Subject: defining molecule clusters after similarity analysis
Posted by Xenia on Mon, 07 Mar 2022 17:22:56 GMT
View Forum Message <> Reply to Message

Hi,

I've run a similarity analysis based on the SkelSpheres descriptor on a set of 50k compounds with 80% similarity limit. Is there a way to obtain a list of all the molecules that are connected to each other? I would like to get a list of separate clusters similar to this:

Cluster 1: molecules A-B-C-D-E (each of them are connected because they have >80% similarity)
Cluster 2: molecules X-Y-Z
...

I can see the neighbor identifiers for each molecule (e.g. molecule B is connected to A and C) but I can't figure out a way to obtain this kind of list. In a small set of molecules I can do it by manually selecting the connected molecules in the neighbor tree view but I would like to find a way to obtain a list of clusters automatically.

Thank you for your help and for developing such an amazing software!

Xènia

Subject: Re: defining molecule clusters after similarity analysis
Posted by thomas on Tue, 05 Apr 2022 13:03:11 GMT
View Forum Message <> Reply to Message

Hi Xènia,

you are not the first to ask for that. Thus, I have updated the similarity/activity-cliff-analysis to generate two more columns: a cluster number and the cluster member count. You can download the updated DataWarrior file (dev version) from the link in the fine-print that appears when clicking the 'read and understood' checkbox on the DataWarrior download page.

Hope this solves the issue for you,

Thomas

Subject: Re: defining molecule clusters after similarity analysis
Posted by Xenia on Fri, 29 Apr 2022 16:09:08 GMT
View Forum Message <> Reply to Message

Hi Thomas,

Awesome – thank you so much for including these two columns in the similarity/activity-cliff

analysis.

It works perfectly and it's saving me a lot of time!

Xènia