



A fast heuristic for the calculation of a cherry-operation distance on phylogenetic networks

Jared Rost, Kaari Landry, Olivier Tremblay-Savard



1 - Introduction

- Phylogenetic networks are better suited than trees to represent complex evolutionary relationships.
- Over time research has focused on solving problems related to phylogenetic networks like construction, and tree containment.
- Cherry-picking sequences (CPSs) are a tool for solving such problems.
- CPSs are made of operations that reduce a network by removing 1 leaf of a simple cherry or 1 edge of a reticulated cherry until the network is fully reduced [1].



2 – Previous Work

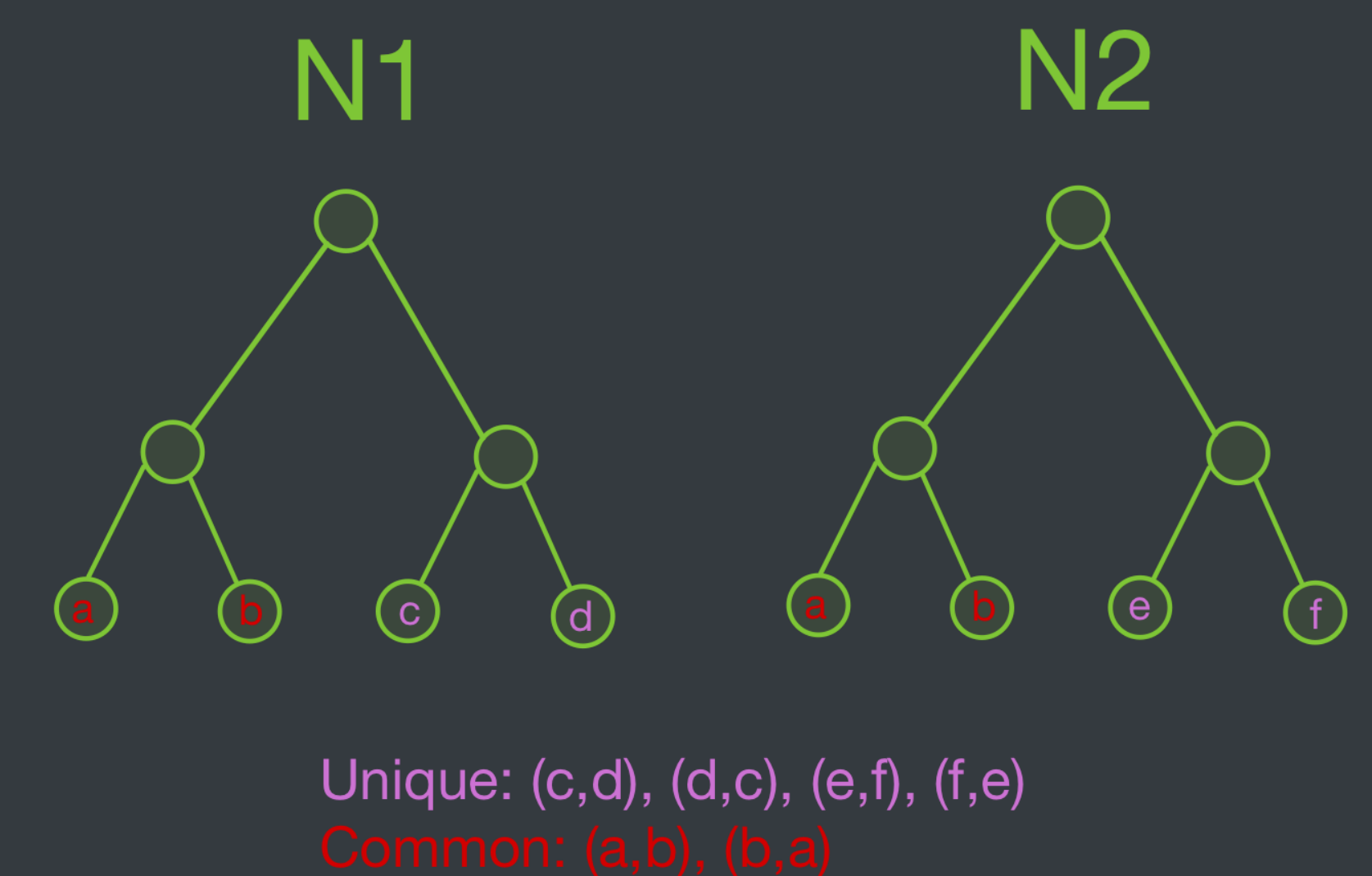
- The tail distance, proposed by Landry et al., is based on the structure of the networks' CPSs and seeks to maximize the common suffix [2].
- The more similar the networks are, the larger the suffix their CPSs will have, and thus, the smaller the distance is between them.
- Calculating the tail distance exactly has been proven to be an NP-hard problem [2].
- Landry et al. put forward a fixed parameter tractable algorithm, in relation to the total number of reticulations in both networks, that calculates the exact distance of two restricted networks [3].

3 – Objective

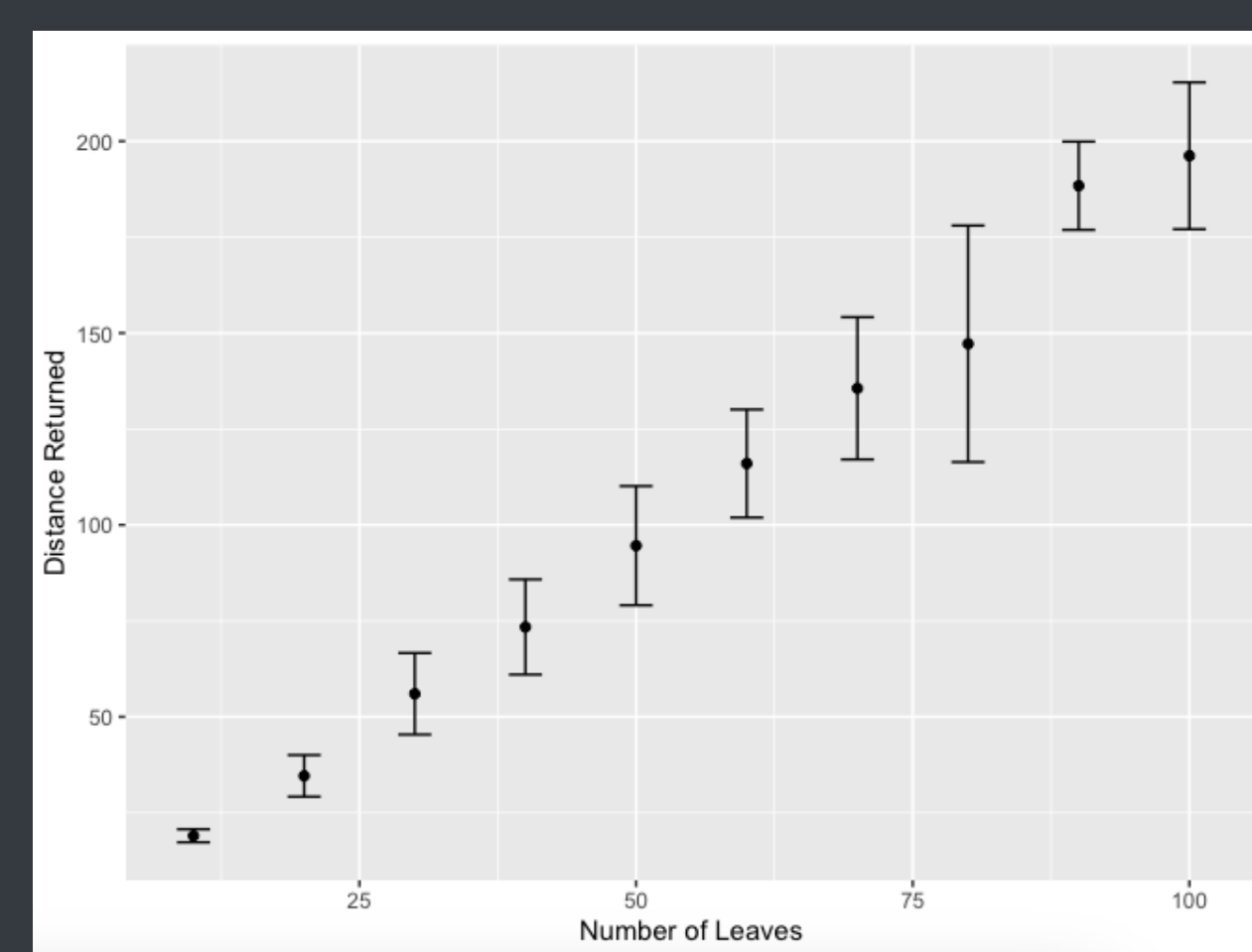
To create a heuristic that calculates the tail distance with reasonable accuracy and is faster than the fixed-parameter tractable time exact algorithm.

4 – Methods

- How it works:
 - Makes a list of common cherries vs unique cherries in each network.
 - Finds the unique/common cherry that, if picked, results in the most common cherries in subnetworks. Depth determines how many cycles ahead it looks.
 - Removes a unique cherry from the larger network, if none are available then it removes a common cherry.

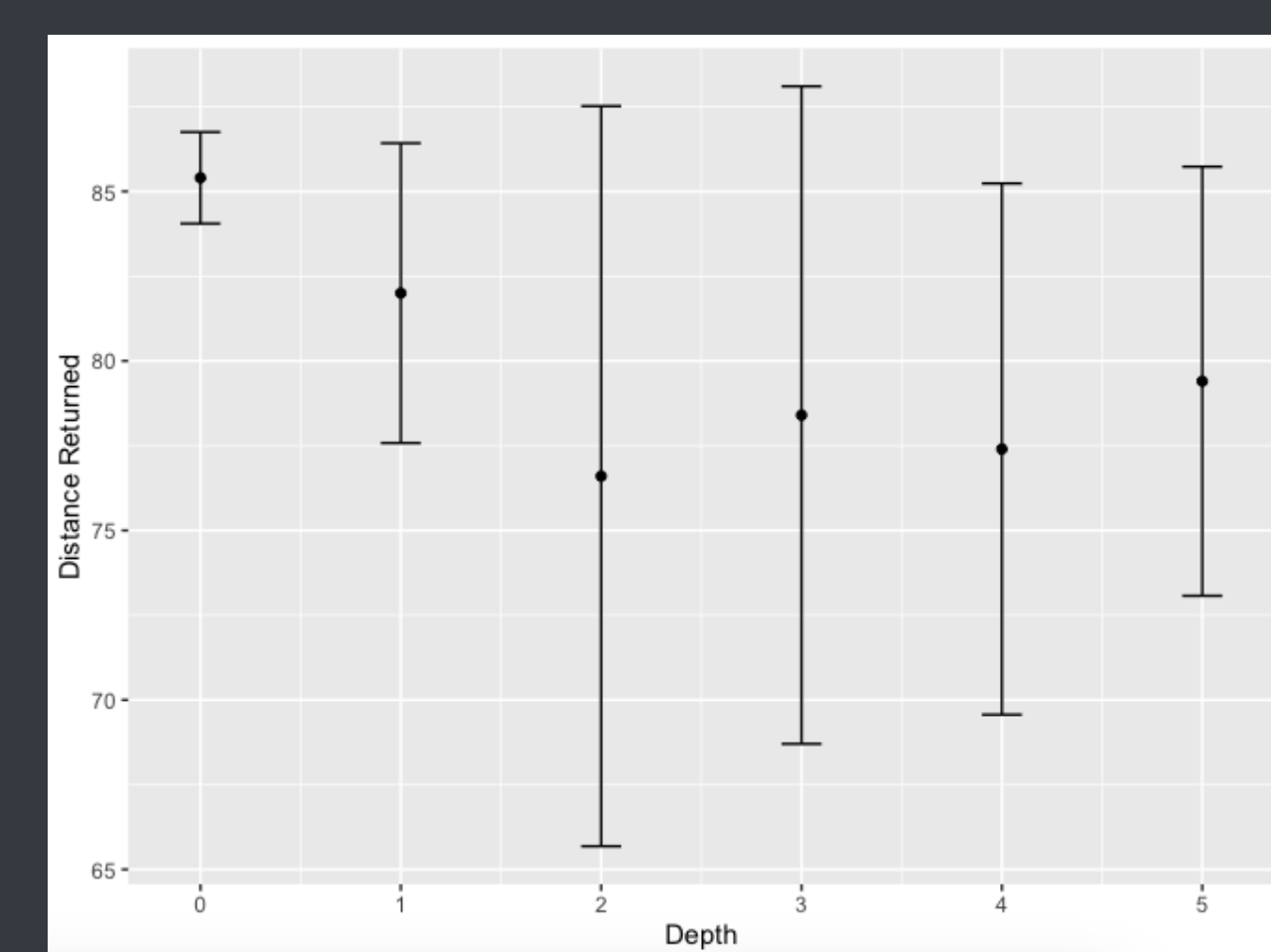


5 – Results



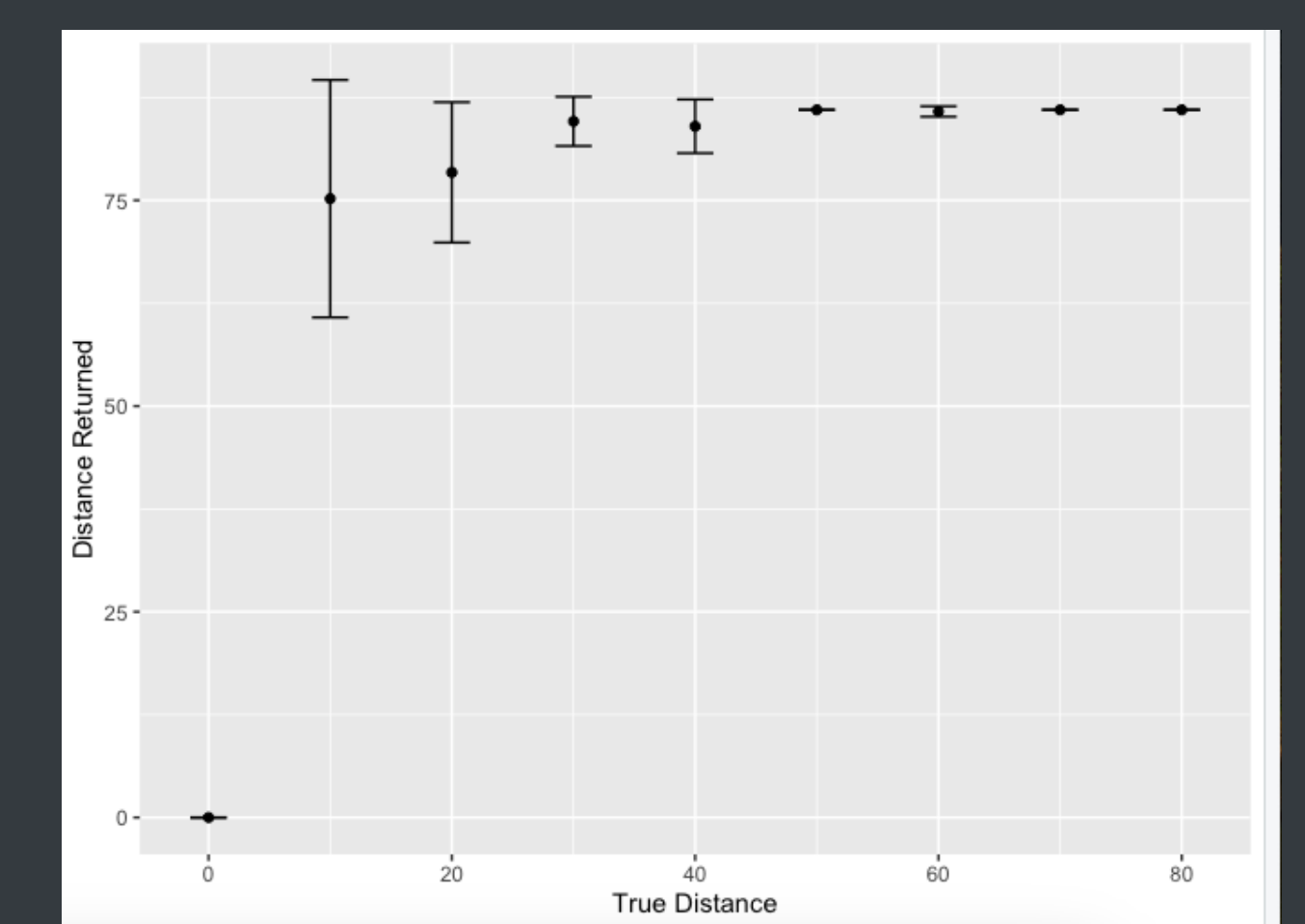
Number of Leaves

- 10% Reticulations
- Distance of 10
- Depth of 3



Depth

- 40 Leaves
- 4 Reticulations
- Distance of 10



Distance

- 40 Leaves
- 4 Reticulations
- Depth of 3

6 – Conclusion

- The average distance returned seems to be slightly under the max possible distance no matter the number of leaves.
- The algorithm is accurate when the two networks are identical and consistently returns near the max possible distance when the true distance approaches the number of leaves.
- It is consistently inaccurate at low depth, but results vary wildly at higher depths.

6 – Citations / Acknowledgements

- [1] Humphries PJ, Linz S, Semple C. Cherry picking: a characterization of the temporal hybridization number for a set of phylogenies. Bull Math Biol. 2013 Oct;75(10):1879-90. doi: 10.1007/s11538-013-9874-x. Epub 2013 Aug 8. PMID: 23925727.
- [2] Landry K, Teodocio A, Lafond M, Tremblay-Savard O. Defining Phylogenetic Network Distances Using Cherry Operations. IEEE/ACM Trans Comput Biol Bioinform. 2023 May-Jun;20(3):1654-1666. doi: 10.1109/TCBB.2022.3162991. Epub 2023 Jun 5. PMID: 35349447.
- [3] Landry, K., Tremblay-Savard, O., Lafond, M. (2023). Finding Agreement Cherry-Reduced Subnetworks in Level-1 Networks. In: Jahn, K., Vinař, T. (eds) Comparative Genomics. RECOMB-CG 2023. Lecture Notes in Computer Science(), vol 13883. Springer, Cham. https://doi.org/10.1007/978-3-031-36911-7_12

This work was supported by grants from NSERC and the University of Manitoba

