In this paper Briand et al introduce a new variant of the Labelled Robinson-Foulds metric. The new distance has several desirable features. Namely, it is computable in linear time, and it restricts to the RF distance when all internal labels are identical. The paper is very clearly written, the proofs all make sense. The simulation study on whether or not denser taxon-sampling improves reconciled gene tree inference is interesting.

The paper builds on recent work by several of the authors in a conference paper from earlier this year where they introduced a variant of the RF distance which allowed edge contracts and expansions as well as node label flips (ELRF). The ELFR turned out to be difficult to compute as sometimes the optimal path from one tree to another required the contraction of an edge that was found in both trees.

In this paper they propose a subtle modification to the allowable edit operations that turns out to simplify things and hence allows for a linear time algorithm to compute the distance between two trees. The edit operations are reformulated to be based on nodes rather than edges. In the earlier approach an edge could only be deleted if it connected nodes with the same label, the new formulation does not have this restriction. Similarly, in the previous formulation an edge insertion created a new node with a label that had to match the "parent" label, whereas in the new edit operations a node insertion can be given any label.

Edit operations between a pair of trees basically consist of contracting "bad edges" (i.e. those that correspond to a bipartition that is not found in both trees) and then popping out required edges in such a way that internal node labels will match. Indeed, the definition of the edit operations leads to a very efficient and easy to understand algorithm that starts by partitioning each tree based on first identifying "good edges" and using these to decompose both trees into islands, these islands can then be paired up based on their terminal labels. To convert one island to its counterpart in the matched pair requires that all the edges be contracted down to a star tree (using the node deletion operation) and then popped back again. This will use 2 operations for each edge in an island. Provided the island pair has at least some overlap in the labels then new nodes can be inserted with the correct label at no extra cost in terms of edit operations, otherwise there will be a cost of 1.

The authors are quite open about the fact that the edit operations and distance defined is not necessarily very biologically reasonable. Indeed, it does seem a bit odd that e.g. the top tree below has the same distance to both of the lower trees.
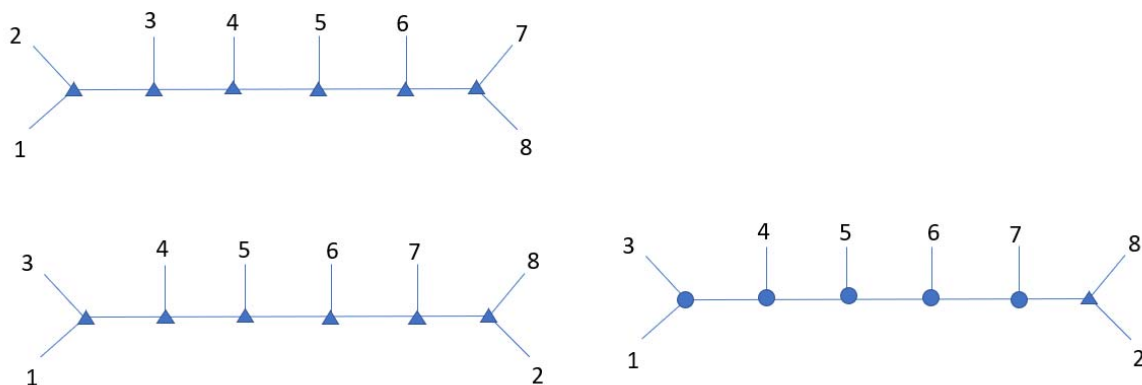
While the distance has been developed with application to gene-tree species tree reconciliation in mind, the applications are obviously much broader. It would be interesting to repeat the simulation study in this paper with a focus on ancestral state reconstruction to address the question of whether denser sampling improves accuracy there too.

Minor typos / queries

Pg 2, column 1, I am not sure what you mean by "node labels in a given tree are pairwise different", just that no labels are repeated?

Pg 2, column 1,

To address the distance drawbacks

-> To address the distance's drawbacks

Pg 2, col 1

Is the node flip operation called a 'flip' because there can only be two kinds of labels? If so, might be worth mentioning this

Pg 3, column 1

IF the node insertion operation was changed to restrict lambda to be lambda(y) would the operations then be equivalent to those for ELRF?

Pg 3, col 2, lemma 1

Is it also worth showing that the space is connected? I had a brief worry about this when I saw that operations were not allowed at the root, but some scribbling convinced me that this doesn't create an issue. I might just be worth adding a sentence or two.

Pg 4, col 1

It may be a start tree

->

It may be a star tree

Pg 5, col 1

Allow us considering each

->

Allow us to consider each

Pg 5, col 2

Not clear what you mean by "leftmost"

Were e={x,y}

->

Where e = {x,y}

Pg 6, col

The next stem -> the next step

Pg 6, col 2

Does labelled gene tree inference benefits from

->

Does labelled gene tree inference benefit from