Program Robinson and Foulds

Vladimir Makarenkov, September 1999
Departement de sciences biologiques
Universite de Montreal
C.P. 6128, succursale Centre-ville
Montréal, Québec H3C 3J7, Canada
e-mail: **makarenv@ere.umontreal.ca**

What does program Robinson and Foulds do?

This program allows to compute in the optimal time *the Robinson and Foulds topological distance* between two or more *additive trees* given their distance matrices. The program is based on the algorithm proposed by Makarenkov and Leclerc (1999, b) for calculation of this distance. The algorithm implemented in this program uses the notion of circular orders to compare the topology of two trees. It employs circular order tree reconstruction to compute an ordered bipartition table of the tree edges for both given distance matrices. These bipartition tables are then compared to determine the Robinson and Foulds topologic distance, known to be an important criterion of tree similarity. The algorithm described in Makarenkov and Leclerc (1999, b) has optimal time complexity, requiring $O(n^2)$ time when performed on two $n \times n$ distance matrices.

The Robinson and Foulds topologic distance is an important and frequently used tool to compare additive (phylogenetic) tree structures (see for instance Robinson and Foulds (1981), Makarenkov and Leclerc (1997)). This distance is equal to the minimum number of elementary operations, consisting of merging or splitting nodes, necessary to transform one tree into the other. As proved in Robinson and Foulds (1981), it is also the number of bipartitions, or Buneman's splits (1971), which belong to exactly one of the two trees. If we deal with two unrooted trees having no internal vertices labeled according to the elements of X, the Robinson and Foulds distance of on the set X of n elements varies between 0 (when the trees are isomorphic) and $2n$-6 (when all non-trivial bipartitions in two trees are different; a trivial bipartition corresponds to an edge incident to a leaf).

The performed algorithm is part of the computer package T-REX (Macintosh and Windows versions available on the World Wide Web at http://alize.ere.umontreal.ca/~casgrain/en/labo/t-rex/index.html) including some popular methods of tree reconstruction such as ADDTREE by Sattath and Tversky (1977), the Neighbor Joining method by Saitou and Nei (1987), the Unweighted

Neighbor Joining method by Gascuel (1997), the Method of Weights by Makarenkov and Leclerc (1999), and others.

Input files

The input data file is an ASCII text file which contains the matrix size *n* followed by a sequence of additive distance matrices $\mathbf{X}_i$ (*n*x*n*), *i*=1...*m*.

Each matrix $\mathbf{X}_i$ is an additive distance matrix verifying the four-point condition.

The data file is organized as follows:

- First, a line the parameter *n*.

  n is the number of rows and columns in all the the matrices $\mathbf{X}_i$ included in the file.

- The additive distance matrices $\mathbf{X}_i$ included follow. A row of data can take as many successive physical lines as needed. Values in the same line are separated by one or more spaces; the number of spaces does not matter.

Options of the program

This program computes in the optimal time the Robinson and Foulds topological distance(s) between two or more matrices provided in the input file. This distance is computed between the first matrix in the input file and all the following ones.

Output file

The output is produced in a separate output file. The output consists of the sequence of *m*-1 values of the Robinson and Foulds topological distances between the first matrix in the input file and all the following ones.

Disclaimer

This program is provided without any explicit or implicit warranty of correct functioning. It has been developed as part of a university-based research program. If, however, you should encounter problems with this program, the author will be happy to help solve them. Researchers may use this

program for scientific purposes, but the source code remains the property of Vladimir Makarenkov. Publications should give proper credit to the method by referring to corresponding author's papers. Users of the program may refer to the present user's manual as follows:

Makarenkov, V., 1999. *Program Robinson and Foulds.* Departement de sciences biologiques, Universite de Montreal. 4 pages.

Technical notes

The program is distributed in a variety of formats:

• C source code for Macintosh and for Windows (the file source.c), which can be compiled using a C/C++ compiler.

• Compiled versions of the program for Win32 compatible computers (RobFoul.exe). The executable file is a Win32 "**console**" executable, not DOS executables. Therefore it cannot run under plain DOS, nor in a DOS window under Windows 3.x, only in Windows 95/98 or Windows NT consoles.

• Compiled version for PowerPC processors for Macintosh (file Robinson_and_Foulds_PPC).

• C source code for UNIX (the file source.c) as well as the corresponding Make File.

References

Buneman, P. 1971. The Recovery of Trees from Measures of Dissimilarity, 387-395. In *Mathematics in Archaeological and Historical Sciences*, eds. F.R. Hodson, D.G. Kendall and P. Tautu, Edinburgh : Edinburgh University Press.

Day W.H.E. 1985. *Optimal Algorithms for Comparing Trees with Labelled Leaves*, J. of Classification 2, 7-28.

Gascuel O. 1997. Concerning the NJ algorithm and its unweighted version UNJ, 149-170 In *Mathematical hierarchies and Biology* (B. Mirkin *et al*., eds.), *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 37, Amer. Math. Soc., Providence, R.I.

Makarenkov V., and Leclerc B. 1997. Tree metrics and their circular orders: some uses for the reconstruction and fitting of phylogenetic trees, 183-208. *In Mathematical hierarchies and*

Biology (B. Mirkin et al., eds.), DIMACS Series in Discrete Mathematics and Theoretical Computer Science 37, Amer. Math. Soc., Providence, R.I.

Makarenkov V., and Leclerc B. 1999 (a). *The fitting of a tree metric to a given dissimilarity with the weighted least squares criterion, Journal of Classification,* 16, pp.3-26.

Makarenkov V., and Leclerc B. 1999 (b). *An optimal way to compare additive trees using circular orders,* accepted for publication in Journal of Computational Biology.

Robinson D.R., and Foulds L.R. 1981. *Comparison of phylogenetic trees,* Mathematical Biosciences 53, 131-147.

Saitou N., and Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Molecular Biology Evolution* 4, 406-425.

Sattath S., and Tversky A. 1977. Additive similarity trees, *Psychometrika* 42, 319-345.