TOPOLOGY-AWARE NEURAL GRAPH OPERATOR (TANGO) FOR MATERIAL CONSTITUTIVE LAWS

Anonymous authors
Paper under double-blind review

ABSTRACT

This work introduces TANGO (Topology-Aware Neural Graph Operator), a graph-based operator learning framework for predicting nonlinear mechanical response in disordered material networks. TANGO represents material networks as graphs, exemplified with polymer gels where nodes encode polymer chains or crosslinkers and edges distinguish chemical bonds from physical entanglements. Strain-conditioned, attention-based message passing propagates topological and local structural information, producing graph embeddings that a decoder maps to continuous stress-strain functions. Trained on coarse-grained molecular dynamics simulations of polymer networks, TANGO accurately predicts full stress-strain curves, recovers key mechanical properties (elastic modulus, ultimate strength, work to failure), and generalizes to unseen chain lengths. Inference is over five orders of magnitude faster than simulations, enabling high-throughput evaluation of network designs. These results demonstrate that topology-aware operator learning can efficiently capture nonlinear, connectivity-dependent mechanics in disordered materials, bridging graph representation learning and constitutive modeling.

1 INTRODUCTION

Predicting the mechanical response of complex material networks from their underlying topology remains a fundamental challenge across length scales, from biological fibrous tissues to architected metamaterials and polymer gels. Classical constitutive models grounded in statistical mechanics provide physical insight but require extensive parameterization and struggle to capture heterogeneity arising from irregular network architectures (Flory (1953); Arruda & Boyce (1993); Ogden (1997); Gent (1996)). High-fidelity simulations offer detailed predictions (Kremer & Grest (1990); Arora et al. (2021); Ge et al. (2021)) but incur computational costs that limit rapid material screening and inverse design.

Recent advances in geometric deep learning offer an alternative. Graph neural networks (GNNs) have achieved success in molecular property prediction (Gilmer et al. (2017); Wieder et al. (2020); Wu et al. (2018)) by operating directly on graph-structured data while preserving symmetries, while neural operators learn mappings between function spaces for physical systems (Li et al. (2020b); Lu et al. (2021); Li et al. (2020a); Yang et al. (2021)). However, GNNs are typically applied to small molecules or crystalline materials with regular structure (Xie & Grossman (2018); Chen et al. (2023)), and neural operators rely on continuum representations. Disordered material networks, where irregular connectivity, structural heterogeneity, and deformation-induced reorganization collectively govern mechanical response, remain largely unexplored.

This work introduces TANGO (Topology-Aware Neural Graph Operator), which bridges these paradigms by learning a mapping from discrete network graphs to continuous mechanical response functions, $\mathcal{G}_\theta : (\mathcal{G}_0, \varepsilon) \mapsto \sigma(\varepsilon)$. Demonstrated on polymer networks, the framework represents molecular-scale connectivity as graphs whose nodes encode constituent-level structural features (polymer chains, crosslinkers), while edges distinguish chemical bonds from physical entanglements (Fig. 1(a)). Strain-conditioned graph embeddings with attention-based message passing capture nonlinear, topology-dependent mechanics across deformation regimes (Fig. 1(b)). Trained on coarse-grained molecular dynamics simulations (CGMD) under tensile deformation (Fig. 1(c)), TANGO accurately reproduces stress-strain responses and generalizes to unseen network topologies.

The key novelties of this work are threefold. **First**, TANGO provides a generalizable framework for learning structure-property relationships in disordered networks by explicitly encoding topology (connectivity patterns, bond types, and local structural features) into the learning process. Unlike conventional GNNs that yield static or single-point predictions, TANGO maps network topology to continuous response functions, generating full stress-strain curves. **Second**, strain-conditioned, attention-based graph convolutions capture nonlinear mechanics arising from network reorganization during deformation, including strain stiffening and topology-dependent failure, beyond the scope of grid-based neural operators designed for regular structures. **Third**, TANGO achieves CGMD-comparable accuracy with up to five orders-of-magnitude speedups, enabling high-throughput screening and inverse design of material networks.

2 METHOD

TANGO learns the mapping $\mathcal{G}_\theta : (\mathcal{G}_0, \varepsilon) \mapsto \sigma(\varepsilon)$ from real polymer network topology $\mathcal{G}_0$ and applied strain ε to a mechanical stress response $\sigma(\varepsilon)$ (Fig. 1(e)). The architecture consists of three key components: (1) a graph encoder that embeds network topology through message passing, (2) a strain encoder that processes the deformation state as a conditioning input, and (3) a decoder that predicts stress from the combined representations.

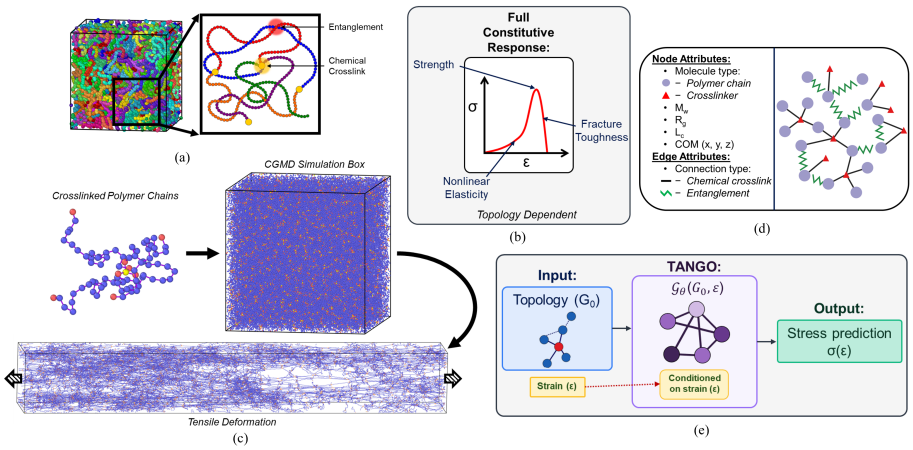

Figure 1: (a) Schematic of a polymer network, showing chains and connectivity. (b) Graph-based representation of the network extracted from CGMD simulations. (c) CGMD simulation of polymer network deformation. (d) Predicted stress-strain response under the trained loading condition. (e) Overview of the TANGO modeling framework, including graph encoding, strain conditioning, and stress prediction.

2.1 NETWORK REPRESENTATION AND GRAPH ENCODING

Each polymer network is represented as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ (Fig. 1(d)), where nodes $v_i \in \mathcal{V}$ represent molecular constituents (polymer chains and crosslinkers) and edges $e_{ij} \in \mathcal{E}$ represent chemical crosslinks or physical entanglements. Node features $\mathbf{x}_i \in \mathbb{R}^7$ encode key structural descriptors, including molecule type, end-to-end length, contour length, radius of gyration, and center-of-mass coordinates. Edge features $\mathbf{e}_{ij} \in 0, 1$ distinguish bond types (see Appendix B).

The graph encoder employs Graph Attention Networks (GAT) (Velickovic et al. (2017)) to aggregate topological information through $L = 3$ message-passing layers with multi-head attention ($H = 2$ heads). Node representations $\mathbf{h}_i^{(\ell)} \in \mathbb{R}^d$ are updated via attention-weighted aggregation followed by layer normalization:

$$\mathbf{h}_i^{(1)} = \mathrm{LN}\left(\sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}^{(1)}\mathbf{W}^{(1)}\mathbf{h}_j^{(0)}\right)\right) \tag{1}$$

where $\mathbf{h}_i^{(0)} = \mathbf{x}_i$, $\mathcal{N}(i)$ denotes the neighborhood of node i, $\mathbf{W}^{(\ell)} \in \mathbb{R}^{d\times d}$ is a learnable weight matrix with hidden dimension $d = 32$, $\sigma(\cdot)$ is the ReLU activation, and $\mathrm{LN}(\cdot)$ denotes layer normalization. Residual connections are applied for $\ell > 1$; full details are provided in Appendix B. The attention coefficients $\alpha_{ij}^{(\ell)}$ are computed using edge-conditioned attention that incorporates both node features and edge attributes:

$$\alpha_{ij}^{(\ell)} = \frac{\exp\left(\mathrm{LeakyReLU}\left(\mathbf{a}^{(\ell)T}[\mathbf{W}^{(\ell)}\mathbf{h}_i^{(\ell-1)} \| \mathbf{W}^{(\ell)}\mathbf{h}_j^{(\ell-1)} \| \boldsymbol{\phi}(\mathbf{e}_{ij})]\right)\right)}{\sum_{k\in\mathcal{N}(i)} \exp\left(\mathrm{LeakyReLU}\left(\mathbf{a}^{(\ell)T}[\mathbf{W}^{(\ell)}\mathbf{h}_i^{(\ell-1)} \| \mathbf{W}^{(\ell)}\mathbf{h}_k^{(\ell-1)} \| \boldsymbol{\phi}(\mathbf{e}_{ik})]\right)\right)} \tag{2}$$

where $\boldsymbol{\phi}(\cdot)$ embeds edge attributes and $\|$ denotes concatenation. A graph-level embedding $\mathbf{g} \in \mathbb{R}^{32}$ is obtained via mean pooling over nodes.

2.2 Strain Conditioning and Stress Prediction

The applied uniaxial strain ε is encoded using a three-layer feedforward network $f_{\text{strain}} : \mathbb{R} \to \mathbb{R}^{32}$, yielding a strain embedding $\mathbf{s} = f_{\text{strain}}(\varepsilon)$ that captures nonlinear deformation effects. The graph and strain embeddings are concatenated and passed to a decoder network to predict the engineering stress,

$$\hat{\sigma} = f_{\text{decoder}}([\mathbf{g}\|\mathbf{s}]), \tag{3}$$

where $f_{\text{decoder}} : \mathbb{R}^{64} \to \mathbb{R}$ is a three-layer feedforward network. This formulation enables TANGO to learn a continuous mapping from network topology and strain to stress, producing complete stress-strain response functions. Full architectural specifications are provided in Appendix B.

3 Results and Discussion

3.1 Constitutive Law Learning

TANGO was trained on 3200 polymer networks with chain lengths $N = 20$, 50, 80, and 100 (Appendix A). Overall performance is summarized in Fig. 2. The pointwise stress parity plot (Fig. 2(a)) shows strong agreement with ground truth (R^2=0.939), with residual scatter attributable to intrinsic noise in finite-size fracture simulations. Predicted stress-strain curves for unseen systems (Fig. 2(b)) accurately reproduce entropic elasticity, strain hardening, and peak stress across all N, with error bands remaining narrow in the elastic regime and widening primarily near fracture. This suggests that further improvements may require richer features encoding fracture initiation or training data spanning a broader range of failure modes. Beyond accuracy, TANGO provides a substantial computational advantage. Predicting a full stress-strain curve requires approximately 4.8 s (1.3×10^{-3} GPU hours) on a single GPU, whereas an equivalent CGMD simulation requires $\sim$200 CPU core hours. When compared on a wall-clock time basis, this represents a speedup of $\sim$ five orders of magnitude, enabling high-throughput exploration of polymer network designs that would be infeasible with direct simulation.

Mechanical properties (elastic modulus, ultimate tensile strength, and work to failure) are extracted post hoc from predicted curves and compared against two baselines (Fig. 2(c-e)). The FFNN baseline replaces TANGO's message-passing encoder with a three-layer feed-forward network operating on statistically pooled node features, while the structure-to-property (S2P) baseline directly predicts scalar properties from pooled graph features, bypassing curve prediction. Both baselines discard explicit topological information (Appendix C).

For elastic modulus (Fig. 2(c)), S2P achieves the highest R^2 (0.890), while TANGO (R^2=0.753) substantially outperforms the FFNN (R^2=0.292). TANGO and S2P attain identical MAE (0.003), indicating that message passing recovers accuracy comparable to a task-specific S2P model, despite

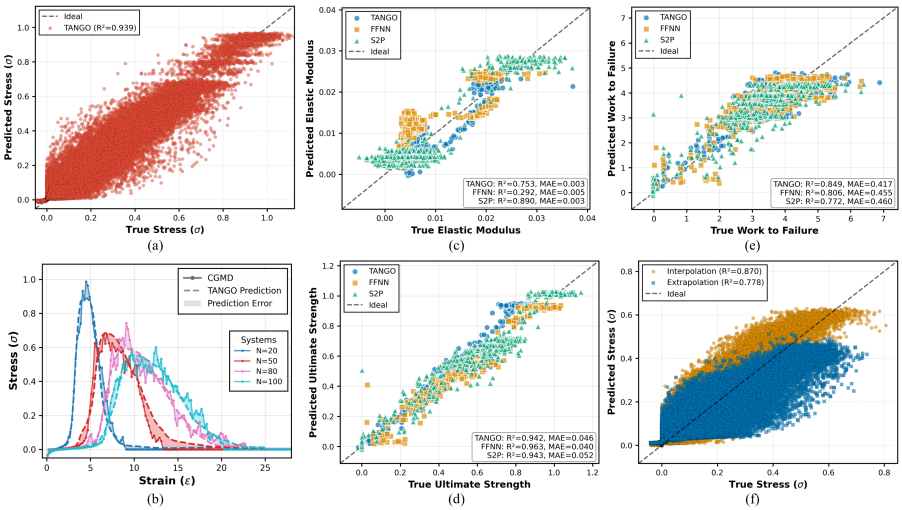

Figure 2: (a) Parity plot of TANGO-predicted versus true pointwise stress (R^2=0.939). (b) TANGO-predicted stress-strain curves (dashed) compared to CGMD ground truth (solid) for unseen N=20, 50, 80, 100 systems; shaded regions denote prediction error. (c) Predicted versus true elastic modulus for TANGO, FFNN, and S2P. (d) Predicted versus true ultimate tensile strength. (e) Predicted versus true work to failure. (f) Parity plots for interpolation (N=80, R^2=0.870) and extrapolation (N=100, R^2=0.778).

predicting the full stress-strain response. The poor FFNN performance highlights the importance of network connectivity for modulus prediction. For ultimate tensile strength (Fig. 2(d)), all models perform similarly (FFNN R^2=0.963, S2P 0.943, TANGO 0.942), suggesting that this property is primarily governed by local chain features accessible without explicit topological encoding. For work to failure (Fig. 2(e)), TANGO performs best (R^2=0.849, MAE=0.417), outperforming both FFNN and S2P. Because work to failure integrates the full stress-strain history, this result underscores the advantage of learning the complete response function rather than isolated scalar targets.

Generalization was evaluated via leave-one-out experiments on chain length. In the interpolation setting, TANGO was trained on N=20, 50, 100 and tested on N=80; in extrapolation it was trained on N=20, 50, 80 and tested on N=100 (Fig. 2(f)). Interpolation performance remains strong (R^2=0.870), while extrapolation exhibits a moderate decline (R^2=0.778), consistent with the challenge of predicting fracture behavior in longer chains beyond the training distribution. Errors are concentrated at high stress, indicating reduced accuracy in the post-peak fracture regime.

4 Conclusion

This work introduces TANGO, a topology-aware neural graph operator that learns constitutive behavior directly from disordered material networks. By encoding network connectivity and conditioning graph representations on strain, TANGO maps molecular-scale topology to macroscopic mechanical response, predicting full stress-strain curves rather than isolated scalar properties. Applied to polymer networks, it accurately captures elastic, strain-hardening, and failure regimes, outperforms structure-agnostic baselines on topology-sensitive metrics, and generalizes to unseen chain lengths. Beyond predictive accuracy, TANGO delivers a dramatic efficiency gain: complete stress-strain responses are obtained in seconds, versus hundreds of CPU hours required for CGMD simulations. This enables high-throughput exploration, screening, and inverse design of complex material networks. More broadly, TANGO bridges graph-based representation learning with operator learning, offering a general framework for constitutive modeling of disordered architectures. Future extensions incorporating richer descriptors of local damage and broader loading conditions could further enhance fidelity and expand applicability to biological tissues, soft composites, and architected metamaterials.

LLM USAGE STATEMENT

The use of large language models (LLMs) in this work was strictly limited to manuscript preparation and presentation-related tasks. Specifically, LLMs were employed to assist with language refinement, improve clarity and coherence of the text, and format results for presentation. LLMs were not used in the development of algorithms, derivation of theoretical results, or execution and analysis of experiments. All scientific concepts, methodologies, and conclusions presented in this work are solely the contributions of the authors.

REFERENCES

Akash Arora, Tzyy-Shyang Lin, and Bradley D Olsen. Coarse-grained simulations for fracture of polymer networks: Stress versus topological inhomogeneities. *Macromolecules*, 55(1):4–14, 2021.

Ellen M Arruda and Mary C Boyce. A three-dimensional constitutive model for the large stretch behavior of rubber elastic materials. *Journal of the Mechanics and Physics of Solids*, 41(2): 389–412, 1993.

Axel Brünger, Charles L Brooks III, and Martin Karplus. Stochastic boundary conditions for molecular dynamics simulations of st2 water. *Chemical physics letters*, 105(5):495–500, 1984.

Saian Chen, Aziguli Wulamu, Qiping Zou, Han Zheng, Li Wen, Xi Guo, Han Chen, Taohong Zhang, and Ying Zhang. Md-gnn: A mechanism-data-driven graph neural network for molecular properties prediction and new material discovery. *Journal of Molecular Graphics and Modelling*, 123: 108506, 2023.

Paul J Flory. *Principles of polymer chemistry*. Cornell university press, 1953.

Ting Ge, Jiuling Wang, and Mark O Robbins. Effects of coarse-graining on molecular simulations of mechanical properties of glassy polymers. *Macromolecules*, 54(5):2277–2287, 2021.

Alan N Gent. A new constitutive relation for rubber. *Rubber chemistry and technology*, 69(1): 59–61, 1996.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. Pmlr, 2017.

Kurt Kremer and Gary S Grest. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *The Journal of Chemical Physics*, 92(8):5057–5086, 1990.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020a.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations. *arXiv preprint arXiv:2003.03485*, 2020b.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Raymond W Ogden. *Non-linear elastic deformations*. Courier Corporation, 1997.

Jörg Rottler, Sandra Barsky, and Mark O Robbins. Cracks and crazes: On calculating the macroscopic fracture energy of glassy polymers from molecular simulations. *Physical review letters*, 89(14):148304, 2002.

Scott W Sides, Gary S Grest, and Mark J Stevens. Surface-tethered chains entangled in a polymer melt: Effects on adhesion dynamics. *Physical Review E*, 64(5):050802, 2001.

Mark J Stevens. Interfacial fracture between highly cross-linked polymer networks and a solid surface: Effect of interfacial bond density. *Macromolecules*, 34(8):2710–2718, 2001.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. *stat*, 1050(20):10–48550, 2017.

Oliver Wieder, Stefan Kohlbacher, Mélaine Kuenemann, Arthur Garon, Pierre Ducrot, Thomas Seidel, and Thierry Langer. A compact review of molecular property prediction with graph neural networks. *Drug Discovery Today: Technologies*, 37:1–12, 2020.

Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530, 2018.

Tian Xie and Jeffrey C Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120(14):145301, 2018.

Yan Yang, Angela F Gao, Jorge C Castellanos, Zachary E Ross, Kamyar Azizzadenesheli, and Robert W Clayton. Seismic wave propagation and inversion with neural operators. *The Seismic Record*, 1(3):126–134, 2021.

A DATASET

The dataset used in this work consists of generic CGMD of polymer networks. All simulations followed a standardized and computationally efficient protocol designed to enable the generation of a large number of statistically independent training samples. Initial configurations were constructed by randomly populating a simulation box with polymer chains and multifunctional crosslinker molecules at a prescribed number density.

Excluded-volume interactions between all bead pairs were modeled using a purely repulsive Lennard-Jones (LJ) potential, with $E_r = 0$ for $r_{ij} > 2^{1/6}\sigma$, and reduced units $\sigma = \epsilon = 1[k_B T]$. Covalent connectivity along polymer backbones was enforced using a quartic bond potential, adopted from prior studies of glassy polymer mechanics, which replaces the conventional FENE potential to permit realistic bond scission under large deformation Rottler et al. (2002); Sides et al. (2001); Stevens (2001). The terminal beads of each polymer chain were assigned a distinct bead type, while retaining identical nonbonded interactions to enable selective end-linking during the formation of spontaneously crosslinked networks. Likewise, crosslinker beads were assigned a separate type to distinguish their functionality during network formation. All simulations were performed under a Langevin thermostat Brünger et al. (1984) at an elevated temperature $T = 1$ with friction coefficient $\xi = 0.01$, and time integration was carried out using an NVE integrator, corresponding to Brownian dynamics in the overdamped limit.

Following initialization, systems with chain length N were subjected to extended equilibration to ensure relaxation into a low-energy, stress-free state. Equilibration was assessed by monitoring the mean-squared internal displacement of polymer segments; the emergence of a plateau in the displacement of terminal beads was taken as an indicator of full configurational relaxation.

Next, each system was simulated in its equilibrium state to sample a large ensemble of statistically independent configurations. Configurations of the polymer melt were periodically recorded at time intervals sufficiently long to ensure decorrelation, thereby yielding n independent equilibrated polymer melts, denoted Ω_n. For each chain length, $n = 400$ independent equilibrium systems were generated.

The equilibrated melts were subsequently crosslinked by enabling designated crosslinker atoms to form up to four new bonds, with a maximum of two bonds per polymer chain end. Bond formation was restricted to chain ends located within a cutoff distance of $r_{\mathrm{cut}} = 2^{1/6}\sigma$, corresponding to the purely repulsive Lennard–Jones interaction range, in order to maximize functionalization while avoiding energetic overlap. Following this gelation process, the formation of a percolated crosslinked polymer network, the systems were annealed to a reduced temperature of $T^* = 0.4$. The resulting quenched networks were then subjected to uniaxial tensile deformation under conditions of constant cross-sectional area, providing a consistent mechanical framework for probing fracture and failure behavior.

To explicitly sample heterogeneously formed polymer networks, an additional protocol was employed following melt equilibration. In this procedure, crosslinking was enabled while configurations were periodically recorded throughout the gelation process, thereby capturing networks spanning a broad range of connectivity, from weakly to highly crosslinked states. These partially formed gels were subsequently subjected to the same thermal quench and uniaxial tensile deformation protocols as the fully crosslinked systems.

Polymer chain lengths of $N = 20, 50, 80$, and 100 were considered, with 800 total samples generated for each Ns. Of these, 400 networks were produced using the fully crosslinked gel protocol, while the remaining 400 were obtained via the connectivity-sampling protocol. This dataset constitutes a diverse ensemble of polymer networks, enabling TANGO to learn meaningful relationships that depend not only on single-chain statistics but also on higher-order network topology and structural heterogeneity.

B TANGO: Implementation Details

B.1 Architecture Specifications

B.1.1 Node and Edge Features

Each node v_i in the polymer network graph is characterized by a 7-dimensional feature vector:

$$\mathbf{x}_i = [\mathbb{I}_{\text{crosslinker}}, \ell_i, \ell_{c,i}, R_{g,i}, x_i, y_i, z_i]^T \tag{4}$$

where $\mathbb{I}_{\text{crosslinker}} \in \{0, 1\}$ distinguishes crosslinker nodes from polymer chains, ℓ_i is the end-to-end length, $\ell_{c,i}$ is the contour length, $R_{g,i}$ is the radius of gyration, and (x_i, y_i, z_i) are the center of mass coordinates. Edge features $\mathbf{e}_{ij} \in \{0, 1\}$ distinguish chemical bonds from physical entanglements.

B.1.2 Message Passing Layers

The edge embedding function transforms scalar edge features to 16 dimensions:

$$\boldsymbol{\phi}(\mathbf{e}_{ij}) = \text{Dropout}_{0.10}\left(\text{ReLU}\left(\mathbf{W}_e \mathbf{e}_{ij} + \mathbf{b}_e\right)\right) \tag{5}$$

Graph attention layers compute edge-conditioned attention coefficients:

$$\alpha_{ij}^{(\ell)} = \frac{\exp\left(e_{ij}^{(\ell)}\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(e_{ik}^{(\ell)}\right)} \tag{6}$$

where

$$e_{ij}^{(\ell)} = \text{LeakyReLU}\left(\mathbf{a}^{(\ell)T}[\mathbf{W}^{(\ell)}\mathbf{h}_i^{(\ell-1)} \| \mathbf{W}^{(\ell)}\mathbf{h}_j^{(\ell-1)} \| \boldsymbol{\phi}(\mathbf{e}_{ij})]\right) \tag{7}$$

Multi-head attention ($H = 2$) outputs are averaged:

$$\mathbf{h}_i^{(\ell),\text{agg}} = \frac{1}{H}\sum_{h=1}^{H}\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(\ell,h)} \mathbf{W}^{(\ell,h)} \mathbf{h}_j^{(\ell-1)} \tag{8}$$

For the first layer ($\ell = 1$):

$$\mathbf{h}_i^{(1)} = \text{LN}\left(\text{Dropout}_{0.2}\left(\text{ReLU}\left(\mathbf{h}_i^{(1),\text{agg}}\right)\right)\right) \tag{9}$$

For subsequent layers ($\ell > 1$), residual connections are applied:

$$\mathbf{h}_i^{(\ell)} = \mathbf{h}_i^{(\ell-1)} + \text{LN}\left(\text{Dropout}_{0.2}\left(\text{ReLU}\left(\mathbf{h}_i^{(\ell),\text{agg}}\right)\right)\right) \tag{10}$$

After $L = 3$ layers, mean pooling produces the graph embedding:

$$\mathbf{g} = \frac{1}{|\mathcal{V}|}\sum_{i=1}^{|\mathcal{V}|} \mathbf{h}_i^{(L)} \in \mathbb{R}^{32} \tag{11}$$

B.1.3 Strain Encoder and Stress Decoder

The strain encoder $f_{\text{strain}} : \mathbb{R} \to \mathbb{R}^{32}$ consists of three fully connected layers (1→64→64→32) with LayerNorm, ReLU, and dropout (0.2). The stress decoder $f_{\text{decoder}} : \mathbb{R}^{64} \to \mathbb{R}$ takes the concatenated embeddings $[\mathbf{g} \| \mathbf{s}]$ and applies three layers (64→64→32→1) with decreasing dropout rates (0.2, 0.1, 0.0).

B.2 Training Procedure

B.2.1 Data Preprocessing

All features and targets are normalized using z-score standardization: $\tilde{\mathbf{x}} = (\mathbf{x} - \boldsymbol{\mu})/(\boldsymbol{\sigma} + 10^{-6})$. Normalization statistics are computed exclusively from training data. For node and edge features, statistics are computed from unique network topologies to avoid overrepresenting graphs with many strain measurements.

B.2.2 Loss Function

The model is trained to minimize the mean squared error (MSE) between predicted and true stress values in the normalized space:

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^{N} (\hat{\sigma}_i - \sigma_i)^2 \tag{12}$$

where N is the mini-batch size, $\hat{\sigma}_i$ is the model prediction, σ_i is the true stress value, and θ represents all learnable parameters.

MSE was chosen over alternative loss functions (MAE, Huber) because: (1) it heavily penalizes large errors, which is critical for accurate stress prediction in failure-critical applications, (2) it provides smooth gradients that facilitate optimization in deep networks, and (3) preliminary experiments showed superior convergence compared to MAE-based training. All predictions are denormalized before computing evaluation metrics.

B.2.3 Optimization

The MSE loss is minimized using AdamW optimizer with decoupled weight decay ($\lambda = 1 \times 10^{-5}$). The learning rate follows cosine annealing with warm restarts:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min}) \left(1 + \cos\left(\frac{T_{\text{cur}}}{T_0}\pi\right)\right) \tag{13}$$

where $\eta_{\max} = 1 \times 10^{-4}$, $\eta_{\min} = 10^{-6}$, and $T_0 = 30$ epochs.

Gradient accumulation over $K = 2$ micro-batches of size $B = 32$ provides an effective batch size of 64. Gradients are clipped to a maximum norm of 1.0. Early stopping with patience $P = 20$ and minimum improvement threshold $\delta = 10^{-6}$ prevents overfitting.

B.2.4 Data Splitting

Partitioning occurs at the network topology level to ensure rigorous generalization testing. Unique network identifiers are split 70/10/20 into train/validation/test sets using a fixed random seed (42). This ensures no network topology appears in multiple splits, preventing the model from simply interpolating between known strain values of memorized structures.

B.3 Implementation

The model is implemented in PyTorch 2.0.1 with PyTorch Geometric 2.3.0, trained on NVIDIA A100 GPUs. Mixed precision training (AMP) with dynamic loss scaling accelerates computation. Training uses 4 parallel data loader workers with pin memory enabled. Training takes ~ 7.5 hours (60–80 epochs with early stopping). All random seeds are fixed at 42 for reproducibility.

C BENCHMARK MODELS

To rigorously evaluate TANGO's graph-based approach, two fair baseline models were developed that use identical raw features and training procedures but replace message passing with simple statistical pooling. All edge information, adjacency, and graph topology are removed, thereby isolating the contribution of information propagation over graph structure.

C.0.1 FEED-FORWARD BASELINE FOR STRESS PREDICTION

The feed-forward neural network (FFNN) baseline uses the same node features as TANGO (7 node features per element) but aggregates them using mean and standard deviation pooling operations. The resulting 14-dimensional feature vector comprises: node features pooled with mean (7) and std (7). This pooled representation is processed through a feed-forward network with architecture matching TANGO: hidden dimension 32, 3 layers with residual connections, LayerNorm, and dropout (0.2). The strain encoder and stress decoder are identical to TANGO.

C.0.2 FEED-FORWARD BASELINE FOR STRUCTURE-PROPERTY PREDICTION

The structure-to-property (S2P) baseline predicts three mechanical properties directly from network structure: elastic modulus (linear regression slope in strain range 1.0–2.0), ultimate stress (maximum stress value), and work to failure (area under the stress-strain curve). The model uses the same pooled node features as the stress-prediction baseline (14 dimensions, corresponding to the mean and standard deviation of seven raw node features). These features are processed through a shared feed-forward trunk with three layers (hidden dimension 32) and residual connections, followed by three task-specific prediction heads, each mapping $32 \to 1$ architecture.

C.0.3 TRAINING CONSISTENCY

Both baselines use identical training procedures to TANGO: AdamW optimizer ($\eta = 10^{-4}$, $\lambda = 1 \times 10^{-5}$), cosine annealing with warm restarts, gradient accumulation (effective batch size 64), gradient clipping (max norm 1.0), early stopping (patience 20), and graph-level data splitting with the same random seed (42). This ensures that performance differences reflect the value of message passing over graph structure rather than implementation or hyperparameter advantages.