

## Reading a Dendrogram

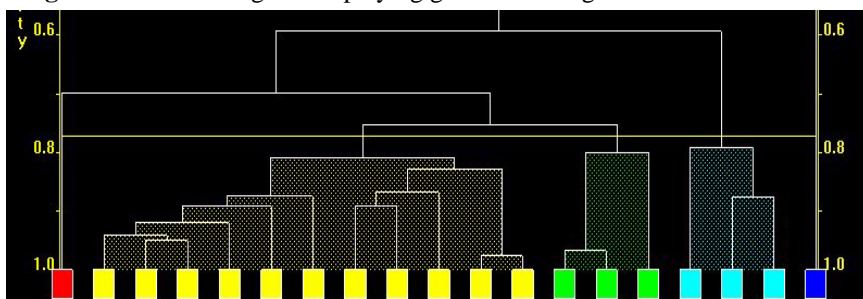
---

Selecting *Dendrograms* from the tab bar running along the top of the screen opens the dendrogram display.

### Basics:

Each box at the bottom of the tree diagram represents a fragment from the database search. The refcode of each is displayed as a label below each box. If the dataset is large it may not be possible to see each box individually and refcodes will not be displayed. Each box has a vertical line coming from it, which is joined to the other vertical lines by horizontal lines called *tiebars*.

**Figure 1.a** - a dendrogram displaying good clustering



There is a *similarity scale* running vertically up the side of the display. This is a measure of how structurally similar any two fragments are, with a value of 1 meaning that the fragments are identical. **The lower a horizontal tie bar connects two fragments, the more similar they are.** Two fragments joined by tie-bars higher up the dendrogram will be less similar.

### Cut-level:

There is a horizontal line that spans the width of the display. This is called the cut-level. It is the cut-off point that determines when fragments are similar enough to be classified together in a group or cluster. Fragments joined by tie-bars below the cut-level are considered to be in the same cluster and are coloured the same to show this. These colours are then used in all of the display screens.

### Good clustering:

A good cluster in the dendrogram should have all the fragment boxes corresponding to a single cluster joined by reasonably low tie bars (having high similarity), while different clusters are separated linked by high tie bars (low similarity). A good indicator of good clustering is when there is a large separation between the tie bars that link fragments within a cluster and the tie bars that link separate clusters (see figure 1.a).

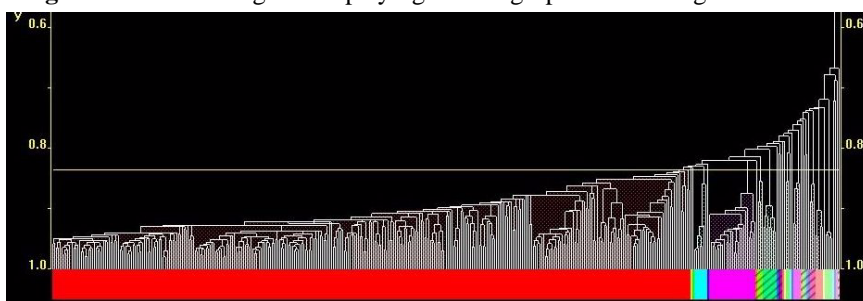
### Bad clustering:

A dendrogram where each fragment is linked to the next by a slightly higher tie bar suggests that there are no clusters in the dataset (see figure 1.b). This is indicative of a single geometry type or of a continuum.

---

A dendrogram where two fragments in the same cluster are joined by a relatively high tie bar may indicate that the analysis should be more discriminate in classifying clusters, and that the cut-level should be lowered to split the cluster into smaller groups.

**Figure 1.b** - a dendrogram displaying chaining - poor clustering



### Manipulating the dendrogram:

The cut-level is calculated for each individual dataset and can be adjusted and experimented with by the user. This is useful as *dSNAP* only defines an initial cut-level according to the results of calculations, while the user can interpret the real underlying structures and decide what would be a sensible and appropriate cut-level based on the level of differences being investigated. There are two ways of doing this:

**1** - Hold down the *Control* key and the left mouse button and drag the mouse up or down to raise or lower the cut-level.

**2** - Select *Set Cut-level to...* from the *Tools* menu. Enter in a numerical value between 0 and 1.0 corresponding to the similarity scale and click *OK*.

The cut-level can be reset to the original calculated value at any time by selecting *Undo Saved Dendrogram Modifications...* from the *Tools* menu and selecting the *Original* setting.

### Example:

In Figure 1.c the fragments in the red cluster are similar and are joined by a low horizontal tie-bar. The fragments in the yellow cluster are also similar to **each other** but are less similar to those in the red cluster as they are joined to the red group by a higher tie-bar, so at this cut-level are in a different cluster. The fragments in the green cluster are even less similar to those in the fragments in the red cluster as they are joined by an even higher horizontal tie-bar (out of shot).

**Figure 1.c** - clusters in a dendrogram

