Subject: Similarity analysis using "find similar compounds..." - slow analysis of libraries
Posted by SM2020 on Tue, 24 Nov 2020 11:35:22 GMT

Hi,

I'm carrying out a similarity analysis by comparing x2 cpd libraries (each ~ 16K) using the "find similar compounds in file..." option using the FragFP descriptor. I am selecting a similarity limit cut-off of ~30% (using the slide bar selector).
I want to obtain the nearest neighbour measure for a given library, so nearest neighbour values can be binned and plotted (note: I realise there are other methods for library comparison as well in DW).

The process is taking an extremely long time on my (admittedly) old computer (4-core/8 threads, i7-3615QM, 16GB Ram, MacOS) - > 24 hrs.

With that in mind, a couple of general questions...

1) What improvements (if any) would help in speeding up this process?
Would increasing the accessible RAM help, or am I generally limited by processor speeds for such analyses?

2) What general hardware upgrades and/or software upgrades would you suggest for dramatically speeding up these types of analyses (<< 24 hrs) in DW?
Faster processor/s with more cores/threads? More RAM? Both...? Other...?

Best Wishes and many thanks.