Tree-like reticulation networks

Andrew R Francis

Centre for Research in Mathematics
University of Western Sydney
Australia

Phylomania 2014.

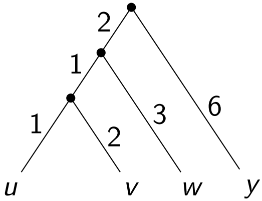
Andrew R. Francis, Mike Steel

Tree-like reticulation networks - when do tree-like distances also support reticulate evolution?

Mathematical Biosciences, in press (arXiv:1405.2965).

Tree metrics

- A phylogenetic tree with edge weights defines a metric on the leaves:
  - add weights on the unique path.

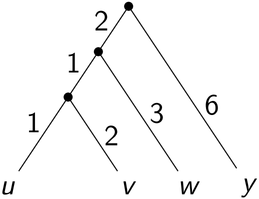


	u	v	w	y
u	0	3	5	10
v		0	6	11
w			0	11
y				0

- A metric that can be placed on a tree is called a tree metric.

Tree metrics

- A phylogenetic tree with edge weights defines a metric on the leaves:
  - add weights on the unique path.



	u	v	w	y
u	0	3	5	10
v		0	6	11
w			0	11
y				0

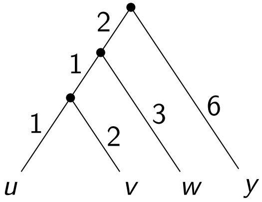
- A metric that can be placed on a tree is called a tree metric.
- A metric d is a tree metric if and only if it satisfies the four point condition:
  - for all quartets of leaves $\{u, v, w, y\}$, two out of

$$d(u,v) + d(w,y), \quad d(u,w) + d(v,y), \quad d(u,y) + d(v,w)$$

are equal, and are greater than or equal to the other one.

Tree metrics

- A phylogenetic tree with edge weights defines a metric on the leaves:
  - add weights on the unique path.



	u	v	w	y
u	0	3	5	10
v		0	6	11
w			0	11
y				0

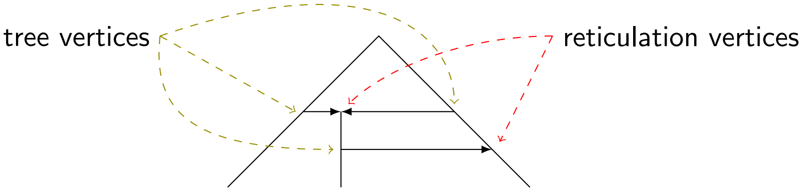
- A metric that can be placed on a tree is called a tree metric.
- A metric d is a tree metric if and only if it satisfies the four point condition:
  - for all quartets of leaves $\{u, v, w, y\}$, two out of

$$d(u,v) + d(w,y), \quad d(u,w) + d(v,y), \quad d(u,y) + d(v,w)$$

  are equal, and are greater than or equal to the other one.
  - In the above we have $3 + 11 = 14$, $5 + 11 = 16$ and $10 + 6 = 16$.

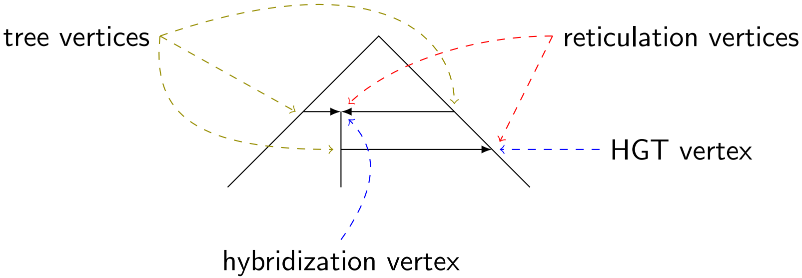
Reticulated networks

Any metric may be able to be represented on a *reticulated network*:



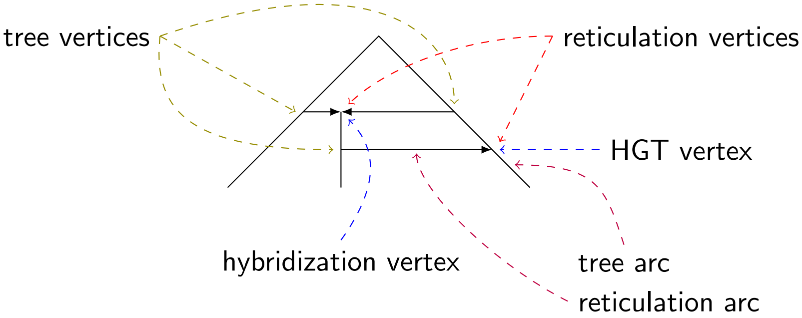
Reticulated networks

Any metric may be able to be represented on a *reticulated network*:



Reticulated networks

Any metric may be able to be represented on a *reticulated network*:



So we have

1. 4PC satisfied $\implies$ there is a tree that can represent the metric.
2. 4PC not satisfied $\implies$ there may be a reticulated network that can represent the metric.

(Note: the 4PC statement is an if-and-only-if).

- What's missing?

So we have

1. 4PC satisfied $\implies$ there is a tree that can represent the metric.
2. 4PC not satisfied $\implies$ there may be a reticulated network that can represent the metric.

(Note: the 4PC statement is an if-and-only-if).

- What's missing?
- The 4PC does not rule out a tree metric also being representable on a reticulated network.
  - One side of the if-and-only-if is an existence statement.

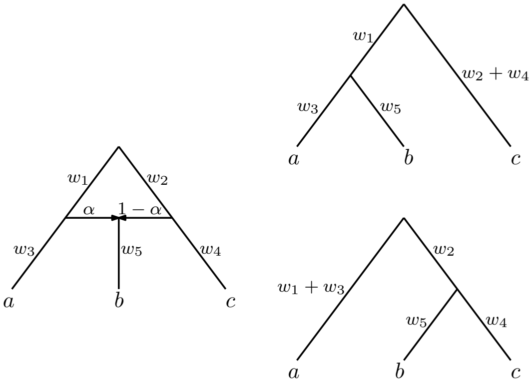
Take home message of this talk:

- A tree metric can also be represented on reticulated networks using average distances.
  - We are able to characterise this precisely in some cases, but not all yet!

. . . now to clarify what is meant by average distances . . .

Metrics on reticulated networks.

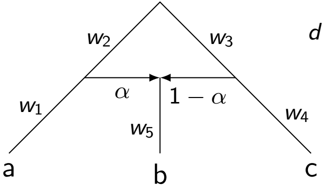
- Let $T(N)$ be the set of trees "displayed" by N. E.g.



- For the purposes of distance, we treat a reticulated network N as the weighted sum of the trees in $T(N)$:

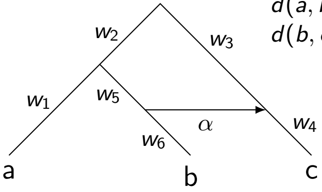
Metrics on reticulated networks.

1. HGT.



$$\begin{aligned} d(a,b) &= \alpha(w_1+w_5)+(1-\alpha)(w_1+w_2+w_3+w_5) \\ &= w_1+w_5+(1-\alpha)(w_2+w_3) \\ d(b,c) &= w_4+w_5+\alpha(w_2+w_3) \\ d(a,c) &= w_1+w_2+w_3+w_4 \end{aligned}$$

2. Hybridization.



$$\begin{aligned} d(a,b) &= w_1+w_5+w_6 \\ d(b,c) &= \alpha(w_4+w_6)+(1-\alpha)(w_2+w_3+w_4+w_5+w_6) \\ &= w_4+w_6+(1-\alpha)(w_2+w_3+w_5) \\ d(a,c) &= \alpha(w_1+w_5+w_4)+ \\ &\quad (1-\alpha)(w_1+w_2+w_3+w_4) \\ &= w_1+w_4+\alpha w_5+(1-\alpha)(w_2+w_3) \end{aligned}$$

Results

Theorem

Suppose that all the trees in $T(N)$ are isomorphic as unrooted phylogenetic X-trees to some tree T. Then d_N is a tree metric that is represented by T.

- For example, if there is a single reticulation near the root, the network is treelike.
- We can be more precise.

Tree-like hybridization networks

Theorem

Let X be a finite set of taxa, and suppose d is a metric on X.

- *If d is a binary tree metric, then it is a metric on a primitive 1-hybridization network N.*
- *If N is a hybridization network, and d is a tree metric on N, then N is either a tree, or is a primitive 1-hybridization network.*

Tree-like hybridization networks

Theorem

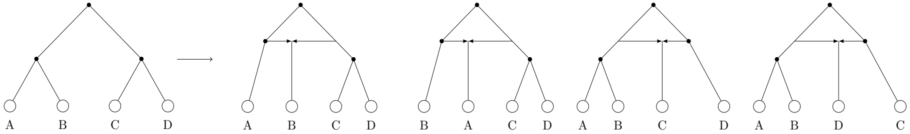
Let X be a finite set of taxa, and suppose d is a metric on X.

- *If d is a binary tree metric, then it is a metric on a primitive 1-hybridization network N.*
- *If N is a hybridization network, and d is a tree metric on N, then N is either a tree, or is a primitive 1-hybridization network.*

Theorem

For each tree metric on n leaves, there are $4(n-3)$ 1-hybridization networks that realise the metric.

Proof:



Tree-like HGT networks

- Let N be an HGT network with T_N the underlying tree (delete all reticulation arcs).

Lemma

If each reticulation arc in N is between adjacent tree-arcs of T_N, then d_N is tree-like on T_N.

- This means we have huge numbers of reticulated (HGT) networks whose metrics are tree-like!

Tree-like HGT networks

- Let N be an HGT network with T_N the underlying tree (delete all reticulation arcs).

Lemma

If each reticulation arc in N is between adjacent tree-arcs of T_N, then d_N is tree-like on T_N.

- This means we have huge numbers of reticulated (HGT) networks whose metrics are tree-like!

Theorem

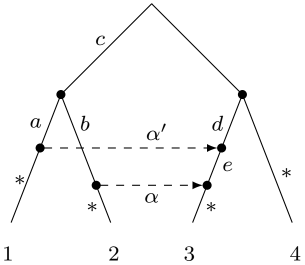
Suppose d_N is a metric from an HGT network N with a single reticulation arc.
Then d_N is tree-like if and only if that arc is either

1. *from one arc to an adjacent arc, or*
2. *between a root arc and one of the two children of the other root arc.*

The only tree that harbours a representation for d_N is T_N.

Strange magic

- There are 2-reticulated HGT networks N that can be represented on T_N and (for other parameter settings) on a tree that is different from T_N, even when the mixing distribution treats the two reticulations independently.



- Setting $\alpha \geq \frac{1}{2}$, $b \geq a$ and $\alpha' a = (1 - \alpha')c$ gives $14|23$.

Further questions

1. Is the following true:
   For any two binary phylogenetic X-trees T_1 and T_2, is there an HGT network N for which $T_N = T_1$ and yet where d_N is representable on T_2 (mixing distribution given by the independence model)?

Further questions

1. Is the following true:
   For any two binary phylogenetic X-trees T_1 and T_2, is there an HGT network N for which $T_N = T_1$ and yet where d_N is representable on T_2 (mixing distribution given by the independence model)?

2. Let $\rho(d)$ denote the minimum number of hybridizations required to represent d on a hybridization or an HGT network.
   - What conditions characterise those metrics d with $\rho(d) = 1$?
   - What about $\rho(d) = k$ for any $k \geq 1$?

   (We know about $\rho(d) = 0$: the 4PC).