SIMPROF on large matrices

The dendrogram itself is rapidly calculated, at least for the agglomerative methods, since no search procedure is involved, and it can thus be constructed for very large numbers of samples – but the SIMPROF routine is highly compute-intensive, given the typical number of permutations (default 999) and recomputations of the similarities which are necessary for each nodal test (CiMC, Chapter 3), and the potentially large number of nodes. PRIMER 7 now allows the option (the default, which would normally be taken) of dividing these calculations among the multiple processors constituting the core of modern PCs, but it is still unwise to take routinely the SIMPROF option with very large resemblance matrices. A selective form of SIMPROF applied to a single selection of samples, and which provides graphical output of the similarity profile, the spread of alternative profiles obtained under permutations of the data matrix and the null hypothesis distribution for that single test, can be found on the **Analyse>SIMPROF** menu, when the active sheet is a (selection of a) data matrix. A possible strategy for large arrays, which are clearly not going to complete all the nodal tests in a viable time, may then be to carry out the clustering, having turned off the ✓SIMPROF test box on the Cluster dialog, and then manually choose a series of nodes from the dendrogram, testing them one at a time by selecting their samples and carrying out individual **Analyse>SIMPROF** tests, thereby getting some idea of how much structure is potentially interpretable from the dendrogram.

Revision #1
Created 13 June 2024 00:19:22 by Arden
Updated 13 June 2024 00:22:42 by Arden