  2. minimize the **crossing number** (not always possible to avoid crossings)



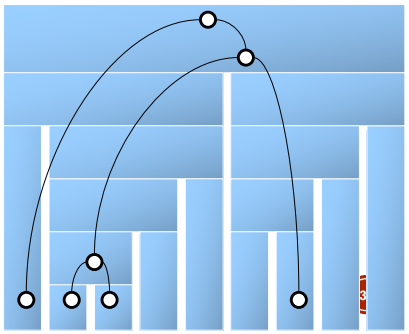
OPTIMIZATIONS (2)

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



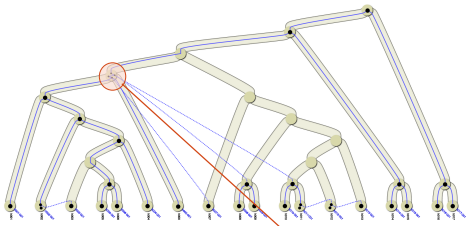


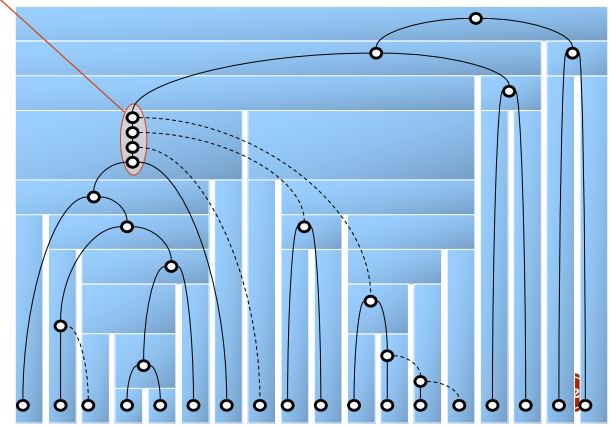
OPTIMIZATIONS (3)



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

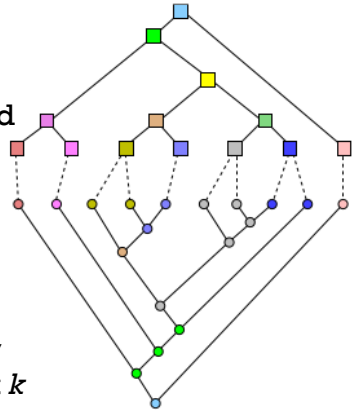
OPTIMIZATIONS (4)



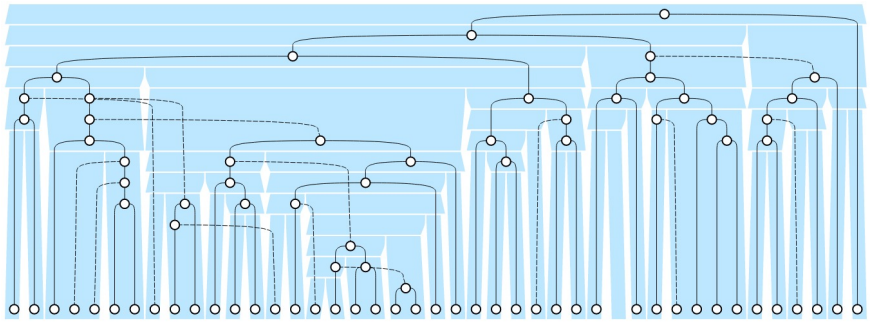


PLANAR INSTANCES

- 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 an drawing with at most k crossings is NP-complete.



AN EXAMPLE



FURTHER PROBLEMS

PROBLEM 5.

- Test the tool:
  - Do biologists like this metaphor?
- Models with more information:
  - handle additional information (e.g. geography) – colors?