Unsupervised Graph Neural Network Reveals the Structure–Dynamics Correlation in Disordered Systems

Vaibhav Bihani

Deparment of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India 110016

Sahil Manchanda

Department of Computer Science and Engineering,
Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India 110016

Sayan Ranu*

Department of Computer Science and Engineering,
Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India 110016 and
Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India 110016

N. M. Anoop Krishnan†

Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India 110016 and
Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India 110016

(Dated: June 28, 2022)

Learning the structure–dynamics correlation in disordered systems is a long-standing problem. Here, we use unsupervised machine learning employing graph neural networks (GNN) to investigate the local structures in disordered systems. We test our approach on 2D binary $A_{65}B_{35}$ LJ glasses and extract structures corresponding to liquid, supercooled and glassy states at different cooling rates. The neighborhood representation of atoms learned by a GNN in an unsupervised fashion, when clustered, reveal local structures with varying potential energies. These clusters exhibit dynamical heterogeneity in the structure in congruence with their local energy landscape. Altogether, the present study shows that unsupervised graph embedding can reveal the structure–dynamics correlation in disordered structures.

I.   INTRODUCTION

Disordered structures exist in distinct phases such as liquid, supercooled and glassy states [1, 2]. Liquids, when cooled fast enough to avoid crystallization, form glasses [2]. While liquids and supercooled liquids are equilibrium states, glasses are considered as metastable state, which continuously relaxes [2]. The cooling pathway, such as the rate and the pressure, directly affects the glass structure and dynamics [3–8]. Thus, although glasses can be considered as "frozen liquids", the dynamics of glasses are notably different from that of liquids. However, it is still debated whether the dynamics of these disordered systems are encoded precisely in their local structure.

Conventional approaches on understanding the structure-dynamics relationships rely on identifying structural signatures such as medium range crystalline orders (MRCO) that relate to the dynamical heterogeneity in disordered systems [9–11]. However, attempts to predict the dynamical heterogeniety directly from the structure have been barely successful, although indirect correlations have been established [12–17]. Recently, data-driven approaches employing machine learning (ML) was used to predict atomic dynamics directly from the structure. Specifically, "softness" [18, 19], a machine learned parameter, was introduced to predict the tendency of atoms to rearrange purely based on their local structure in a disordered system. Further, Bapst et al. [20] used a graph neural network (GNN) to learn the structural descriptors capable of predicting propensity [15, 21] of atoms to move under shear stress. Although both these works [18, 20] confirmed that the dynamical heterogeneity could be predicted directly from the structure, they employed a supervised ML approach [22, 23], where the models were trained explicitly to learn the descriptors that predict the dynamics of atoms. In other words, the data on both structure and dynamics was used to learn their interrelationship through supervised ML, which was then used to model unseen systems. This infuses an inherent bias to learn the features that best represent the individual dynamics of atoms. Hence, the descriptors may not be suitable for other tasks such as predicting the collective dynamics such as self-organization or even learning the local motifs in glass structures [24, 25]. Further, training the model require a priori knowledge of both the structure and the dynamics, which may not be available for many systems.

---

* sayanranu@iitd.ac.in
† krishnan@iitd.ac.in

Here, using GNNs, we learn the structural features of disordered systems directly in an unsupervised fashion [26]. Specifically, by considering disordered structures as graphs [27], we allow the model to learn local atomic environment in an unsupervised fashion. We show that the GNN is able to identify the features that, in turn, discovers local clusters in the disordered structures reminiscent of the MRCO. We test our approach on a well established 2D binary LJ glass of composition $A_{65}B_{35}$ to form glasses from the melt at different cooling rates ranging over four orders of magnitude. The clusters discovered by the GNN exhibit a good correlation in terms of their dynamics and energy, revealing the existence of a structure–dynamics correlation.

II. METHODOLOGY

To learn the local topology of disordered systems, we use GNNs. Figs. 1(a)-(b) describe the transformation from disordered structure to a graph. First, we transform each of the disordered structures as an undirected graph [27] $G = (V, E)$. The node set V and edge set E represent the atoms and the set of bonds between these atoms, respectively. Here, the bonds of the atoms are with their first neighbors with the cutoff distance as the first minima of the partial pair distribution functions (PDFs). Thus, $\mathcal{N}(v)$ represents the set of nodes that are neighbors to the node v in graph G. Further, each node $v \in V$ consists of the following node attributes (h_v^0): **(1)** one-hot encoding representing the atom (node) type (*i.e.*, type **A** or **B**). **(2)** Coordinates of the node v in two dimensions (x_v, y_v).

Our goal is to understand the local structure of such disordered systems. Towards this end, we utilize GRAPH-SAGE [26] to learn the node representation in a way that captures topological structure of each node's neighborhood (see Fig. 1(c),(d)). Hence, for each node $v \in V$, we learn how to aggregate information from its neighboring nodes $\mathcal{N}(v)$ as

$$\mathbf{h}_v^k \leftarrow F\left(\mathbf{W^k} \cdot \text{MEAN}\left(\left\{\mathbf{h}_v^{k-1}\right\} \cup \left\{\mathbf{h}_u^{k-1}, \forall u \in \mathcal{N}(v)\right\}\right)\right) \tag{1}$$

where h_v^0 represents the raw feature information of nodes comprising of the atom type and coordinates, h_v^k represents the node information after $k-$rounds of message passing with the neighbors, and F represents an activation function which in our case is the sigmoid function.

The purpose of the GNN is to construct representations for each node integrating both the nodes' initial raw feature information and the information about the local graph structure that surrounds them. Specifically, each node aggregates its own information and information from its $k-hop$ neighborhood to update its representation, i.e., $h_v^1...h_v^k$ where k denotes the depth of the GNN and h_v^k denotes the final representation of the node v. The term $\mathbf{W}^k$ is a learnable parameter of the GNN at depth k. In the case of a disordered atomic structure, $\mathbf{W}^k$ learns to aggregate feature information from neighboring atoms and pass this information to generate an improved representation for the central atom v incorporating the local neighborhood features. With a higher depth k, nodes incrementally gain more and more information from further reaches of the graph. For instance, $k = m$ aggregates information from the mth hop neighborhood of the node v. However, higher values of k in GNNs, such as $k > 3$ could lead to poor performance [28].

Next, we train the GNN by minimizing an objective function. Since our aim is to learn the local features of the structure without any additional information, we resort to unsupervised learning. In order to train the model, we generate a fixed-length (L) random walk [26, 29] from each node v in the graph, which visits different nodes in v's neighborhood. This random walk allows the exploration of the local structure surrounding node v. Fig. 1(c) depicts 2 random walks of length 3 from a node v. Note that in contrast to previous approaches, here we are not forcing the GNN to learn any task-specific embedding such as predicting the dynamics. The embeddings are purely a representation of the local neighborhood of a node and hence is purely structural in origin. In other words, the representation learned by GNN is purely based on the atomic structure and has no bias toward their dynamics.

Through these random walks, we identify nodes $u \in V$ that co-occur in v's neighborhood. Without loss of generality, we consider these node pairs (v, u) as *positive samples*, that is representing the neighborhood of v. Node pairs (v, u_n) that do not co-occur in random walks are considered as *negative samples*, which is representative of a random node absent from v's neighborhood. Next, our unsupervised objective function given by Eq. 2 is optimized to learn the parameters of the Graph neural network.

$$\mathcal{L}\left(z_u\right) = -\log\left(\sigma\left(z_v^{\ T} z_u\right)\right) - Q \cdot E_{u_n \sim p_n(u)} \log\left(\sigma\left(z_v^{\ T} z_{u_n}\right)\right). \tag{2}$$

Here, $z_v = h_v^k$, where k is the number of layers of the GNN and Q determines the number of negative samples from the negative distribution $p_n(u)$ (consisting negative samples). The minimization of the objective function encourages nodes that are in same neighborhood (*positive samples*) to have similar representation and vice-versa. As evident from Eq. 2, the dot-product (hence, similarity) of representations of positive node pairs (v, u) that co-occur in random walks is encouraged to be higher, while the opposite is true for the representation of negative node pairs (v, u_n).
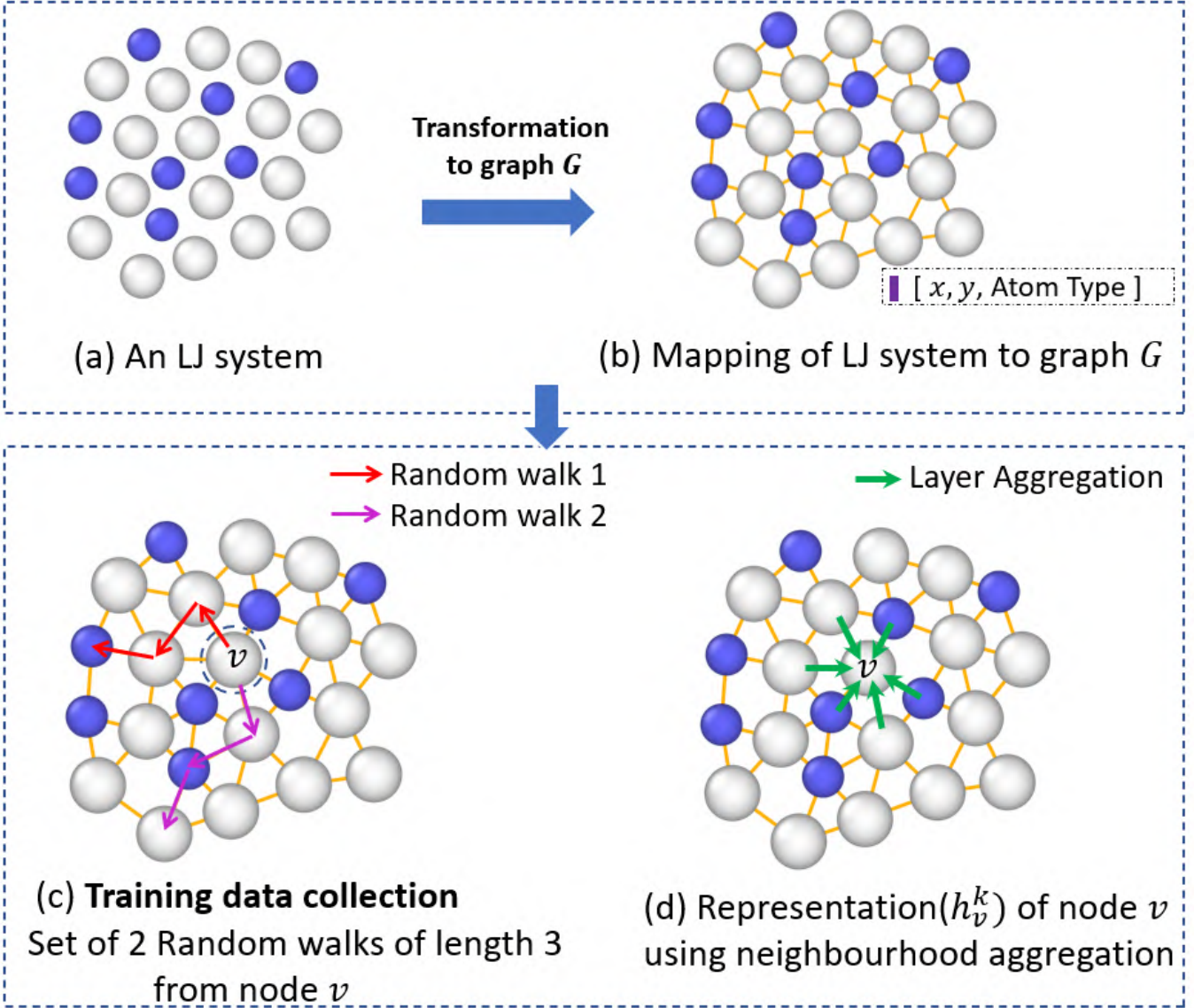
FIG. 1: Proposed framework for learning the structural representation of disordered systems.

III.   RESULTS AND DISCUSSION

We test our approach on a 2D binary LJ system. We perform MD simulation of melt-quenching LJ system [30] of 500 atoms with composition $A_{65}B_{35}$ to prepare glasses at different cooling rates. All the simulations are carried out using the LAMMPS package [31]. The interaction between the particles is governed by

$$V_{\mathrm{LJ}}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right] \tag{3}$$

where r refers to the distance between two particles, σ is the distance at which inter-particle potential energy is zero and ε refers to the depth of the potential well. Here, we use the LJ parameters $\varepsilon_{AA} = 1.0$, $\varepsilon_{AB} = 1.5$, $\varepsilon_{BB} = 0.5$, $\sigma_{AA} = 1.0$, $\sigma_{AB} = 0.8$ and $\sigma_{BB} = 0.88$. The mass for all particles is set to 1.0. All the quantities are expressed in reduced units with respect to σ_{AA}, ε_{AA}, and Boltzmann constant k_B. We set the interaction cutoff $r_c = 2.5\sigma$ [32] and the time step $dt = 0.003$ for simulations.

We perform all the simulations at constant volume. For preparing the glass structure, first the ensemble is equilibrated at high temperature $T = 2.0$ where it equilibrates in the liquid state. Next, it is cooled down to low temperature of $T = 0.05$, at 4 different cooling rates ranging from 3.33×10^{-3} to 3.33×10^{-6}. The cooling is carried out at NVT,

using Nose-Hoover thermostat in temperature steps of 0.05. At each of these steps, the atomic structure of the system along with their density and enthalpy is noted. In order to avoid the noise due to thermal fluctuations, the glass structures are minimized at 0 atm pressure to obtain the ground state enthalpy and density [32]. Accordingly, we obtain 160 minimized glass structures, 40 from each cooling rate, representing the ground state structures at different temperatures and states.

To train the GNN, we set $k = 2$, representing the information aggregation from up to the 2nd neighbors. We observe that higher values of k could accumulate noise, which is analogous to the PDF in disordered systems which at larger distances saturates to 1. The choice of $k = 2$ also resulted in learning better representations as evidenced by the performance on the validation set as discussed in Fig. 10 in section S4 in the Appendix. For performing random walk, we set $L = 4$ and $Q = 10$ and number of random walks starting from each node to be 10. Effect of random walk length parameter is shown in section S7 of Appendix. We set the size of node representation to be 64. Finally, we use the training graphs to optimize the parameters of GNN. Using the learnt parameters of GNN, we generate node representations for graphs that were not seen during the training phase. The parameters used for training GNN are specified in detail in section S4 in Appendix.
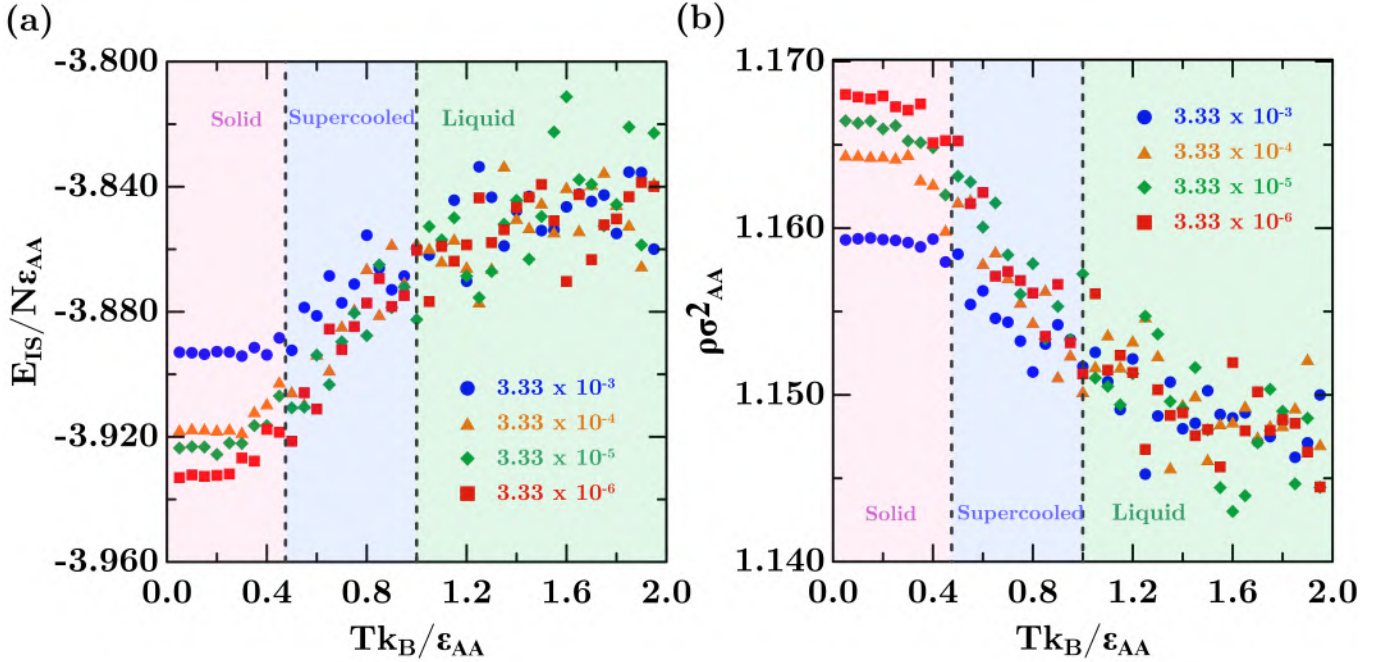

FIG. 2: Variation in **(a)** potential energy, and **(b)** density with respect to temperature for different cooling rates of the binary LJ system. The liquid, supercooled liquid, and the solid (glassy) regions are highlighted.

First, we analyse the change in the enthalpy and density of the disordered structures, when they are cooled at different rates. Figs. 2(a) and (b) represent the variation in enthalpy and density of the structures upon quenching. In agreement with theoretical understanding [4, 5, 8, 32], we observe that lower cooling rates lead to more stable glasses with lower enthalpy and higher densities. We also observe that the density and enthalpy of the structures in the liquid ($T > 1.0$) and supercooled ($0.6 < T < 1.0$) regimes are indistinguishable, despite the difference in their cooling rates—thanks to their equilibrium nature [2]. In contrast, the solid (glassy, $T < \sim 0.6$) state exhibits clear variation for the structures obtained with different cooling rates. Note that the enthalpy of a system have major contributions from the short-range interactions in a structure, while the density could be attributed to the medium- and long-range order of the system. Accordingly, we observe that the disordered structures simulated have notable differences both in terms of their short-range interactions and the long-range order.

Next, in Fig. 3 we compare the total and partial PDFs of the structures obtained at different temperatures and cooling rates. Since liquid and supercooled structures does not show any significant difference of cooling rate (see Fig. 2), only one PDF is plotted representing them. We observe that despite the denser and more stable structure (in terms of enthalpy) of glass obtained at lower cooling rate, the total PDFs obtained are almost identical. Except a minor broadening of the first peak of B-B partial PDFs, the partial PDFs of different glassy structures are also comparable. This results shows that merely distance distribution cannot capture the effect of cooling rate on the local structure. Note that extensive studies have been carried out to understand the local structure of LJ glasses
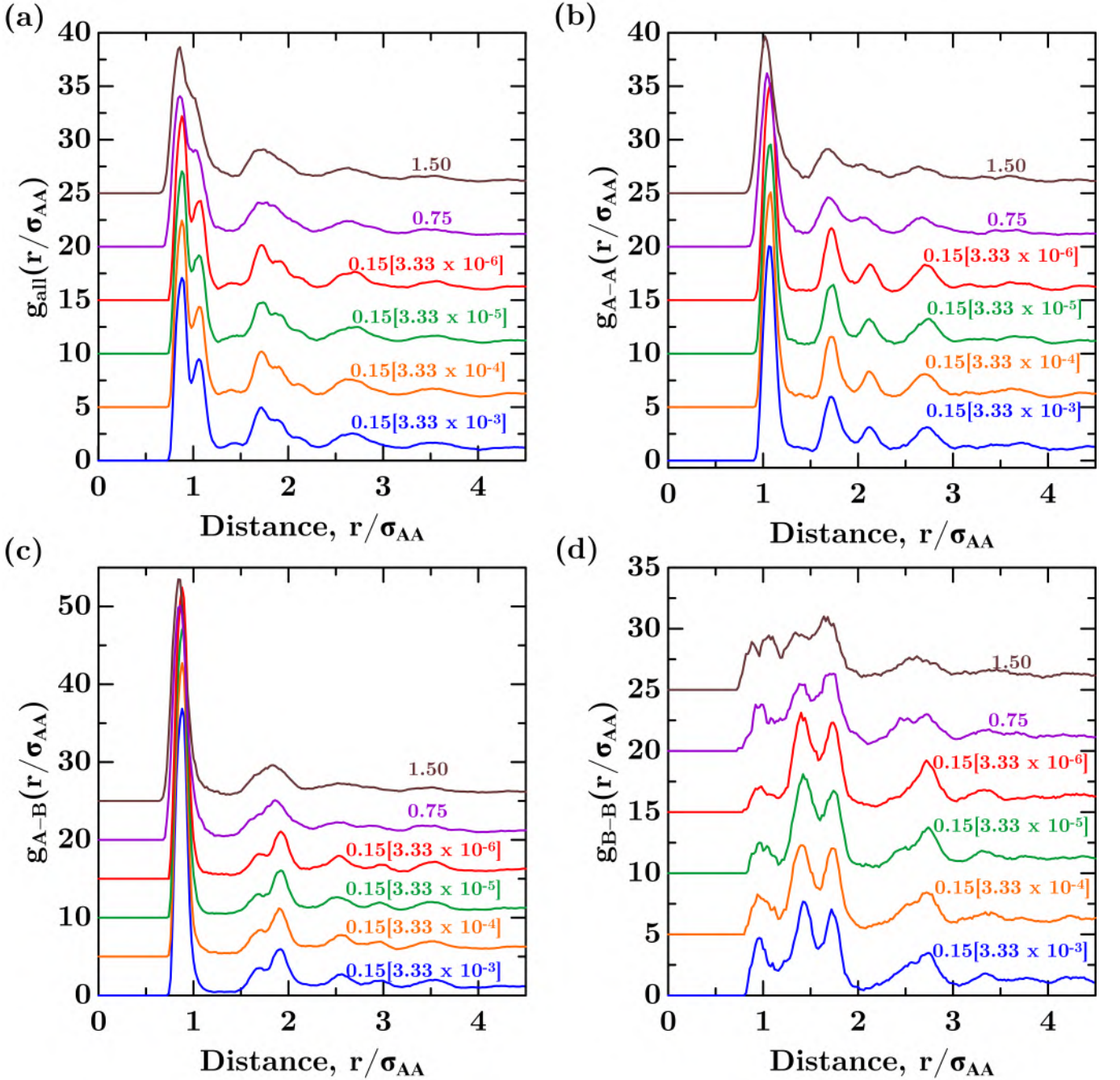
FIG. 3: ((a) Total, (b) A-A, (c) A-B, and (d) B-B partial pair distribution functions for the binary LJ system in the liquid, supercooled liquid and glassy states. An offset in the y-direction is provided for different systems to improve clarity in representation.

with different cooling rates and readers are referred to these works for further differences and similarities on the local structure with respect to cooling rate [5].

Towards our goal of gaining better understanding of the different local regions in the disordered structure (graph), we use the unsupervised graph embeddings learned by the GNN. Specifically, we cluster the nodes of a graph based on their node representations using a density-based clustering algorithm OPTICS [33] and cosine similarity as our distance function between two nodes.

Specifically, nodes (atoms) in the graph are ordered such that spatially closest points become neighbors in the ordering. Then based upon a distance parameter, representing the density, a decision is made to accept/reject points

to be part of a cluster. Details of the OPTICS clustering algorithms and hyper-parameters used are given in section S2,S3 and S5 in Appendix.
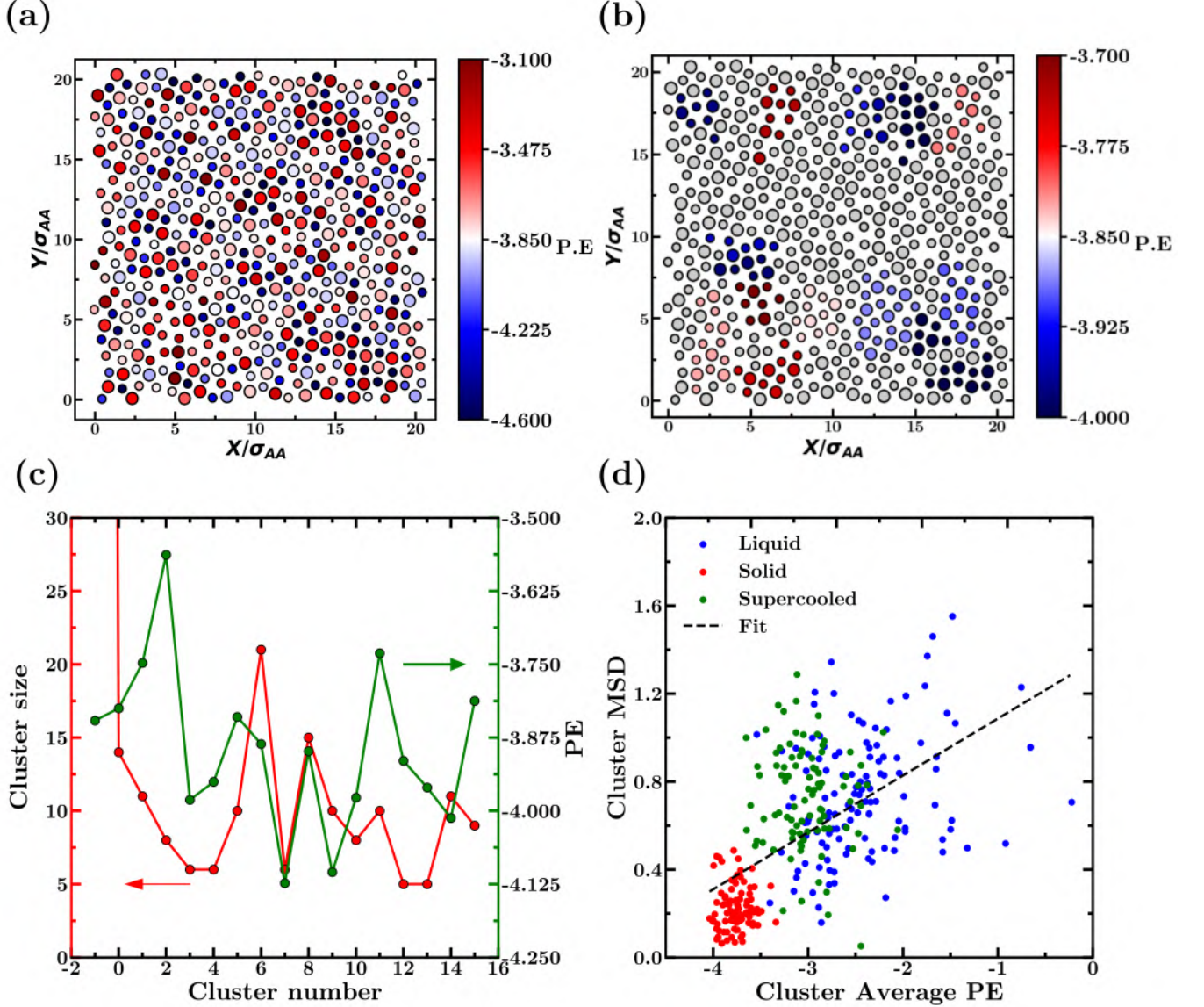

FIG. 4: (a) Potential energy per atom, (b) average potential energy of the clusters obtained. (c) Cluster size and potential energy of each obtained cluster. (d) Average potential energy of the cluster with respect to the MSD.

Now, we analyze the LJ system when atoms are clustered based upon the approach described above. Fig. 4(a) shows the potential energy per atom for an arbitrary glass structure from the simulations. We observe that the energy distribution seems to be random with no local order. In contrast, Fig. 4(b) shows the average potential energy of the clusters obtained from OPTICS clustering of node representations generated by the trained GNN. Note that nodes colored in grey are not part of any cluster. i.e., they are outliers. Additionally, we also show the reachability distance plot generated by OPTICS in section S6 in Appendix. Fig. 4(c) shows the cluster size and average potential energy of the clusters. We clearly observe several high and low energy clusters in the structure. It can be observed from the limits of color bar in Fig. 4(b) that the variation in average potential energies of the obtained clusters, although small in comparison to that of individual atoms, is of higher range than that observed in the cooling curves of Fig. 2. Interestingly, the local clusters observed are reminiscent of the MRCO observed in supercooled and glassy systems based on orientational order parameters [9–11]. This suggests that the unsupervised embeddings learned by the GNN is able to capture the local structural order (and disorder) in these systems.

In order to analyze whether the learnt cluster correlates to the dynamical heterogenity observed in these systems,

we study the curvature of the local energy landscape of the clusters following an established methodology [2, 34]. First, we perform multiple energy minimizations to ensure that the samples are at their ground state. Then, we provide a sudden energy bump to the system corresponding to a temperature of $T = 0.25$. The temperature is chosen to be high enough to allow potential motion between low-energy barriers but low enough to avoid glass transition or melting of the system. Finally, we allow the system to evolve in the micro-canonical ensemble (NVE) for 100 thousand steps. The MSD $(r^2)_C$ of a cluster C is obtained as

$$(r^2)_C = \left[\frac{1}{N_C} \sum_{i=1}^{N_C} (\vec{r_i} - \vec{r_{i,0}})^2 \right] \tag{4}$$

where N_C represents the number of atoms in a cluster, r_i are the positions of each atom after 100 thousand steps of dynamics following the energy bump, and $r_{i,0}$ are the positions in the inherent structure.

Stable low-energy structures represented by narrow and deep local energy landscape exhibit low MSD, while high energy structures, exhibiting a shallow and broad landscape, exhibit high MSD. To understand the correlation between the clusters identified by GNN and dynamical heterogeniety, we plot the MSD of the clusters with respect to the average potential energy of the clusters (see Fig. 4(d)). Note that each data point in the plot corresponds to a cluster, and the color represent the state (liquid, supercooled, solid) from which it was obtained. We observe a direct positive correlation with the cluster potential energy and the MSD with a Pearson coefficient of 0.747, representing high statistical significance. The variations in the cluster MSD even within a given disordered structure (see Fig. 5 in section S1 of Appendix), in congruence with the local potential energy, suggests the existence of dynamical heterogeneity in these systems. These heterogenieties are captured effectively by the graph embeddings.

IV. CONCLUSION

Altogether, present study reveals a novel method, employing unsupervised GNN, for learning the local structure in disordered systems. We demonstrate that these local structures exhibit a direct correlation with the dynamical heterogenieties present in the disordered systems. It is worth noting that in contrast to previous approaches, where ML was trained against a data to predict dynamics from structure, the present approach reveal local structural motifs in an unbiased manner purely from the structure. Further analysis of these clusters could potentially reveal hidden order within a disordered structure, that, in turn, govern their dynamics. In addition, the learned embedding could further be potentially used as a representation or a reduced-order feature of the disordered system to predict their properties. Our approach thus paves way for gaining an understanding on the structure–dynamics correlation in disordered systems that are not achieved by conventional methods.

The code is available at : `https://github.com/M3RG-IITD/ML-DOS`

[1] C. A. Angell, Formation of glasses from liquids and biopolymers, Science **267**, 1924 (1995).
[2] P. G. Debenedetti and F. H. Stillinger, Supercooled liquids and the glass transition, Nature **410**, 259 (2001).
[3] K. Vollmayr, W. Kob, and K. Binder, Cooling-rate effects in amorphous silica: A computer-simulation study, Phys. Rev. B **54**, 15808 (1996).
[4] X. Li, W. Song, K. Yang, N. A. Krishnan, B. Wang, M. M. Smedskjaer, J. C. Mauro, G. Sant, M. Balonis, and M. Bauchy, Cooling rate effects in sodium silicate glasses: Bridging the gap between molecular dynamics simulations and experiments, The Journal of chemical physics **147**, 074501 (2017).
[5] K. Vollmayr, W. Kob, and K. Binder, How do the properties of a glass depend on the cooling rate? a computer simulation study of a lennard-jones system, The Journal of chemical physics **105**, 4714 (1996).
[6] C. T. Moynihan, A. J. Easteal, M. A. De Bolt, and J. Tucker, Dependence of the fictive temperature of glass on cooling rate, Journal of the American Ceramic Society **59**, 12 (1976).
[7] M. M. Smedskjaer, R. E. Youngman, S. Striepe, M. Potuzak, U. Bauer, J. Deubener, H. Behrens, J. C. Mauro, and Y. Yue, Irreversibility of pressure induced boron speciation change in glass, Scientific reports **4**, 1 (2014).
[8] P. Bhaskar, R. Kumar, Y. Maurya, R. Ravinder, A. R. Allu, S. Das, N. N. Gosvami, R. E. Youngman, M. S. Bødker, N. Mascaraque, *et al.*, Cooling rate effects on the structure of 45s5 bioglass: Insights from experiments and simulations, Journal of Non-Crystalline Solids **534**, 119952 (2020).
[9] K. Watanabe and H. Tanaka, Direct observation of medium-range crystalline order in granular liquids near the glass transition, Physical review letters **100**, 158002 (2008).
[10] T. Kawasaki, T. Araki, and H. Tanaka, Correlation between dynamic heterogeneity and medium-range order in two-dimensional glass-forming liquids, Physical review letters **99**, 215701 (2007).

[11] I. Tah, S. Sengupta, S. Sastry, C. Dasgupta, and S. Karmakar, Glass transition in supercooled liquids with medium-range crystalline order, Physical review letters **121**, 085703 (2018).

[12] D. Coslovich and G. Pastore, Are there localized saddles behind the heterogeneous dynamics of supercooled liquids?, EPL (Europhysics Letters) **75**, 784 (2006).

[13] A. Widmer-Cooper, P. Harrowell, and H. Fynewever, How reproducible are dynamic heterogeneities in a supercooled liquid?, Physical review letters **93**, 135701 (2004).

[14] A. Widmer-Cooper and P. Harrowell, Predicting the long-time dynamic heterogeneity in a supercooled liquid on the basis of short-time heterogeneities, Physical review letters **96**, 185701 (2006).

[15] A. Widmer-Cooper and P. Harrowell, On the study of collective dynamics in supercooled liquids through the statistics of the isoconfigurational ensemble, The Journal of chemical physics **126**, 154503 (2007).

[16] E. D. Cubuk, S. S. Schoenholz, J. M. Rieser, B. D. Malone, J. Rottler, D. J. Durian, E. Kaxiras, and A. J. Liu, Identifying structural flow defects in disordered solids using machine-learning methods, Physical review letters **114**, 108001 (2015).

[17] A. Malins, J. Eggers, H. Tanaka, and C. P. Royall, Lifetimes and lengthscales of structural motifs in a model glassformer, Faraday discussions **167**, 405 (2013).

[18] S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras, and A. J. Liu, A structural approach to relaxation in glassy liquids, Nature Physics **12**, 469 (2016).

[19] S. S. Schoenholz, E. D. Cubuk, E. Kaxiras, and A. J. Liu, Relationship between local structure and relaxation in out-of-equilibrium glassy systems, Proceedings of the National Academy of Sciences **114**, 263 (2017), https://www.pnas.org/doi/pdf/10.1073/pnas.1610204114.

[20] V. Bapst, T. Keck, A. Grabska-Barwińska, C. Donner, E. D. Cubuk, S. S. Schoenholz, A. Obika, A. W. Nelson, T. Back, D. Hassabis, *et al.*, Unveiling the predictive power of static structure in glassy systems, Nature Physics **16**, 448 (2020).

[21] L. Berthier and R. L. Jack, Structure and dynamics of glass formers: Predictability at large length scales, Physical Review E **76**, 041509 (2007).

[22] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, nature **521**, 436 (2015).

[23] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Reviews of Modern Physics **91**, 045002 (2019).

[24] E. B. Jones and V. Stevanović, The glassy solid as a statistical ensemble of crystalline microstates, npj Computational Materials **6**, 1 (2020).

[25] P. Boolchand, G. Lucovsky, J. Phillips, and M. Thorpe, Self-organization and the physics of glassy networks, Philosophical Magazine **85**, 3823 (2005).

[26] W. Hamilton, Z. Ying, and J. Leskovec, Inductive representation learning on large graphs, Advances in neural information processing systems **30** (2017).

[27] R. J. Trudeau, *Introduction to graph theory* (Courier Corporation, 2013).

[28] M. Liu, H. Gao, and S. Ji, Towards deeper graph neural networks, in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (2020) pp. 338–348.

[29] D. Aldous and J. Fill, Reversible markov chains and random walks on graphs (1995).

[30] R. Brüning, D. A. St-Onge, S. Patterson, and W. Kob, Glass transitions in one-, two-, three-, and four-dimensional binary lennard-jones systems, Journal of Physics: Condensed Matter **21**, 035117 (2008).

[31] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, Comp. Phys. Comm. **271**, 108171 (2022).

[32] S. Singh, M. D. Ediger, and J. J. De Pablo, Ultrastable glasses from in silico vapour deposition, Nature materials **12**, 139 (2013).

[33] M. Ankerst, M. Breunig, H. Kriegel, R. Ng, and J. Sander, Ordering points to identify the clustering structure, in *Proc. ACM SIGMOD*, Vol. 99 (2008).

[34] N. M. A. Krishnan, B. Wang, Y. Yu, Y. Le Pape, G. Sant, and M. Bauchy, Enthalpy landscape dictates the irradiation-induced disordering of quartz, Phys. Rev. X **7**, 031019 (2017).

[35] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, A density-based algorithm for discovering clusters in large spatial databases with noise., in *kdd*, Vol. 96 (1996) pp. 226–231.

[36] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, Optics: Ordering points to identify the clustering structure, ACM Sigmod record **28**, 49 (1999).

[37] OPTICS Clustering Scikit-Learn, `https://scikit-learn.org/stable/modules/generated/sklearn.cluster.OPTICS.html/`, [Online; accessed 07-June-2022].

Supplementary Material

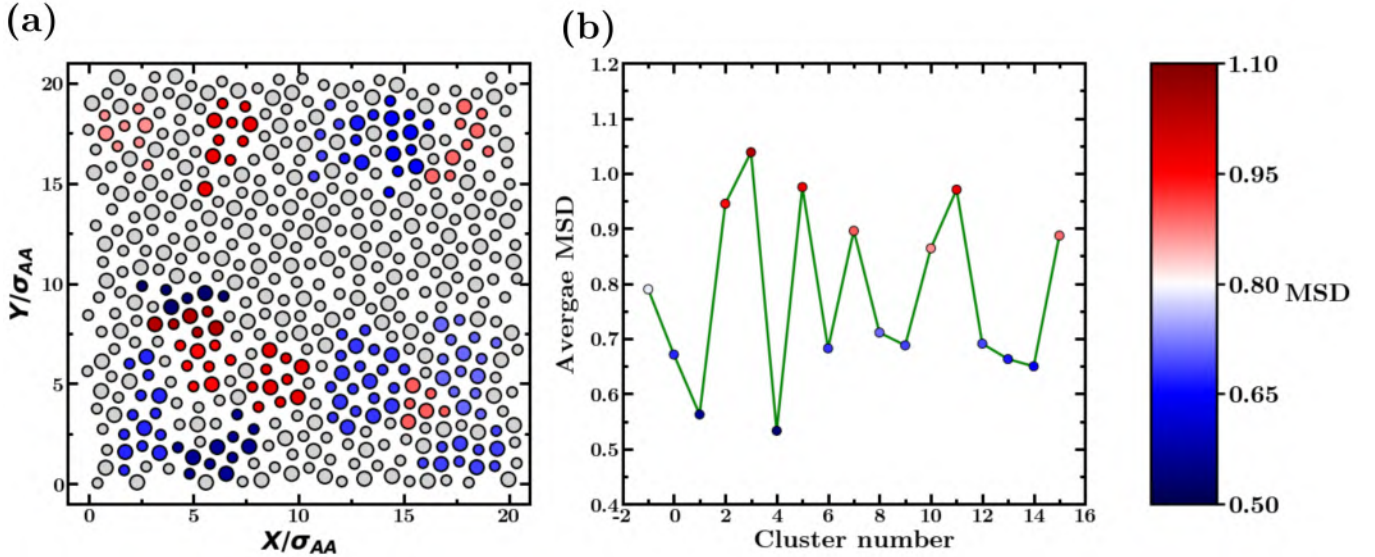## S1. CLUSTER MEAN SQUARED DISPLACEMENT IN 2D LJ SYSTEM

FIG. 5: (a)Cluster mean squared displacement in 2D LJ system, (b) Cluster mean squared displacement vs Cluster number

In Fig. 5(a) and (b) we show the obtained mean squared displacement of clusters in our 2D LJ system. Clustering the graph embeddings thus captures the existence of dynamic heterogeneity in these disordered systems.

S2. DENSITY BASED CLUSTERING:

Density based clustering identifies groups/clusters in data based upon the intuition that each cluster is a contiguous region of high point density, separated from other clusters by regions of low point density. Fig. 6 shows an example of density based clustering obtained using the DBSCAN algorithm [35]. We can clearly see that DBSCAN generated clusters where density of points is high and is separated by regions of low density of points. Density based clustering can generate clusters of any shape, which has advantage of other methods such as the popular *centroid-based* technique *K-means*, which generates clusters that are more or less spherical.

However, DBSCAN fails to identify clusters when data has clusters of varying density. As an example, let us consider Fig. 7(a), which consists clusters of varying densities. DBSCAN fails to identify one of the clusters. The prime reason responsible for this behavior is that DBSCAN uses a fixed cutoff distance between two points to consider them as neighbors. However, a *larger* distance in one cluster could be a *shorter* distance in another cluster as can be seen in Fig. 7(a). Further, a solution that uses larger cutoff could lead to merging of multiple clusters and a lower cutoff could lead to missing out points to be part of a cluster as in Fig. 7(a). Hence, finding the optimal cutoff is non-trivial.

To overcome this limitation, the OPTICS (Ordering Points To Identify the Clustering Structure) [36] algorithm was proposed. Specifically, instead of relying on a fixed cutoff distance, points are ordered linearly such that spatially closest points become neighbors in the ordering. This ordering is then utilized to perform clustering. As we can see in Fig. 7(b), using OPTICS, we are able to identify regions having varying densities.

S3. OPTICS CLUSTERING ALGORITHM

As discussed above, OPTICS [36] is a density-based clustering algorithm, like DBSCAN, with the additional advantage of being able to identify clusters of varying densities. We next introduce few definitions for better understanding of the algorithm. Please refer to Fig. 8a for a visual depiction.
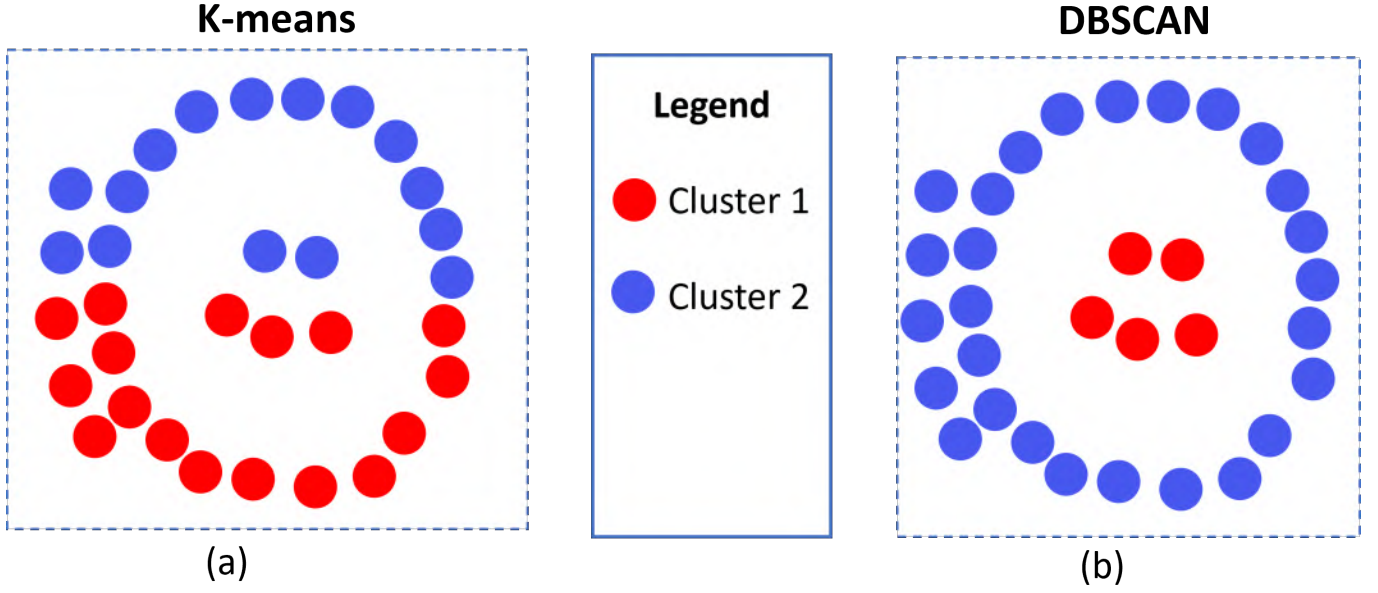
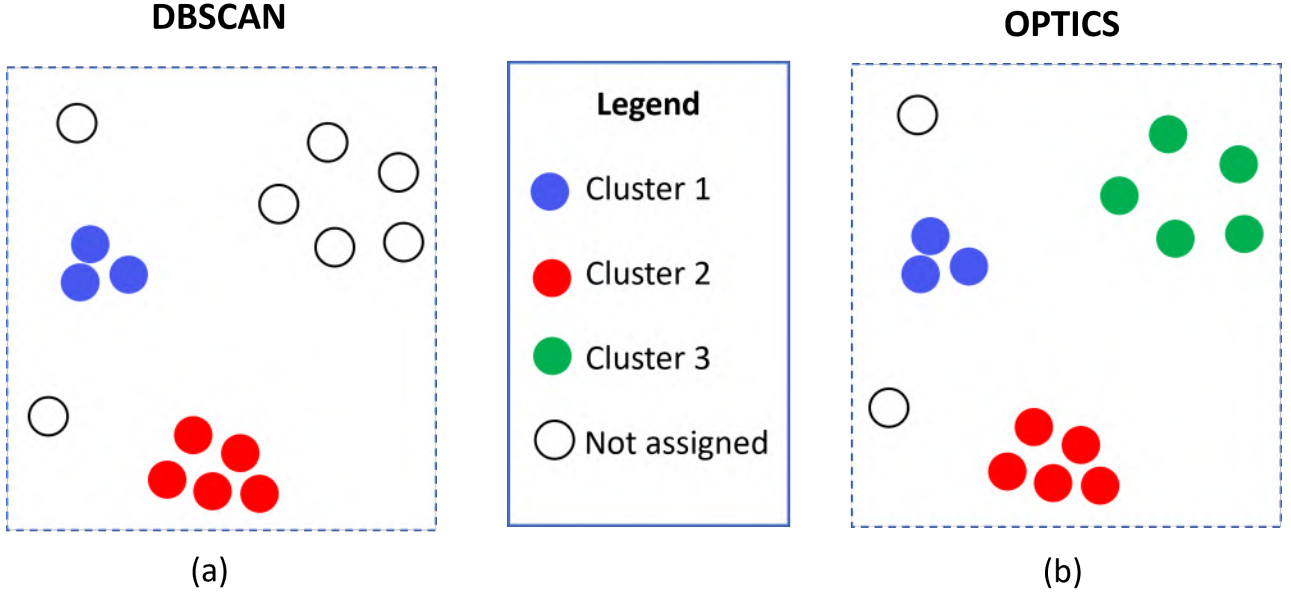FIG. 6: An example of DBSCAN vs centroid based clustering (K-means) for points in 2D space.

FIG. 7: An example of DBSCAN vs OPTICS for points in 2D space.

- ϵ parameter: Two points are considered neighbors if their distance is less than ϵ. In our case it translates to the cosine distance between the learned representations of two atoms (nodes). Specifically for our case, without loss of generality, an atom p is neighbor of an atom q if $CosineDistance(h_p^k, h_q^k) < \epsilon$. Here h_p^k refers to the representation of the pth atom at k^{th} layer of the GNN.

- ϵ-Neighborhood $N_\epsilon(p)$: The points that are within ϵ distance of a node p.

- $MinPts$: Minimum number of points that must be within ϵ distance of an atom p in order to form a cluster.

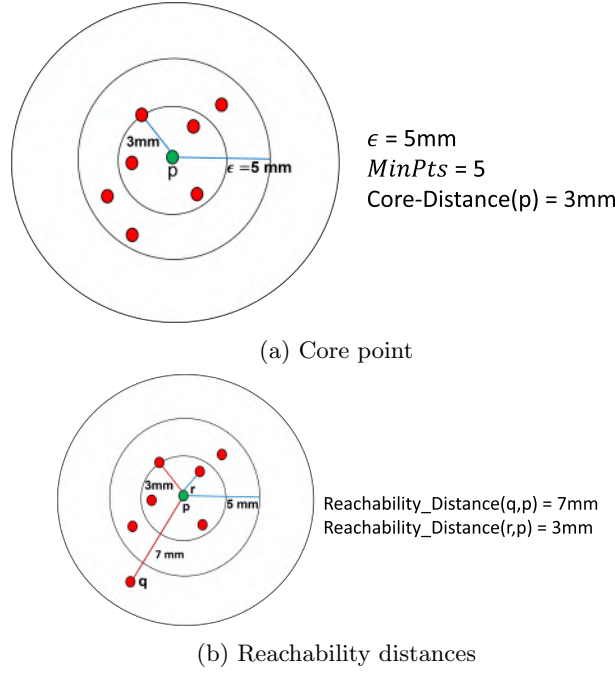- **Core Point:** A point is considered a core point if it has more than $MinPts$ points within ϵ distance.

(a) Core point

(b) Reachability distances

FIG. 8: OPTICS: Core points and reachability distance of points

- **Core Distance:** Minimum ϵ required to make a point p a core point.

$$
\text{core-dist}_{\epsilon, MinPts}(p) =
$$
$$
\begin{cases}
undefined, & \text{if } |N_\epsilon(p)| < MinPts \\
MinPts\text{-th smallest distance in } N_\epsilon, \text{ else}
\end{cases}
\tag{5}
$$

- **Reachability Distance:** Reachability distance between a point p and o is the maximum of the *Core Distance* of p and the distance between p and o. For point o that is reachable from a point p:

$$
\text{reach-dist}_{\epsilon, MinPts}(o, p) =
$$
$$
\begin{cases}
\text{undefined}, & \text{if } |N_\epsilon(p)| < MinPts \\
\max(\text{core-dist}_{\epsilon, MinPts}(p), dist(p, o)), & \text{otherwise.}
\end{cases}
\tag{6}
$$

Fig. 8b shows examples of reachability distances for some points.

In OPTICS clustering, first, core distances of all points are computed. Then, we calculate their reachability distances iteratively. In this iterative process, the data points are chosen in ascending order of their lowest reachability distance to the points already visited until that time. Finally, based upon the reachability distances of all points, we construct a reachability plot as in Fig. 9(c).

In Fig. 9(c), we observe valleys of points (sequence of low reachability distances sandwiched between high reachability distances), which represent high density regions. Valleys get separated by regions of low-density. These valleys can be considered to be part of the same cluster(Fig. 9(b). For example, the colored points(other than gray) are those identified as clusters, while gray ones represent noise. For more details on OPTICS, we refer to the original paper [36].

S4. PARAMETER DETAILS OF GRAPHSAGE

Hyper-parameters used for training the GraphSage GNN model are given below. Note that these parameters were optimized based on the results on the training and validation data.
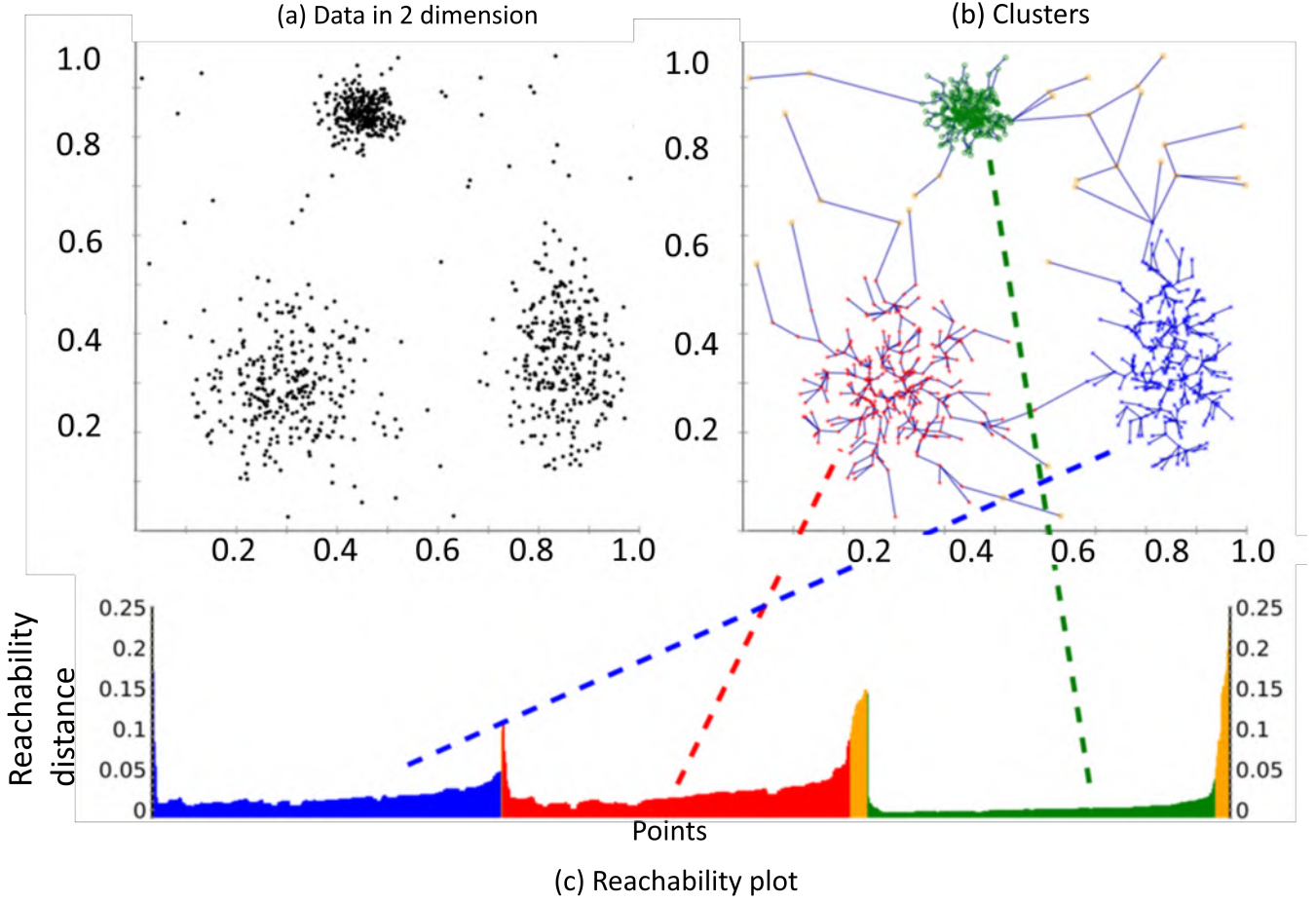
FIG. 9: OPTICS: (a) Sample data in 2 dimension, (b) Clusters, (c) Reachability plot for all points in the sample data. Fig. source. https://github.com/christianversloot/machine-learning-articles/blob/main/performing-optics-clustering-with-python-and-scikit-learn.md

- Number of GNN Layers$(k) = 2$

- Hidden layer size $= 64$

- Random Walk Length$(L) = 4$

- No. of Random Walks per node $= 10$

- Number of Negative Samples $(Q) = 10$

- Optimizer $=$ Adam

- Learning rate $= 0.01$

- Graph batch size $= 5$

- Node batch size $= 250$

- Neighborhood aggregation function $=$ Mean

- Gradient clipping max norm $= 5$

Choice of number of layers (k): As discussed in the main paper, each node aggregates information from its k-hop neighborhood. To identify a suitable value of k, we evaluate the model by performing a random walk test on the set of validation graphs (not seen during training). As per our training loss in Eq. 2, the cosine similarity

of embeddings of a pair of nodes co-occurring in random walks is expected to be higher than cosine-similarity of embeddings of node pairs that do not co-occur in random walk. In Fig. 10, we show the Area under cover- Receiver Operating characteristics (AUC-ROC) score obtained from random walk test for GNN model trained at different number of GNN layers (k). We observe that the performance of the model at $k = 2$ is significantly higher than at $k = 1$. Further, the performance saturates at $k = 2$. Hence, we use $k = 2$, which provides a good balance between efficacy and efficiency (computation cost grows with the number of layers).
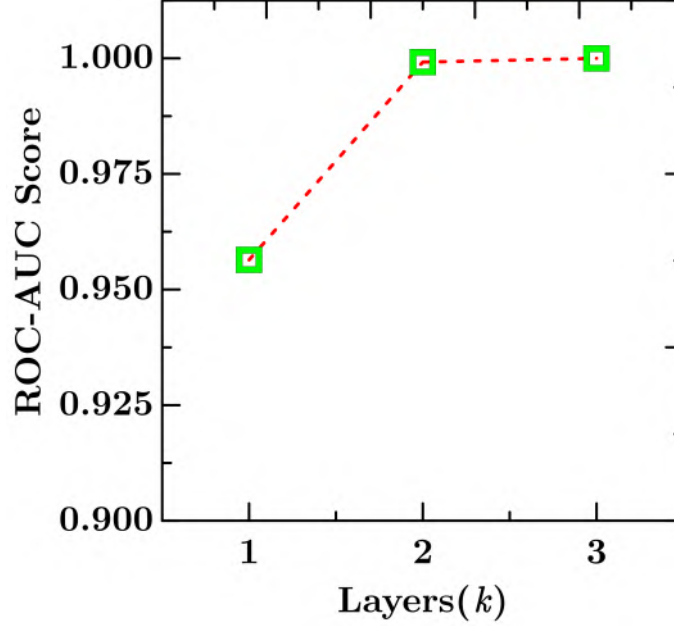

FIG. 10: ROC-AUC score at different number of GNN Layers(k) for random walk test

S5. PARAMETER DETAILS OF OPTICS

We used OPTICS clustering from Scikit-learn library[37]. We mention the parameters used below:

- *MinPts*: Minimum number of samples in a neighborhood for a point to be considered as a core point. We set this to 5.

- ϵ: Maximum distance between two samples for one to be considered as in the neighborhood of the other. We set this to 5.

- Distance metric: Cosine similarity

S6. REACHABILITY PLOT FOR NODES OF 2D LJ SYSTEM

In Fig. **??** we show the reachability distance plot obtained by OPTICS algorithm for nodes of our 2D LJ system (same as shown in Fig. 4.b in main paper). Each color represents a cluster generated by OPTICS as in Fig. 4.b. The gray colored points(noise) instead of forming a cluster themselves, act as a separator between regions of high density.

S7. EFFECT OF RANDOM WALK LENGTH

We also study the effect of length of random walk in Table **??**. We observe that the correlation between MSD of clusters and average cluster potential energy increases monotonically with the length of random walk till 4, after which it drops. A possible reason is that when the length of the random walk increases beyond 4, far away nodes could get incorporated in the set of *positive samples* pairs, and the model might accumulate noise from these node pairs at larger distances.
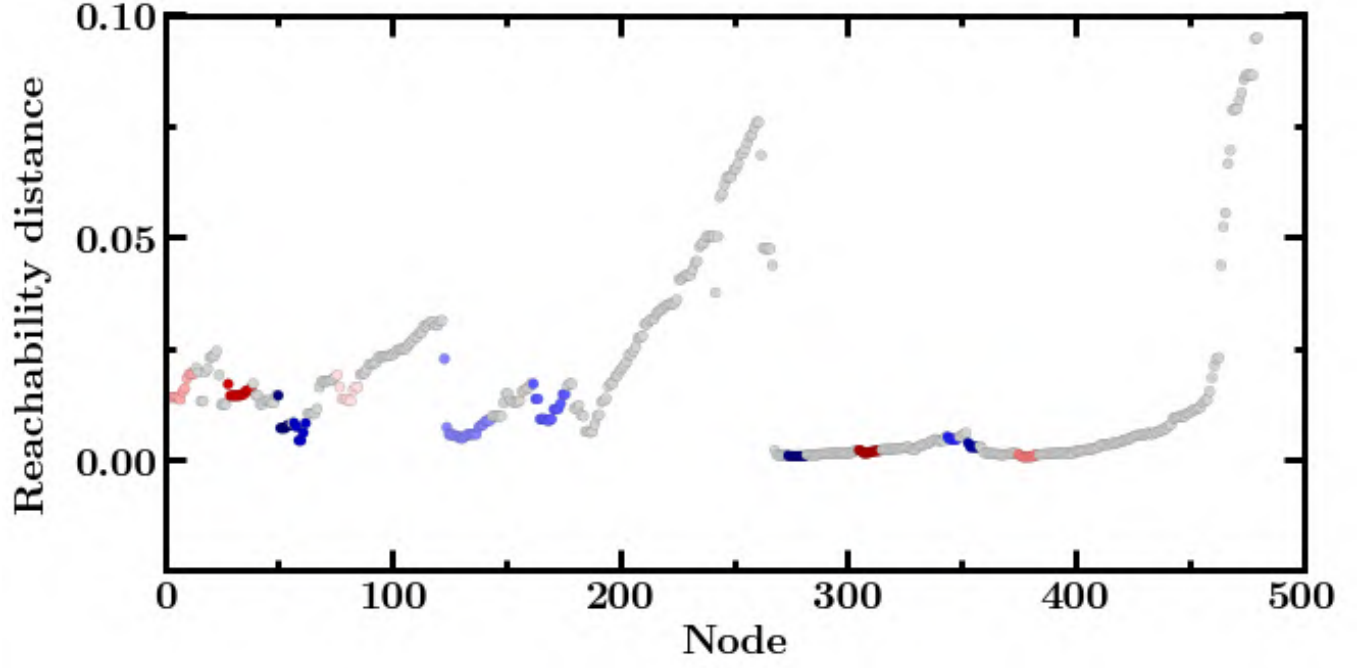
FIG. 11: Reachability distance plot generated by OPTICS for nodes of 2D LJ system

Random Walk Length	Correlation
2	0.658
3	0.714
4	0.747
5	0.627

TABLE I: Pearson's correlation coefficient between MSD and average cluster potential energy