dendextend: an R package for easier manipulation and visualization of dendrograms

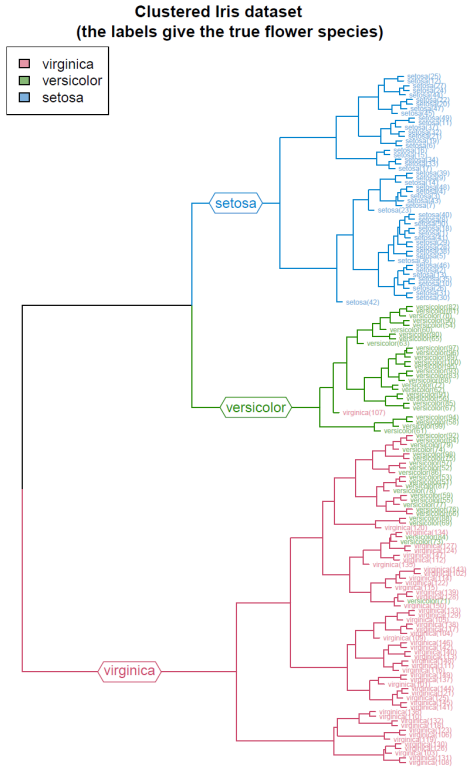
Tal Galili [1,*]

1. Tel Aviv University
*Contact author:Tal.Galili@gmail.com

Keywords: dendrogram, hierarchical clustering, hclust, visualization, tanglegram

In this talk I will introduce the **dendextend** package [1] which extends the palette of functions and methods for the `dendrogram` object.

A dendrogram is a tree diagram which is often used to visualize a hierarchical clustering of items. Dendrograms are used in many disciplines, ranging from Phylogenetic Trees in computational biology to Lexomic Trees in text analysis. Hierarchical clustering in *R* is commonly performed using the `hclust` function. When a more sophisticated visualization is desired, the `hclust` object is often coerced into a `dendrogram` object, which in turn is modified and plotted. While **base** *R* comes with several very useful methods for manipulating the `dendrogram` object (namely: `plot`, `print`, `[[`, `labels`, `as.hclust`, `cophenetic`, `reorder`, `cut`, `merge`, `rev`, and `str`), still - the current palette of functions leaves a lot to be desired.



The novel **dendextend** package offers functions and methods for the `dendrogram` object, allowing for easier manipulation of a dendrogram's shape, color and content through functions such as `rotate`, `prune`, `labels<-`, `labels_colors`, `cutree`, `color_branches`, and more. **dendextend** also provides the tools for comparing the similarity of two dendrograms to one another either graphically using a `tanglegram` plot, or statistically with association measures ranging from `cor_cophenetic` to `Bk_plot`, while enabling bootstrap and permutation tests for comparing the trees.

Since tree structure often requires the use of recursion, which can be slow in *R*, some of the more computationally intensive aspects of the **dendextend** package can be handled with its sister package, **dendextendRcpp** [2], which overrides several basic functions (namely: `cut_lower_fun`, `heights_per_k.dendrogram`, `labels.dendrogram`), with their *C++* implementation.

References

[1] Tal Galili (2014). dendextend: Extending R's dendrogram functionality, http://cran.r-project.org/web/packages/dendextend

[2] Tal Galili (2014). dendextendRcpp: Faster dendrogram manipulation using Rcpp, http://cran.r-project.org/web/packages/dendextendRcpp