CH5650: Molecular Data Science and Informatics
Quiz (Maximum Marks – 15)

Start Time: 8 AM, March 28, 2023; Exam End Time: 8 PM, March 28, 2023

Data: Molecular dynamics trajectory of a molecule in an infinitely dilute solution condition. The data file consists of ~10000 snapshots of a molecule at different times. Each frame consists of the cartesian coordinates and velocities (ux, uy, uz, vx, vy, vz) of all the constituting atoms of the molecule. All the quantities are in reduced unit. Each snapshot in the file consists of 9 comment lines followed by 100 lines, which are the position and velocity coordinates of 100 atoms. Also, following are the descriptions of the attributes, which are there in the data file.

id: Atom id. It varies from 1 to 100, as there are hundred atoms in the system.
mol: Molecule id. It is always 1, as the file contains one molecule.
type: Type of atom. It is always 1, as all the atoms in the molecule are same type.

Problem Statement: The molecular trajectory consists of multiple phases/structures. Apply the concept or dimensionality reduction and clustering to classify the phases. Based on your analysis, report the number of phases present in the trajectory. Draw a lower dimensional representation of the data to show all possible phases.