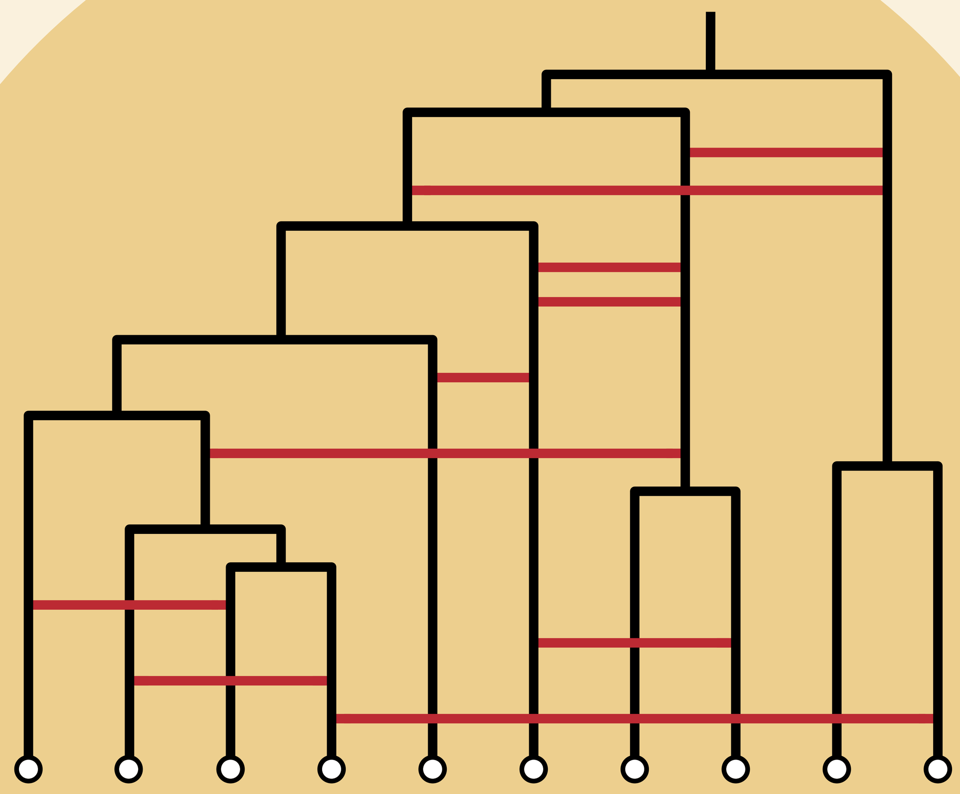
Minimising Crossings in a Tree-Based Network

Jonathan Klawitter • Peter Stumpf

Phylogenetic trees and **networks** are rooted, binary graphs with labelled leaves used to visualise evolutionary histories.

A **tree-based network** N is a phylogenetic network that has a spanning tree T which is the subdivision of a phylogenetic tree [1]. We call edges not covered by T **cross edges**.

A tree-based network is used to model reticulate events that extend a phylogenetic tree. Depending on the context, endpoints of cross edges may have the same height, e.g. for horizontal gene transfer [3], or different heights, e.g. for recombination [4].



What is the lowest number of crossings you can achieve?

We consider drawings of a tree-based network N on tree T where

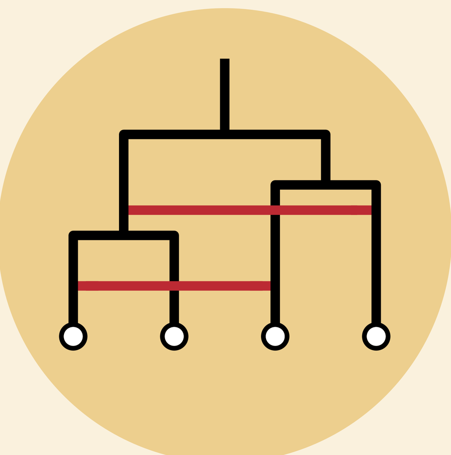
- T is drawn planar,
- leaves are equidistant, and
- heights are preserved.

We assume that heights are the same for leaves, but distinct otherwise, except for the two endpoints of horizontal cross edges.

Our goal is to **minimise the number of crossings**, which as we note is fully determined by

- the order of the leaves, or equivalently
- the rotations of the vertices of T.

Horizontal Drawing Style



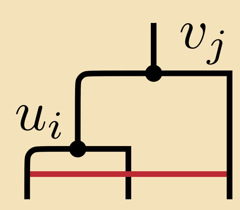
- Cross edges are horizontal lines

Note: Rotation of one vertex can effect best rotation of other vertices.

NP-complete

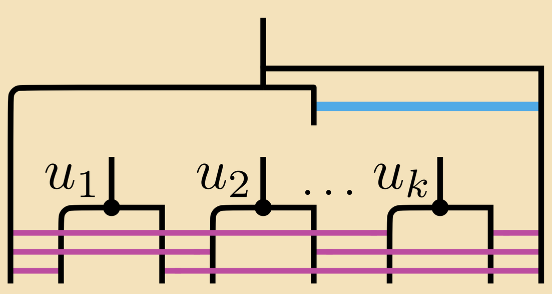
Proof by reduction of MAX-CUT

Gadget for edge $\{u, v\}$



adds 1 crossing iff u_i and v_i have "same" rotation

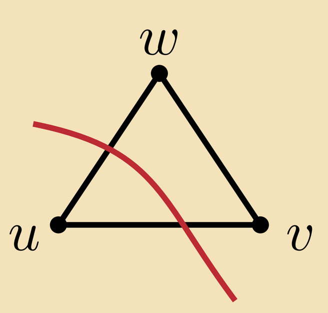
Gadget for vertex u

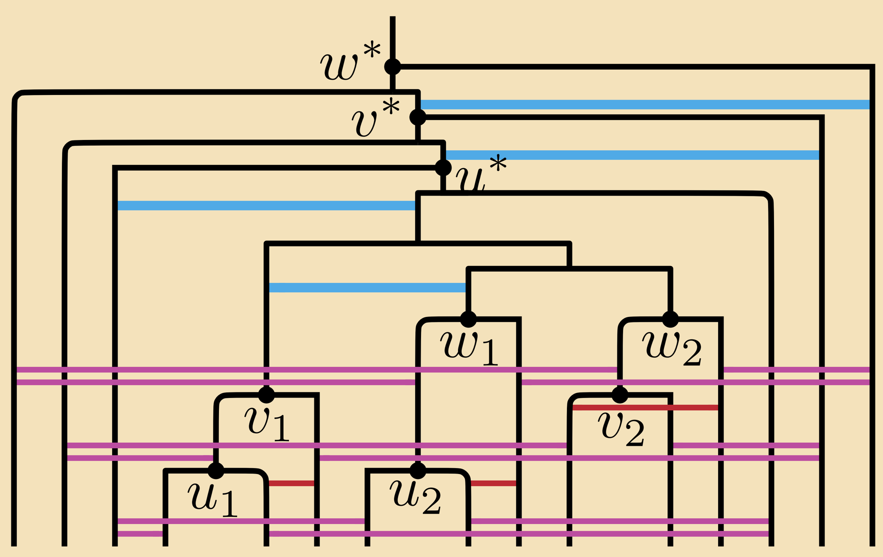


u_i's forced to have same rotation with purple bundles

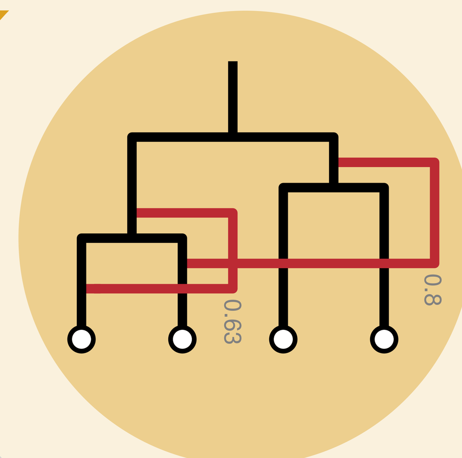
thick blue bundles fix upper part of tree

Example





Ears Drawing Style



- Cross edges drawn with two bends
- Vertical segment right of subtree containing endpoints
- Allows labelling with statistical support

Note: Best rotation of vertex determined by width of right subtree where cross edges leave the left subtree and vice versa.

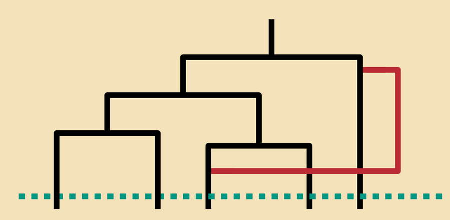
Quadratic time solvable

Sweep from leaves to root

Let N have n vertices and k cross edges.

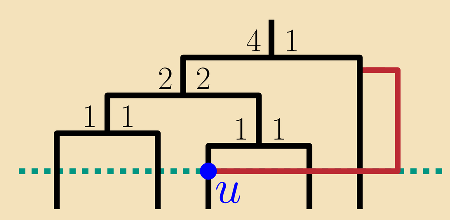
Vertices of T store

- potential crossings from left/right subtree through right/left subtree.

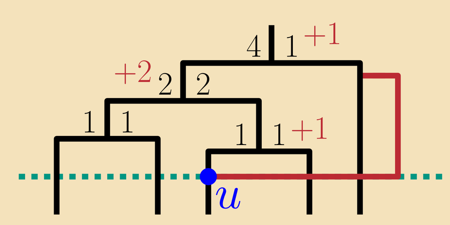


Reaching u of cross edge $\{u, v\}$

- compute width of left/right subtree for v of T at height of u in $O(n)$, and

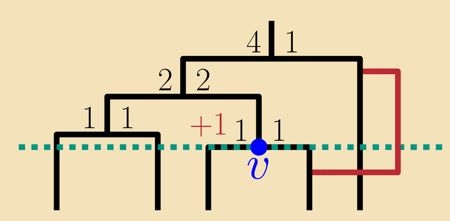


- propagate potential crossings with other subtrees up to lca(u, v) in $O(n)$ [2].



Reaching vertex v of T

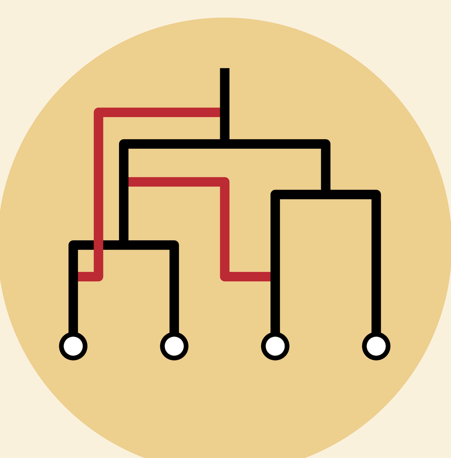
- decide rotation based on number of potential crossings.



Lastly, extend partial order of nested vertical segments to total order of all. Algorithm runs in $O(nk)$.

By propagating potential crossings up to the root, the algorithm also works for a modified ears drawing style, where the vertical segments are to the far right of T.

Snake Drawing Styles



Cross edges drawn x-monotone with/as
• two bends, • curve, • straight line.

Reduces to horizontal style ⇒ NP-complete

1. A. R. Francis and M. Steel. Which Phylogenetic Networks are Merely Trees with Additional Arcs? *Systematic Biology*, 64(5):768–777, 2015.
2. H. N. Gabow and R. E. Tarjan. A Linear-time Algorithm for a Special Case of Disjoint Set Union. In *STOC '83*, p246–251.
3. V. Kumar, F. Lammers, T. Bidon, M. Pfenninger, L. Kolter, M. A. Nilsson, and A. Janke. The evolutionary history of bears is characterized by gene flow across species. *Scientific Reports*, 7(1), 2017.
4. T. G. Vaughan, D. Welch, A. J. Drummond, P. J. Biggs, T. George, and N. P. French. Inferring Ancestral Recombination Graphs from Bacterial Genomic Data. *Genetics*, 205(2):857–870, 2017.