Predicting the entire glassy dynamics from static structure by machine learning relative motion

Hayato Shiba,1, ∗ Masatoshi Hanai,1 Toyotaro Suzumura,2,1 and Takashi Shimokawabe1

1Information Technology Center, University of Tokyo, Chiba 277-0882, Japan
2Graduate School of Information Science and Technology, University of Tokyo, Tokyo 133-8658, Japan
(Dated: September 8, 2022)

The recent developments of machine learning have enabled accurate predictions of glassy dynamics. However, existing machine-learning models are designed so that they learn a single dynamic quantity and related it to the structural features of glassy liquids. In this Letter, we propose a graph neural network model which learns relative motion between neighboring pairs of particles, in addition to self motion of particles. By relating the structural features to these two different dynamical variables, the model autonomously acquires the ability to discern how two different types of dynamic processes, strain fluctuations and particle rearrangements, affect self motion of particles undergoing slow relaxation. Our model affords a methodology to treat these two at the same time and realizes highly precise predictions of glassy structural relaxation over the entire temporal range.

When a liquid is rapidly cooled, it freezes into a glass which retains a random amorphous structure. This phenomenon, referred to as the glass transition, is ubiquitous, but its origin still remains to be clarified. The elucidation of this origin has been an important research topic for decades [1, 2]. Along with the drastic slowdown of atomic motion accompanying the glass transition, localized domains of particles that rearrange more preferentially than others grow up [3–6]. Identifying structural origin for this phenomenon has long been an central problem in the field. Structural analyses of local geometric orders [7–9] and of real-space normal modes of vibrations derived from static structures [10–14] have accumulated evidences for structure–dynamics correspondence. However, the random featureless structure of glasses still prevents us from fully understanding the origin.

Recently, machine-learning methodologies have enabled the accurate determination of structure–dynamics correspondence in glassy systems [15–22]. Among them, graph neural networks (GNNs) [23], a deep-learning algorithm that operates on a graph, have been found to be an suitable methodology [20]. GNNs had already proven to be useful for making predictions of materials properties by directly representing structures of molecules or atomic clusters as nodes and edges of the graph [24–27]. Bapst et al. [20] have leveraged such ability of GNNs to the task of predicting particle mobility in a glass-forming liquid from its static structure, and have shown that GNNs outperform other machine-learning methodologies. This high accuracy in the predictions of glassy dynamics has enabled development of inverse modeling as a further application [28]. It has also led to the development of a cost-effective linear-regression model that can overcome the high computational cost of GNNs but still achieves an equivalent precision in its prediction [21, 22].

In the study by Bapst et al. [20], the prediction accuracy of GNNs is found to outperform not only the that of other machine-learning methods but also of the normal mode analysis, at all the temperature accessible by simulations and across all the temporal scales up to the structural relaxation time. The normal mode analysis is a method to predict how short-time vibrations are spatially distributed. Because short-time vibrations (soft modes) are intimately linked to the slow α relaxations [9, 29–33], the normal mode analysis yields a strong baseline of physics-based prediction of glassy dynamics from static structure [10, 34]. The prediction accuracy of GNNs, nevertheless, becomes relatively low in the fast β dynamics than in the slow α relaxation [20, 22]. We may therefore conjecture that the dynamics is made difficult to predict due to the elastic vibrations, which are erased out after coarse-graining the short-time dynamics. Such accuracy low may possibly be retrieved, as shown by a recent study [35], if we encode short-time vibrations obtained from the simulations [35]. However, there are still no methodologies to enable make predictions of the slow dynamics directly from the static structure.

Here, we propose a GNN model, named “BOnd TAargeting Network (BOTAN)”, which realizes highly precise predictions of the entire glassy relaxation from static structure. BOTAN can learn a target characteristic quantity assigned on edges corresponding to pairs of close-by particles. This model can be positioned as an extension of the GNN used in the previous study [20], which we term as NT-GNN (“node-targeting GNN”) because it was trained using only the particle propensity for motion at each node of the graph. BOTAN and NT-GNN are based on the interaction network [36, 37] consisting of an encoder–process–decoder architecture, where features of nodes and edges in a graph are mutually computed by exchanging messages between a pair of two-layer multi-layer perceptrons (MLPs) assigned for nodes and edges. This architecture is suitable in extracting intricate relationships from simulation data computed with high precision. Because the network treats node and edge features equivalently, regressing the network toward the target quantities on edges as well as nodes [38] is possible by decoding edge embeddings by an MLP [39].

As the new target quantity for learning on edges, we introduce pair-distance change after time \( t \),

\[
E_{ij}(t) = r_{ij}(t) - r_{ij}(0),
\]

which characterize relative motion between neighbor pairs. Here, \( r_{ij}(t) \) denotes the pairwise distance between particles \( i \) and \( j \) at time \( t \). Only the pairs are considered within a threshold distance \( r_e = 2.0\sigma_{AA} \) at the initial time, \( t = 0 \). The dataset for training is trajectory data of a three-dimensional
(3D) binary Kob–Andersen Lennard–Jones mixture [40] generated from 400 initial configurations, wherein the pairwise force is smoothed at its cutoff [41, 42]. The actual target quantity of machine learning is the ensemble average over the result of 32 independent simulation runs generated from each of initial configurations, so-called the isoconfigurational ensemble [10, 30, 34, 43]. In the remainder of this Letter, we use the dimensionless units based on the Lennard-Jones potential and target quantities of learning that are always the average on isoconfigurational ensembles for which explicit notations are abbreviated. Machine learning procedures are similar to NT-GNN [20], except that training with edge targets $\mathcal{E}_{ij}(t)$ is started with pretrained model parameters for reasons described later. The standard $L_2$-norm loss function is used for learning and the input graph is constructed in a manner that pairs of nodes are connected by an edge when the corresponding pair distance is closer than $r_s$. Prediction of the model is examined with test dataset generated from an independent set of 100 initial configurations [39].

In Fig. 1, prediction results of the pair-distance change $\mathcal{E}_{ij}(t)$, for $t = (a) 130$ and (b) 4120 using BOTAN trained at respective time points and at the lowest temperature $T = 0.44$, is shown for a particle configuration in the test dataset. The data are shown in color maps projected onto segments satisfying $r_{ij}(t = 0) < 1.35$, in a two-dimensional (2D) cross section. The “actual” pair-distance changes (the isoconfigurational ensemble average) from the simulation are also shown for each as the ground truth. The time points correspond to 0.03$\tau_\alpha$ and $\tau_\alpha$, with $\tau_\alpha$ the $\alpha$-relaxation time. BOTAN well discerns which specific pairs become separated, particularly for the short time in (a). Over a long time, (b), the prediction deviates from the ground truth, but it still captures pair-level propensity for distance changes as well as the spatial contrast between mobile and immobile regions. The prediction accuracy of BOTAN over $\mathcal{E}_{ij}(t)$ can be quantified by using the Pearson correlation coefficient [20–22] that expresses the proximity between the predicted and ground truth data. This coefficient should be quantified carefully to ensure that the comparison is made over equivalent particle pairs. Therefore, among pairs of type A particles (pairs related to type B are excluded), pairs within the first neighboring distance $r_{ij}(t = 0) < 1.35$ are chosen in this comparison, because neighbor pairs in the first and second neighbor shells are expected to exhibit different distance changes on average. The prediction accuracy thus evaluated is high in the short time, and monotonically declines with time. The high prediction accuracy in the short time originates from a simple anti-correlation relation between pair-distance change $\mathcal{E}_{ij}(t)$ and the initial distance $r_{ij}(t = 0)$; in Fig. 1 (d), a probability distribution function (PDF) map is shown as functions of these quantities for closest neighbor pairs of type A particles in the training dataset at $t = 130 (0.03\tau_\alpha)$ at $T = 0.44$. The pair-distance changes, for most of neighbor pairs, remain one order of magnitude smaller than the particle radii. This quantity, therefore, characterizes strains induced by the elastic vibrations. The anti-correlation gradually disappears but persists up to the $\alpha$-relaxation time as the time increases, making the relation between the two more nontrivial [39]. The anti-correlation relation determines the global minima which BOTAN is optimized into by learning pair-distance changes. Because it becomes more difficult to optimize BOTAN to predict anti-correlation relation for longer time $t$, we pretrain BOTAN with the short-time pair-distance changes at $t = 13$, $T = 0.44$ for the sake of efficient training around the global minima and then use the weight parameters for training with varying $t$.

BOTAN is modeled by adding a decoder applied to edge embeddings, whereas a decoder can be retained for node embeddings. It then becomes possible to train BOTAN with target quantities for both, that is, particle propensity for the self-

FIG. 1. Distribution of predicted and the “actual” (ground truth) pair distance change $\mathcal{E}_{ij}(t)$ at a temperature $T = 0.44$ for $t = (a) 130 (0.03\tau_\alpha)$ and (b) 4120 ($\tau_\alpha$), respectively, which are plotted in color maps projected on each segment. Segments represent pairs within the distance of $r_s = 1.35$. A specific cross section $(11.1 < z < 11.9)$ is cut out of the 3D box system. (c) Pearson correlation coefficients between predicted and ground truth values of $\mathcal{E}_{ij}(t)$ plotted as a function of time $t$, at $T = 0.44$, 0.50, and 0.64. Errorbars depict the median, best, and worst of five independently trained models. Two dotted lines indicate 0.03$\tau_\alpha$ and $\tau_\alpha$ for $T = 0.44$, and arrows show the $\alpha$-relaxation times for respective temperatures. (d) Probability distribution map, plotted using the training dataset (ground truth data) as a function of $\mathcal{E}_{ij}(t)$ and $r_{ij}(t = 0)$, for $t = 130$ at $T = 0.44$. 

\begin{align}
\mathcal{E}_{ij}(t) &= \frac{1}{N} \sum_{i,j} \delta_{ij}(t) \quad \text{(ground truth)} \\
\rho(\mathcal{E}_{ij}(t)) &= \frac{1}{N-1} \sum_{i,j} \delta_{ij}(t) \quad \text{(predicted)} \\
\end{align}
We compare the prediction ability of BOTAN with that of NT-GNN by Bapst et al. In Fig. 2, prediction results for (1) $t = 130 \times (0.03\tau_a)$ and (2) $4120 \times (\tau_a)$ at $T = 0.44$ of (a) NT-GNN and of (b) BOTAN, in addition to (c) the “actual” propensity directly evaluated from the trajectory data in the test dataset. As all the snapshots are shown in the same cross section as in Fig. 1 (a), the spatial correspondence can be seen between the predictions of $S_i(t)$ and $\mathcal{E}_{ij}(t)$. BOTAN clearly makes better prediction of spatially heterogeneous patterns of particle propensity than NT-GNN. The error maps of predictions as the as the difference between the predicted and ground truth propensity map, plotted in (d) for NT-GNN. For both $t = 130$ and $4120$, large scale spatial patterns appear in the error map, showing that NT-GNN fail to capture the full details of heterogeneity in the particle motion. In (e), 3D vector plots are also created to show displacement field locally averaged over nearest neighbors [39]. These plots can extract particle motion that taking place as a consequence of quasi-localized and phonon vibrations [44–46] after removing the effect of interparticle rearrangement. In the both, $t = 130$ and $4120$, we observe collective motion on the length scale considerably exceeding the particle size. The spatial patterns seem similar in between, whereas the flows are more aligned in the former suggesting the effect of phononic motion. The locally-averaged displacements tend to be parallel with the contour lines of the prediction error map especially in the region where the displacements vectors are aligned in parallel. The clear spatial correlation suggests that the collective fluctuations, which is less intimately related to the local structures, is the possible reason for the decline in prediction accuracy of NT-GNN [20, 22].

We assess the prediction accuracy of BOTAN in comparison with the previous model. Because particles of types A and B are expected to exhibit different diffusivity over a long period, the prediction accuracy is evaluated with the Pearson correlation coefficient using only the data for type A particles. Figure 3 shows how accurately BOTAN and NT-GNN predict the particle propensity $S_i(t)$, for three different temperatures, $T = (a) 0.44$, (b) 0.50, and (c) 0.64. The trend of time dependence for NT-GNN is in agreement with the previous results [20]: in the shorter time before reaching the plateau region ($t < 0.2$), the Pearson correlation coefficients assume high values, then it falls below 0.5, and afterwards gradually increases to reach its peak at around the $\alpha$-relaxation time. Conversely, BOTAN outperforms NT-GNN in its predictive accuracy over the entire temporal range and for all the temperatures under investigation.

To this end, we note that the pair-distance change in Eq. (1) have its own significance in spite of the extreme simplicity of the definition. Because the pair-distance changes characterize relative motion, they can capture structural rearrangements being free from the effect of vibrations, which is a feature distinct from standard quantities defined on the basis of

\[
S_i(t) = |r_i(t) - r_i(0)|
\]

in addition to the pair-distance changes $\mathcal{E}_{ij}(t)$. Here, $r_i(t)$ denotes the 3D position of particle $i$ at time $t$. The actual target quantities are again their isoconfigurational ensemble averages. The simultaneous learning of these quantities can beget quantities are again their isoconfigurational ensemble averages. The clear spatial correlation suggests that the collective fluctuations, which is less intimately related to the local structures, is the possible reason for the decline in prediction accuracy of NT-GNN [20, 22].
particle displacements. Similar ideas have also been implemented in the way of counting the replacement of neighboring pairs [10, 47, 48], also termed as “bond breakage” [5, 49, 50]. Furthermore, the pair-distance changes can characterize how largely each pair tends to change to the local shear deformation around rearranging cores; such an aspect of structural changes has long been addressed in the concept of “Shear Transformation Zones” [45, 51–53]. By learning how local strains of non-rearranging regions are distributed and how rearranging hot spots are localized, BOTAN autonomously fixes the errors in the prediction of particle displacements, and overall acquires unprecedented predictive ability on the glassy dynamics.

We further investigate how well NT-GNN can learn two additional quantities that are less affected by elastic vibrations than the self displacement $S_i$. These quantities are both defined to quantify difference between the actual displacement and the uniform strain displacement of the neighbors

$$R_i(t) = \frac{1}{N_{a,n}} \sum_{j \in n} Y_{ij}(t), \quad D_i(t) = \sqrt{\frac{1}{N_{a,n}} \sum_{j \in n} Y_{ij}^2(t)},$$

$$T_{ij}^\alpha(t) = [r_{ij}^\alpha(t) - r_{ij}^\alpha(0)] - [r_{ij}^\alpha(0) - r_{ij}^\alpha(0)] \cdot (\delta_{a} + \epsilon_{a})$$

The vector and tensor components $\alpha, \beta \in \{x, y, z\}$ are explicitly denoted for clarity. Here, $N_{a,n}$ denotes the number of nearest-neighbor particles and $\epsilon_{a}$ represents the local strain tensor that minimizes $D_i(t)$ with respect to local particle displacements, with $\alpha, \beta \in \{x, y, z\}$ the spatial components. In the short-time limit $t \to 0$ where this local strain reduces to zero, the former quantity $R_i(t)$ reduces to relative displacement with respect to neighbor environments [39, 54]. The latter quantity $D_i(t)$ is a variant of $R_i(t)$ having qualitatively similar meaning. This $D_i(t)$ is widely used in the literature [14, 48, 51, 52, 55], as it is susceptible to hot spots where particle rearrangement preferably occurs. In Fig. 3, Pearson correlation coefficients for type-A particles are also shown for predictions of $R_i(t)$ and $D_i(t)$ by NT-GNN. The predictions of $R_i(t)$ and $D_i(t)$ improve especially in the time regions $t < 0.1 \tau_\alpha$, for these quantities less affected by collective thermal fluctuations, suggesting that the short-time vibrations are actually the cause.

In conclusion, we have introduced a GNN model that realizes faithful predictions of glassy dynamics over the entire temporal range, with far more improved accuracy than ever. The key to the improvement is that the model learns relative motion between neighbor pairs related to it to two-body structural correlation, in addition to the self motion of the particles under the influence of vibrations. The GNN model then acquires ability to autonomously “interpret” how particle motion is affected by different dynamical effects, strain fluctuations and particle rearrangements, transparently reconciling the fast $\beta$ and slow $\alpha$ relaxation dynamics. As a consequence, the model affords a method, more reliable than ever, to predict how a structural relaxation proceeds in space and time in a glassy system.

In a broader context, the key design principle of our GNN model lies in incorporating and resolving interplay of different types fluctuations on different spatial scales. As an immediate application, similar machine-learning method can be applied for studying various phenomena where collective flow takes place – deformations under external strain or more diverse problems including e.g. dense active matter [56]. Moreover, the high fidelity of prediction by our GNN model may potentially be used for finding the “reaction coordinates” along which the fluctuations should be enhanced for efficient sampling of molecular trajectory. Extensions to the recently discussed machine-learning-aided sampling technique may also be interesting [57, 58].

The source code for training performed in this Letter, based on PyTorch Geometric [59], is openly available from the GitHub repository [60].

We thank T. Kawasaki for critical inputs into this research. We also thank Y. Asahi and J. J. Molina for helpful discussions. H.S. and T.Sh were supported by JSPS KAKENHI Grant Number JP19H05662 and by “Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures” and “High Performance Computing Infrastructure” in Japan (Project ID: jh220052-NAH). M.H. and T.Su were supported by “Advanced Research Infrastructure for Materials and Nanotechnology” from MEXT, Japan. Training and testing are performed using NVIDIA A100 Tensor Core GPUs (40GB SXM) on Wisteria/BDEC-01 Aquarius subsystem at Information Technology Center, University of Tokyo. For generation of the dataset for training and testing, usage of the following computational resources are acknowledged: Oakforest-PACS at the Joint Center for Advanced High Per-
formance Computing (JCAHPC), Oakbridge-CX at Information Technology Center, University of Tokyo, Supercomputer System B at Institute for Solid State Physics, University of Tokyo, and MASAMUNE-IMR at Institute for Materials Research, Tohoku University.

* E-mail: shiba@cc.u-tokyo.ac.jp

[1] C. A. Angell, Formation of glasses from liquids and biopolymers, Science 267, 1924 (1995).
[2] F. G. Debenedetti and F. H. Stillinger, Supercooled liquids and the glass transition, Nature 410, 259 (2001).
[3] M. D. Ediger, Spatially heterogeneous dynamics in supercooled liquids, Annu. Rev. Phys. Chem. 51, 99 (2000).
[4] W. Kob, C. Donati, S. J. Plimpton, P. H. Poole, and S. C. Glotzer, Dynamical heterogeneities in a supercooled lemnard-jones liquid, Physical Review Letters 79, 2827 (1997).
[5] R. Yamamoto and A. Onuki, Heterogeneous diffusion in highly supercooled liquids, Physical Review Letters 81, 4915 (1998).
[6] L. Berthier and G. Biroli, Theoretical perspective on the glass transition and amorphous materials, Reviews of Modern Physics 83, 587 (2011).
[7] H. Tanaka, H. Tong, R. Shi, and J. Russo, Revealing key structural features hidden in liquids and glasses, Nature Reviews Physics 1, 333 (2019).
[8] C. P. Royall and S. R. Williams, The role of local structure in dynamical arrest, Physics Reports 560, 1 (2015).
[9] H. Tong and H. Tanaka, Revealing hidden structural order controlling both fast and slow glassy dynamics in supercooled liquids, Physical Review X 8, 011041 (2018).
[10] A. Widmer-Cooper, H. Perry, P. Harrowell, and D. R. Reichman, Irreversible reorganization in a supercooled liquid originates from localized soft modes, Nature Physics 4, 711 (2008).
[11] A. Tanguy, B. Mantisi, and M. Tsamados, Vibrational modes as a predictor for plasticity in a model glass, EPL 90, 16004 (2010).
[12] K. Chen, M. L. Manning, P. J. Yunker, W. G. Ellenbrook, Z. Zhang, A. J. Liu, and A. G. Yodh, Measurement of correlations between low-frequency vibrational modes and particle rearrangements in quasi-two-dimensional colloidal glasses, Physical Review Letters 107, 108301 (2011).
[13] A. Ghosh, V. Chikkadi, P. Schall, and D. Bonn, Connecting structural relaxation with the low frequency modes in a hard-sphere colloidal glass, Physical Review Letters 107, 188303 (2011).
[14] S. S. Schoenholz, A. J. Liu, R. A. Riggleman, and J. Rottler, Understanding plastic deformation in thermal glasses from single-soft-spot dynamics, Physical Review X 4, 031014 (2014).
[15] P. Ronthove, S. Chakraborty, D. Hu, M. Sahu, K. K. Sahu, K. F. Kelton, N. A. Mauro, and Z. Nussinov, Detecting hidden spatial and spatio-temporal structures in glasses and complex physical systems by multisresolution network clustering, European Physical Journal E 34, 105 (2011).
[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] 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).
[18] 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 of the United States of America 114, 263 (2017).
[19] E. Boattini, S. Marín-Aguilar, S. Mitra, G. Foffi, F. Smallenburg, and L. Filion, Autonomously hidden local structures in supercooled liquids, Nature Communications 11, 5479 (2020).
[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, and P. Kohli, Unveiling the predictive power of static structure in glassy systems, Nature Physics 16, 448 (2020).
[21] E. Boattini, F. Smallenburg, and L. Filion, Averaging local structure to predict the dynamic propensity in supercooled liquids, Physical Review Letters 127, 088007 (2021).
[22] R. M. Alkemade, E. Boattini, L. Filion, and F. Smallenburg, Comparing machine learning techniques for predicting glassy dynamics, Journal of Chemical Physics 156, 204503 (2022).
[23] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, The graph neural network model, IEEE Transactions on Neural Networks 20, 61 (2009).
[24] T. Xie and J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Physical Review Letters 120, 145301 (2018).
[25] L. Chansrot, A. Das, S. Goyal, T. Lavril, M. Shuaihi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Pulizhiti, A. Sriram, B. Wood, J. Yoon, D. Purik, C. L. Zitnick, and Z. Ulissi, Open catalyst 2020 (oc20) dataset and community challenges, ACS Catalysis 11, 6059 (2021).
[26] V. Fung, J. Zhang, E. Juarez, and B. G. Sumpter, Benchmarking graph neural networks for materials chemistry, npj Computational Materials 7, 84 (2021).
[27] S. Takamato, C. Shinagawa, D. Motoki, K. Nakago, W. Li, I. Kurata, T. Watanabe, Y. Yamaama, H. Iriuchi, Y. Asano, T. Onodera, T. Ishii, T. Kudo, H. Ono, R. Sawada, R. Ishitani, M. Ong, T. Yamaguchi, T. Kataoka, A. Hayashi, N. Charoenphakdee, and T. Ibuka, Towards universal neural network potential for material discovery applicable to arbitrary combination of 45 elements, Nature Communications 13, 2991 (2022).
[28] Q. Wang and L. Zhang, Inverse design of glass structure with deep graph neural networks, Nature Communications 12, 5359 (2021).
[29] U. Buchenau and R. Zorn, A relation between fast and slow motions in glassy and liquid selenium, EPL 18, 523 (1992).
[30] 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).
[31] J. C. Dyre, Colloquium: The glass transition and elastic models of glass-forming liquids, Reviews of Modern Physics 78, 953 (2006).
[32] L. Larini, A. Ottochian, C. D. Michele, and D. Leporini, Universal scaling between structural relaxation and vibrational dynamics in glass-forming liquids and polymers, Nature Physics 4, 42 (2008).
[33] H. W. Hansen, B. Frick, T. Hecksher, J. C. Dyre, and K. Niss, Connection between fragility, mean-squared displacement, and shear modulus in two van der waals bonded glass-forming liquids, Physical Review B 95, 104202 (2017).
[35] A. Tripodo, G. Cordella, F. Puosi, M. Malvaldi, and D. Leporini, Neural networks reveal the impact of the vibrational dynamics in the prediction of the long-time mobility of molecular glassformers, International Journal of Molecular Sciences 23, 9322 (2022).

[36] P. W. Battaglia, R. Pascanu, M. Lai, D. Rezende, and K. Kavukcuoglu, Interaction networks for learning about objects, relations and physics, Advances in Neural Information Processing Systems 29, 4502 (2016).

[37] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, C. Gulcehre, F. Song, A. Ballard, J. Gilmer, G. Dahl, A. Vaswani, K. Allen, C. Nash, V. Langston, C. Dyer, N. Heess, D. Wierstra, P. Kohli, M. Botvinick, O. Vinyals, Y. Li, and R. Pascanu, Relational inductive biases, deep learning, and graph networks (2018), arXiv:1806.01261 [cs.LG].

[38] G. DeZoort, S. Thais, J. Duarte, V. Razavimaleki, M. Atkinson, I. Ojalvo, M. Neubauer, and P. Elmer, Charged particle tracking via edge-classifying interaction networks, Computing and Software for Big Science 5, 26 (2021).

[39] See Supplemental Material [URL will be inserted by publisher] for the following details: GNN model architecture, dataset generation, training procedures, correlations between pair-distance changes and the initial distances with varying time, and the definition of locally-averaged displacement field.

[40] W. Kob and H. C. Andersen, Testing mode-coupling theory for a supercooled binary lemnard-jones mixture: The van hove correlation function, Physical Review E 51, 4626 (1995).

[41] S. Toxvaerd and J. C. Dyre, Communication: Shifted forces in molecular dynamics, The Journal of Chemical Physics 134, 081102 (2011).

[42] M. Shimada, H. Mizuno, and A. Ikeda, Anomalous vibrational properties in the continuum limit of glasses, Physical Review E 97, 022609 (2018).

[43] A. Widmer-Cooper, P. Harrowell, and H. Fynnewever, How reproducible are dynamic heterogeneities in a supercooled liquid?, Physical Review Letters 93, 135701 (2004).

[44] H. Mizuno, H. Shiba, and A. Ikeda, Continuum limit of the vibrational properties of amorphous solids, Proceedings of the National Academy of Sciences of the United States of America 114, E9767 (2017).

[45] D. Richard, G. Kapteijns, J. A. Giannini, M. L. Manning, and E. Lerner, Simple and broadly applicable definition of shear transformation zones, Physical Review Letters 126, 015501 (2021).

[46] Y.-C. Hu and H. Tanaka, Origin of the boson peak in amorphous solids, Nature Physics 18, 669 (2022).

[47] P. Yunker, Z. Zhang, K. B. Aptowicz, and A. G. Yodh, Irreversible rearrangements, correlated domains, and local structure in aging glasses, Physical Review Letters 103, 115701 (2009).

[48] V. Chikkadi and P. Schall, Nonaffine measures of particle displacements in sheared colloidal glasses, Physical Review E 85, 031402 (2012).

[49] H. Shiba, T. Kawasaki, and A. Onuki, Relationship between bond-breakage correlations and four-point correlations in heterogeneous glassy dynamics: Configuration changes and vibration modes, Physical Review E 86, 041504 (2012).

[50] B. Guiselin, C. Scalliet, and L. Berthier, Microscopic origin of excess wings in relaxation spectra of supercooled liquids, Nature Physics 18, 468 (2022).

[51] M. L. Falk and J. S. Langer, Dynamics of viscoplastic deformation in amorphous solids, Physical Review E 57, 7192 (1998).

[52] F. Shimizu, S. Ogata, and J. Li, Theory of shear banding in metallic glasses and molecular dynamics calculations, Materials Transactions 48, 2923 (2007).

[53] M. R. Hasyim and K. K. Mandadapu, A theory of localized excitations in supercooled liquids, Journal of Chemical Physics 155, 044504 (2021).

[54] H. Shiba, T. Kawasaki, and K. Kim, Local density fluctuation governs divergence of viscosity underlying elastic and hydrodynamic anomalies in a 2d glass-forming liquid, Physical Review Letters 123, 265501 (2019).

[55] H. L. Peng, M. Z. Li, and W. H. Wang, Structural signature of plastic deformation in metallic glasses, Physical Review Letters 106, 135503 (2011).

[56] C. Bechinger, R. D. Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, Active particles in complex and crowded environments, Reviews of Modern Physics 88, 045006 (2016).

[57] F. Noé, S. Olsson, J. Köhler, and H. Wu, Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning, Science 365, eaaw1147 (2019).

[58] M. Gabrié, G. M. Rotsko, and E. Vanden-Eijnden, Adaptive monte carlo augmented with normalizing flows, Proceedings of the National Academy of Sciences of the United States of America 119, e2109420119 (2022).

[59] M. Fey and J. E. Lenssen, Fast graph representation learning with PyTorch Geometric, in ICLR Workshop on Representation Learning on Graphs and Manifolds (2019).

[60] H. Shiba, https://github.com/h3-open-bdec/pyg_botan (2022).
I. MACHINE LEARNING MODEL ARCHITECTURE

In this section, we explain the neural network architecture of BOTAN and NT-GNN, and also provide information that was not provided in the previous study conducted by Bapst et al. [S1]. Let $G(V, E)$ be a graph that includes a set of nodes $V$ and edges $E$. For a pair of nodes $v, u \in V$, an edge between $v$ and $u$ is represented as $e_{v,u}$. The neighboring edges of a vertex $v \in V$ are represented as $N(v) \subseteq E$. Given input feature vectors $h_{v}^{in}$ and $h_{e}^{in}$ for a node $v \in V$ and an edge $e \in E$ respectively, the goal is to compute the corresponding output features $h_{v}^{out}$ and $h_{e}^{out}$. As shown in a block diagram in FIG. S1, and by denoting an encoder, a multi-layer perceptron (MLP) and a decoder as $EN(\cdot), MLP(\cdot)$ and $DE(\cdot)$, each layer to compute feature vectors for node $v$ is formulated as follows. First, input feature vectors are encoded by respective encoders

$$h_{v}^{0} = EN(h_{v}^{in}), \quad h_{e}^{0} = EN(h_{e}^{in}). \quad (S1)$$

Then, update of the edge and node features via MLPs are repeated for $n$ times

$$h_{v,e,u}^{m} = MLP(h_{v,e,u}^{m-1} \oplus h_{v}^{m-1} \oplus h_{e,u}^{0}) \quad (S2)$$

$$h_{v}^{m} = MLP\left(h_{v}^{m-1} \oplus \sum_{e \in N(v)} h_{e}^{m} \oplus h_{v}^{0}\right) \quad (S3)$$

where $m$ stands for the iteration index of $n$ repeat cycle ($0 < m \leq n$). In each cycle, messages are passed between nodes and edges, wherein encoded feature vectors are concatenated together before every update. The decoders placed as the final layer yields the output features on each node and edge

$$h_{v}^{out} = DE(h_{v}^{n}), \quad h_{e}^{out} = DE(h_{e}^{n}). \quad (S4)$$

In practice, $EN(\cdot), MLP(\cdot)$ and $DE(\cdot)$ are all implemented by two hidden layers of 64 units with rectified linear units (ReLU) non-linearity. At every repeat cycle of update on edges and nodes in Eqs. (S2) and (S3), information of the neighbor shell (inside the distance of 2.0$r_{AA}$) is propagated. We set the number of this repetition as $n = 7$ so that the information propagates over the whole system (with a box length of 15.06) in a single epoch of learning.

Figure S1. Block diagram of the GNN indicating the steps of computation in an epoch. “EN” and “DE” are abbreviations of “encoder” and “decoder”, respectively. The GNN has an encoder-process-decoder structure consisting of the connected layers of MLPs with two hidden layers of 64 units with ReLU activation.
In addition to the node output feature $h_{\text{out}}^v$ which have been exploited to learn particle propensity in previous studies [S1–S3], we introduced the decoder for the edge feature $h_{\text{out}}^e$ in this model, which is an essential innovation in the present study. It is essential that edges receive messages from both neighbor nodes $v, u$ in each repetitive cycle of message passing in Eq. (S2), though this detail does not affect the results and was not explicitly addressed in the study by Bapst et al. [S1].

In this study, a graph data for learning is constructed based on the particle configuration wherein the particles are represented by its nodes. Edges are assigned to pairs of particles $i$ and $j$ within a distance threshold value $r_{ij}(t = 0) \leq r_c$, so that the neighborhood relation can be represented. We chose a threshold length $r_c = 2.0\sigma_{AA}$ that has the best predictive performance on particle propensity [S1]. Then these graphs are fed into the GNN by encoding particle types (A and B) and 3D relative positions $\mathbf{r}$ of pair species depending on the combination of pair species as $\mu, \nu \in \{A, B\}$, so that the force and the potential vanish at the cutoff length $r_c$.

The neighborhood relation can be represented. We chose a threshold length $r_c = 2.0\sigma_{AA}$ that has the best predictive performance on particle propensity [S1]. Then these graphs are fed into the GNN by encoding particle types (A and B) and 3D relative positions between particles independently into the nodes and edges as single floating-point numbers, wherein the types (A and B) are converted into 0 and 1.

II. DATASET DETAILS

We employ a Kob–Andersen-type binary Lennard–Jones mixture in three dimensions [S1, S4] which is usually given by an interatomic pairwise potential

$$u(r) = 4\epsilon_{\mu\nu} \left[ \left( \frac{\sigma_{\mu\nu}}{r} \right)^{12} - \left( \frac{\sigma_{\mu\nu}}{r} \right)^{6} \right].$$

$\mu, \nu \in \{A, B\}$ denotes the particle types, where the number of particles belonging to type A (denoted as $N_A$) is 80% of the total number $N = 4096$. The interaction energy and particle size are defined as $\epsilon_{AA} = 1.0$, $\epsilon_{AB} = 1.5$, $\epsilon_{BB} = 0.5$, $\sigma_{AA} = 1.0$, $\sigma_{AB} = 0.8$, and $\sigma_{BB} = 0.88$. In this study, distances, time, and temperature are denoted in units of $\sigma_{AA}$, $\tau = \sqrt{m\epsilon_{AA}/k_B}$, with $k_B$ the Boltzmann constant.

For improving the accuracy of numerical integration over a large number of steps, the pairwise potential is modified in the track of previous studies as [S5, S6]

$$U(r) = u(r) - u(r_c) - (r - r_c) \frac{du(r)}{dr} \bigg|_{r=r_c},$$

so that the force and the potential vanish at the cutoff length $r = r_c$. The cutoff lengths are set differently for different combinations of pair species depending on the combination of pair species as $r_c = 2.5\sigma_{\alpha\beta}$.

The procedure of dataset generation is similar to that in the previous study [S1] except that the system size is fixed. For all the simulations, the number density of the system is fixed at $N/V = 1.2$. Time step of numerical integration is kept at $\Delta t = 10^{-3}$ throughout. We set three target temperatures, $T = 0.44$, 0.50, and 0.64. In Fig. S2, the self-part of the intermediate scattering functions for particles of type A

$$F_s(k, t) = \frac{1}{N_A} \sum_{j \in A} \exp \left[ i k \cdot (\mathbf{r}_j(t + t_0) - \mathbf{r}_j(t_0)) \right],$$

Figure S2. Intermediate scattering function $F_s(k, t)$ in Eq. (S6) is shown for temperatures $T = 0.44$, 0.47, 0.50, 0.56, and 0.64 as functions of time. Vertical dotted lines indicate time points corresponding to $0.03\tau_\alpha$ and $\tau_\alpha$ for $T = 0.44$, which are mainly investigated in the Letter.
are shown for these temperatures (and additionally $T = 0.47$ and 0.56), where the wavenumber $k$ is set at $2\pi/\sigma_{\text{AA}}$. The $\alpha$-relaxation times $\tau_\alpha$ are evaluated as 4120, 217, and 16.5, for $T = 0.44$, 0.50, and 0.64, respectively. In the Letter, time points $t = 130$ (0.03$\tau_\alpha$) and 4120 ($\tau_\alpha$) are mainly investigated at the lowest temperature $T = 0.44$, which is indicated in the same Fig. S2 with vertical dotted lines.

For each of these three target temperatures, 500 independent particle configurations are generated as follows. After keeping the temperature at $T = 5.0$ for the time lapse of $t_0 = 10^4$, the system is cooled rapidly (in $t_{\text{cool}} = 20$) to the target temperature. The temperature is subsequently kept constant until the system reaches the steady state up to time scale of $40\tau_\alpha$. The final particle configuration thus obtained is used as the initial one at $t = 0$ in the production run for generating the dataset.

From the 500 initial configurations, we have introduced the isoconfigurational ensemble [S7, S8] by running 32 separate microcanonical (NVE) simulations all starting from the same configuration; however, the initial velocities are randomly given from the Maxwell-Boltzmann distribution at the target temperature. These simulations are conducted approximately up to $30\tau_\alpha$, where the particle configurations are sampled at logarithmically increasing time points as summarized in Table I. Among the 500 initial configurations, 400 configurations (12,800 runs) are used for training and the other 100 (3,200 runs) are for testing.

### III. CORRELATION BETWEEN PAIR-DISTANCE CHANGE AND INITIAL DISTANCE WITH VARYING TIME

BOTAN, the GNN model we propose in this study, learns pair-distance changes $E_{ij}(t)$, defined in Eq. (1) in the Letter. In the short time region corresponding to the fast $\beta$ relaxation, the high prediction accuracy of BOTAN originates from the simple anti-correlations between $E_{ij}(t)$ and the initial distances $r_{ij}(t = 0)$. As will be shown in the next section, this anti-correlation relationship defines the global minima to which weight parameters of BOTAN should be optimized. In Fig. S3, the relation between these quantities are shown in the form of probability distribution functions (PDF) for all particle pairs of type A (within the distance $r_{ij}(t = 0)$), over all the training dataset for the temperature $T = 0.44$ (400 initial configurations, namely, 12800 runs), at different time points $t = 13.0, 130, 412, 1300, 4120,$ and $13000.$ The anti-correlation remains at in the time shorter than the $\alpha$-relaxation time $t = 4120$, however, becomes less clear as the time increases. BOTAN leverages the anti-correlation for improving the predictions of particle propensity for motion.

### IV. TRAINING PROCEDURE

The learning setups and hyperparameters in the present study are similar to those used by Bapst et al. [S1]. The training dataset is learned to minimize the loss function, wherein the neural network is optimized by Adam algorithm without norm regularization with the learning rate of $10^{-4}$ with a standard implementation in PyTorch. The data loaded for training are augmented by applying random rotation of the simulation box to the particle positions, as all the target quantities for learning are scalar variables that are rotationally invariant.

The loss functions for learning are defined by

$$\mathcal{L}_M = p\mathcal{L}_S + (1 - p)\mathcal{L}_E,$$

(S7)

where $\mathcal{L}_S$ and $\mathcal{L}_E$ are standard $L_2$-norm loss functions respectively defined for deviations of particle propensity $S_i(t)$ and pair-distance changes $E_{ij}(t)$. These may be written as

$$\mathcal{L}_S = \sum_{j=1}^{N} \left[ \hat{S}_j(t; \{r_i\}) - \langle S_j(t) \rangle_{\text{IC}} \right]^2,$$

(S8)

$$\mathcal{L}_E = \sum_{j=1}^{N} \sum_{k \in \text{nn.}} \left[ \hat{E}_{jk}(t; \{r_i\}) - \langle E_{jk}(t) \rangle_{\text{IC}} \right]^2,$$

(S9)
Figure S3. The PDF showing the relationship between the initial pair distance $r_{ij}(t = 0)$ at $T = 0.44$, for time points $t = (a) 13.0$, (b) 130 (the same as Fig. 1 (d) in the Letter), (c) 412, (d) 1300, (e) 4120, and (f) 13000.

where $\langle \cdot \rangle_{IC}$ represents the average over the isoconfigurational ensemble and $\hat{S}_j(t, \{r_j\})$ and $\hat{E}_{jk}(t, \{r_i\})$ denote the values of particle propensity and pair-distance change as the output from the GNN. By using the loss function $L_M$, the model can be trained simultaneously via both $S_j(t)$ and $E_{jk}(t)$, and further, these quantities can be inferred simultaneously as well. $p$ is treated as a hyperparameter ($0 \leq p \leq 1$) as it decides the weight of learning between nodes and edges. Two limiting cases are $p = 0$ wherein the model only learns pair-distance changes $E_{ij}(t)$ (corresponding to Fig. 1 in the Letter), and $p = 1$ which reduces to NT-GNN [S1]. For the simultaneous learning on nodes and edges ($0 < p < 1$), we make the choice of $p = 0.4$ in this study. This choice is justified by an ablation experiment in which Pearson correlation coefficients are calculated with changing $p$ for $t = 130$ and $T = 0.44$ (starting not with the pretrained network parameters but with the random initial weight), as shown in Fig. S5.

It takes a large number of epochs (typically larger than 3000) to train BOTAN with pair-distance changes $E_{ij}(t)$ from random initial weights especially in the short time. As demonstrated in Sec. III, pair-distance changes $E_{ij}(t)$ have strong anti-correlation with the initial distance $r_{ij}(t = 0)$. While this relation defines where the global minima for optimization of BOTAN is located, for a neural network to search for such global minima requires a large number of optimization steps. Therefore, we start all

Figure S4. Training and test losses (left axis) for $t = 13$ at $T = 0.44$ for the pretraining of pair-distance changes $E_{ij}(t)$ as an edge feature, with $p = 0$. The Pearson correlation coefficient $\rho$ representing the predictive accuracy of the model is evaluated at every epoch using the test data (100 configurations), and plotted as a function of epoch (right axis).
the trainings of BOTAN by pair-distance changes (when \( p = 0 \) and \( p = 0.4 \)) with a model pretrained over 2000 epochs using the training data for short time dynamics \( t = 13.0 \) at \( T = 0.44 \) (corresponding to Fig. S3 (a)). In Fig. S4, we show learning curve (epoch-dependent training and test losses) of this pretraining and the Pearson correlation coefficients \( \rho \) (for pairs of type-A particles within the distance 1.35), which represents predictive accuracy of the model at every epoch of training. Then afterwards, we load this pretrained model parameters and train BOTAN with varying the time \( t \) and the temperature \( T \) over 1000 epochs. All these training is performed without minibatching of the graph input. In Fig. S5, we show results of the ablation experiment on the parameter \( p \), in which Pearson correlation coefficients are evaluated for \( t = 130 \) and \( T = 0.44 \).

In the training with NT-GNN (\( p = 1 \)), the minibatch size of the graph input are fixed at 5 and the number of epochs for learning are fixed at 1000. Although Bapst et al. applied no minibatching in their previous study [S1], this difference does not bring much differences in the results presented in this Letter.

In Fig. S6 and S7, the training and and test loss functions are plotted as a function of epochs (so called the learning curves), for \( p = 0 \) (corresponding to Fig. 1 of the letter) and \( p = 0.4 \). Also, Pearson correlation coefficients \( \rho \), the proximity between the ground truth and the predictions of the GNN model at each epoch, are also shown.

In all these training scenarios, all the particles (particle pairs) are learned simultaneously, regardless of the particle types. The training with NT-GNN over 1000 epochs required approximately 2 hours when using one NVIDIA A100 Tensor Core GPU (40 GB SXM). Essentially the same amount of time is required for training of BOTAN, except that there may exist a substantial overhead arising from computation of interparticle distances, of which the actual elapsed time depends on its implementation.

\[ \langle \Delta r_i^{\text{ave}}(t) \rangle_{\text{IC}} = \left\{ \frac{\sum_j \left[ r_i(t) - r_j(0) \right] \Theta \left( r_c - r_{ij}(0) \right)}{\sum_j \Theta \left( r_c - r_{ij}(0) \right)} \right\}_{\text{IC}}, \quad (S10) \]

where \( \Theta(x) \) denotes Heaviside’s step function and \( r_c \) is the cutoff length of coarse-graining \( r_c \) set at 1.35. Each of them is plotted in a cross-section with a thickness of 1.2, where the 3D vectors are enlarged for better visibility.

\[ \rho \]

VI. CODE AVAILABILITY

The source code used for training and evaluation is available at GitHub repository https://github.com/h3-open-BDEC/pyg_botan. This code is an implementation of BOTAN using PyTorch Geometric with extensions from the original code for TensorFlow provided by Bapst et al. [S1]. Data for evaluations will also be made available via the hyperlink in the same GitHub repository for three years after the publication of the Letter.

Figure S5. Dependence of Pearson correlation coefficients (see the Letter for the definitions) between prediction result of BOTAN and the ground truths plotted as a function of \( p \), a parameter determining the weight of learning between nodes and edge target quantities, namely, particle propensity \( S_i(t) \) and pair-distance changes \( E_{ij}(t) \). Tests are for the data for \( t = 130 \) at \( T = 0.44 \). When \( 0 < p < 1 \), these quantities are learnt simultaneously.
Figure S6. Training and test losses (left axis) for $t = (a) 130$ and (b) 4120 at $T = 0.44$ for the learning of pair-distance changes $E_{ij}(t)$ as an edge feature ($p = 0$, corresponding to Fig. 1 of the Letter), in addition to the resultant Pearson correlation coefficient (right axis) between the predicted and the ground truth data.

Figure S7. Training and test losses (left axis) for $t = (a) 130$ and (b) 4120 at $T = 0.44$ for the simultaneous learning of particle propensity $S_i(t)$ and pair-distance changes $E_{ij}(t)$ (with $p = 0.4$, corresponding to Fig. 2 (a) and (e), of the Letter), in addition to the resultant Pearson correlation coefficients (right axis) between the predicted and the ground truth data for both $S_i(t)$ and $E_{ij}(t)$.

[S1] V. Bapst, T. Keck, A. Grabska-Barwińska, C. Donner, E. D. Cubuk, S. S. Schoenholz, A. Obika, A. W. Nelson, T. Back, D. Hassabis, and P. Kohli, Unveiling the predictive power of static structure in glassy systems, Nature Physics 16, 448 (2020).
[S2] E. Boattini, F. Smallenburg, and L. Filion, Averaging local structure to predict the dynamic propensity in supercooled liquids, Physical Review Letters 127, 088007 (2021).
[S3] R. M. Alkemade, E. Boattini, L. Filion, and F. Smallenburg, Comparing machine learning techniques for predicting glassy dynamics, Journal of Chemical Physics 156, 204503 (2022).
[S4] W. Kob and H. C. Andersen, Testing mode-coupling theory for a supercooled binary lennard-jones mixture: The van hove correlation function, Physical Review E 51, 4626 (1995).
[S5] S. Toxvaerd and J. C. Dyre, Communication: Shifted forces in molecular dynamics, The Journal of Chemical Physics 134, 081102 (2011).
[S6] M. Shimada, H. Mizuno, and A. Ikeda, Anomalous vibrational properties in the continuum limit of glasses, Physical Review E 97, 022609 (2018).
[S7] A. Widmer-Cooper, P. Harrowell, and H. Fynewever, How reproducible are dynamic heterogeneities in a supercooled liquid?, Physical Review Letters 93, 135701 (2004).
[S8] A. Widmer-Cooper, H. Perry, P. Harrowell, and D. R. Reichman, Irreversible reorganization in a supercooled liquid originates from localized soft modes, Nature Physics 4, 711 (2008).
[S9] M. Fey and J. E. Lenssen, Fast graph representation learning with PyTorch Geometric, in ICLR Workshop on Representation Learning on Graphs and Manifolds (2019).