Subject: t-SNE
Posted by sansun on Fri, 15 Jul 2022 07:48:38 GMT
View Forum Message <> Reply to Message

Hi

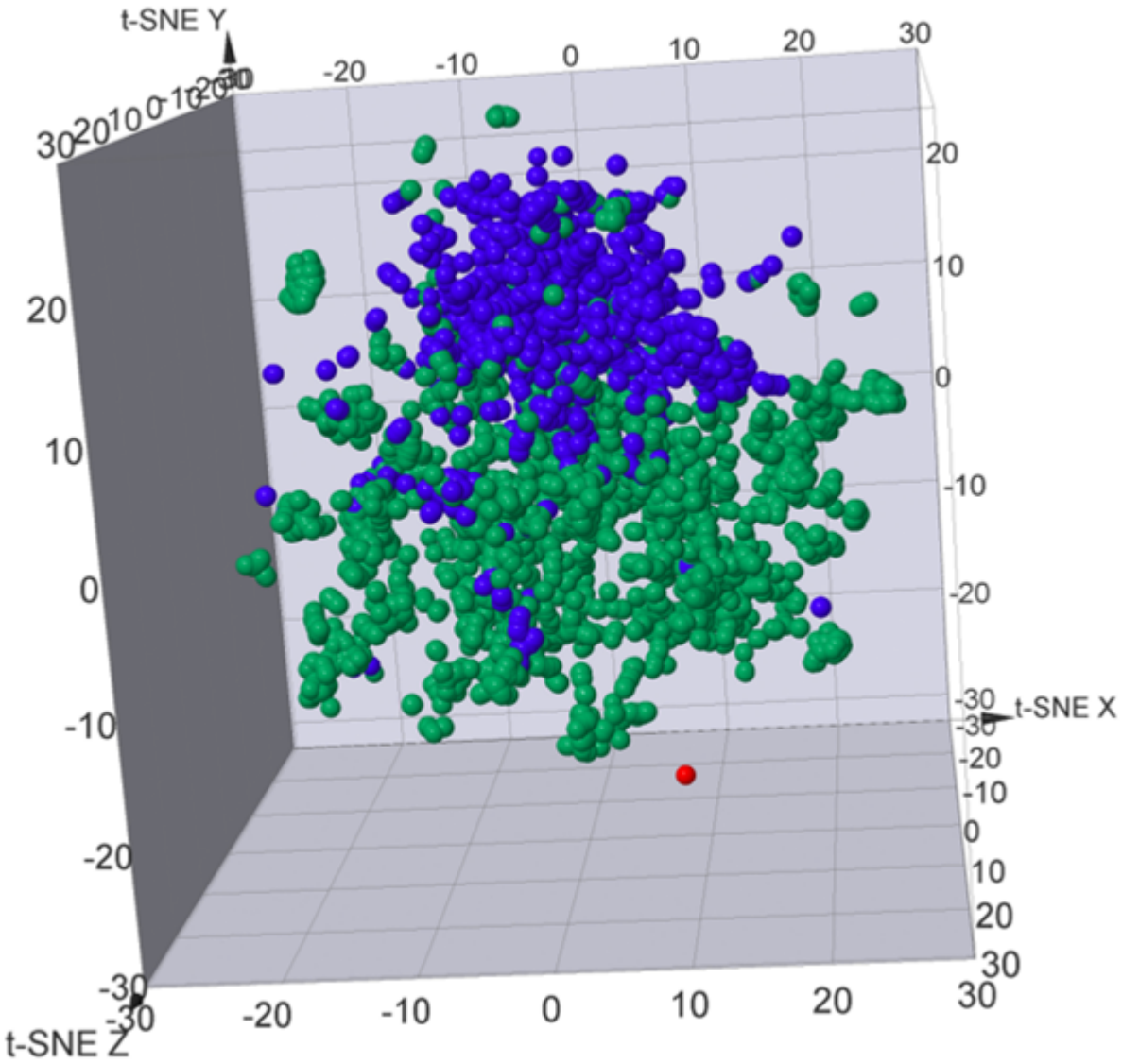
I have generated t-SNE 3D plot to compare two libraries (green and blue) using SkelSphere descriptor.
The molecules in the two libraries are kind of making separate clusters. Can we say the chemical space of the two libraries do not overlap considerably or they contain different type of structures?

If yes, how can we represent this conclusion quantitatively using t-SNE?

Thanks
SS

File Attachments

1) tSNE.png, downloaded 1800 times



Subject: Re: t-SNE
Posted by nbehrnd on Tue, 19 Jul 2022 06:12:38 GMT
View Forum Message <> Reply to Message

Sansun,

in a first approximation, you could determine the centres of gravity for each cluster, and from there the spread along (the projections of) x, y, z. You then could build a tetragonal parallelepiped of known dimension for each cluster and check if the volume one of these boxes is to some part enclosed by the box about the other cluster, if one cluster's box is totally enclosed by the box of the other (i.e., a sub-set of the space), or if the two are separate from each other.

The likely better approach were to perform a principal component analysis (PCA) to determine

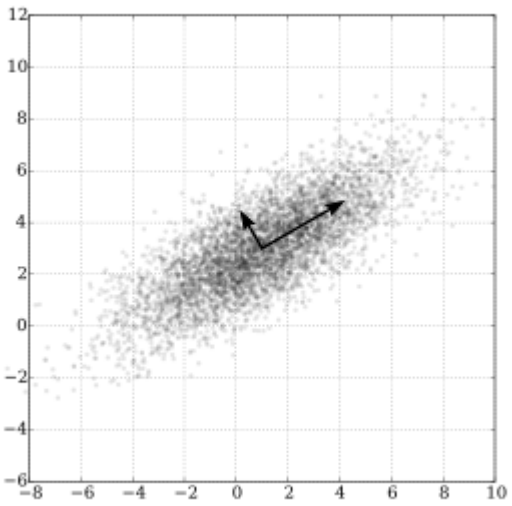
centre of gravity of each cluster, and to determine the corresponding eigenvectors. This would offer the advantage that the orientation of the vectors to construct an envelope of the cluster no longer is the projection along x, or y, or z, but that these vectors may have any orientation in space. (If you enter a 3D cluster each, I would constrain the PCA to only consider three dimensions.) PCA includes some vector transformation to centre (and normalize) the scattered clouds in first place; I do not recall if DW already is set up to perform these operations with a function from the GUI.

Norwid

image credit: https://en.wikipedia.org/wiki/Principal_component_analysis

File Attachments

1) GaussianScatterPCA.svg.png, downloaded 1679 times



Subject: Re: t-SNE
Posted by sansun on Tue, 19 Jul 2022 07:35:30 GMT
View Forum Message <> Reply to Message

Hi Norwid

Thanks for the detailed response. I'll try that. I don't think it can be performed in DW though.
