Executive Summary

Final Project

University of Cambridge

Department of Physics

Prepared by:
Sabahattin Mert Daloglu

Supervised by:
Miles Cranmer

Word Count: 999

June 30, 2024

1 Model Architecture

This project reproduces and extends the main results from **Discovering Symbolic Models from Deep Learning with Inductive Biases** by Cranmer et al [1]. The study focuses on using both explicit (architecture bottlenecks) and implicit (regularization) inductive biases to create sparse latent representations in Graph Networks (GNs). The internal architecture of these GNs is shown in Figure 1.

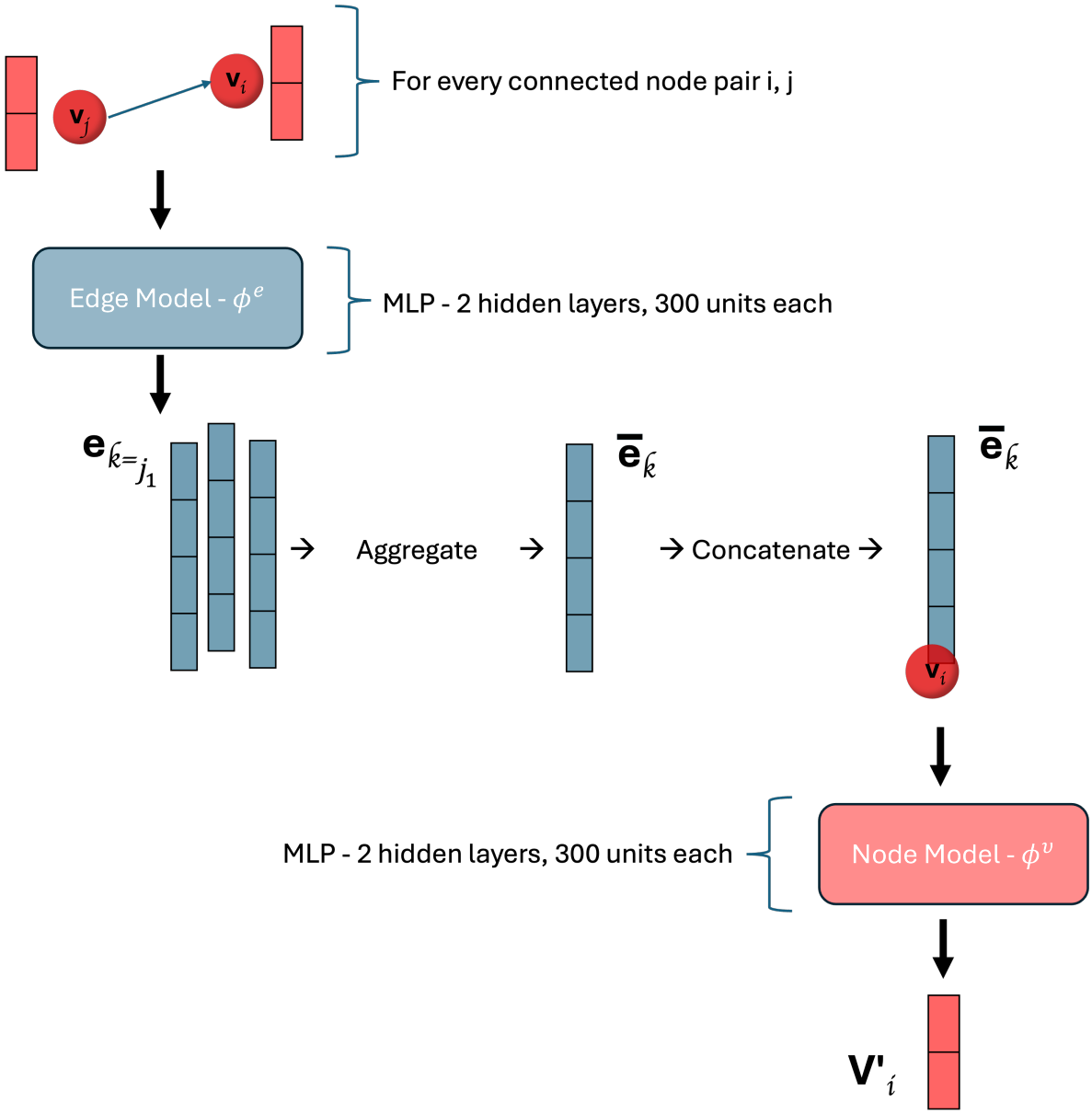

Figure 1: Message passing Graph Network architecture utilized in the study.

2 Datasets

The GN models were trained on 2-dimensional 4-particle classical mechanical systems, including gravitational dynamics with $1/r$ and $1/r^2$ potentials, charge systems, and spring systems. Inputs to the GN included independent snapshots of simulations with instantaneous position, velocity features, and acceleration labels. The graph structure input to the GN model is depicted in Figure 2

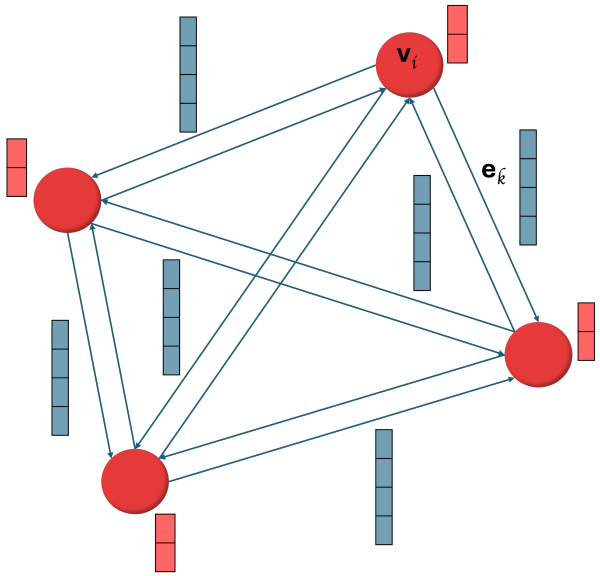

Figure 2: Graph structure that is used to represent a snapshot of a simulation data input to the GN model.

3 Experiments and Results

We implemented L1 regularization to penalize the magnitude of edge messages and a Bottleneck Model to constrain the dimensions of the message vector to the true dimension of the system. We tested if the learned messages were a linear transformation of the pairwise forces, dependent on the latent space dimensions matching the actual system dimensions. This was empirically validated by fitting the most important components of the edge messages to the true forces. The R^2 fit results for the L1 regularization with a constant coefficient ranged from 0.545 to 0.850 for different systems, differing from the near-perfect fit results reported in the original paper. Additionally, the edge message sparsity did not converge to the true system dimensions, as shown in the boxplot of Figure3 illustrating the standard deviations of the first 15 message components for the spring system.

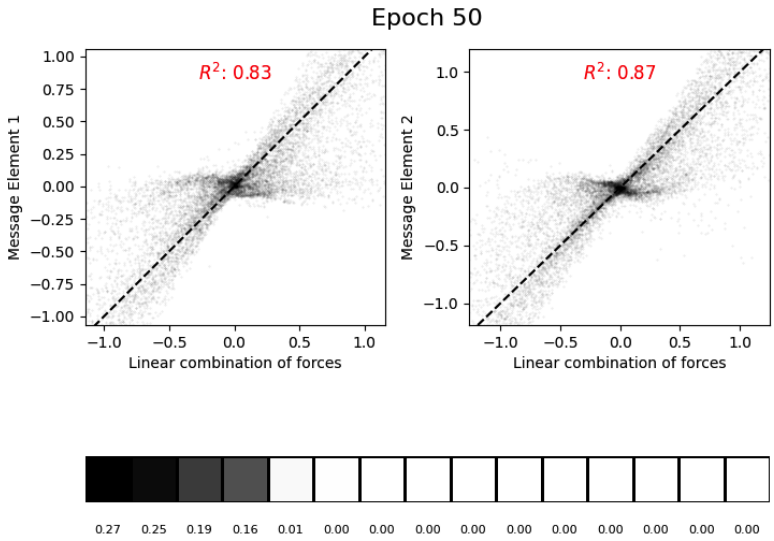

Figure 3: Correlation of linear transformation of true forces with learned messages for the L1 model with constant coefficient trained on the spring system.

Discussions with the original paper's first author led to the implementation of linear and triangular schedules for the L1 coefficient, improving the R^2 results and indicating a stronger correlation between

the learned messages and the true forces as shown in Table 1.

Simulation	L_1-Constant	L_1-Linear	L_1-Triangle
Charge-2	0.800	**0.810**	0.800
r^{-1}-2	0.545	**0.550**	0.500
r^{-2}-2	**0.545**	0.530	0.410
Spring-2	0.850	**0.970**	0.840

Table 1: The average R^2 value of a fit of a linear combination of true force components to the message components for different L_1 regularization schedules, applied on pruned data values across two dimensions. Numbers close to 1 indicate the messages and true force are strongly correlated.

The linear schedule was adopted for further analysis due to its superior performance in the majority of the systems. Moreover, the correct dimensionality of the latent space was achieved with the use of the linear schedule as seen in the boxplot in Figure 4.

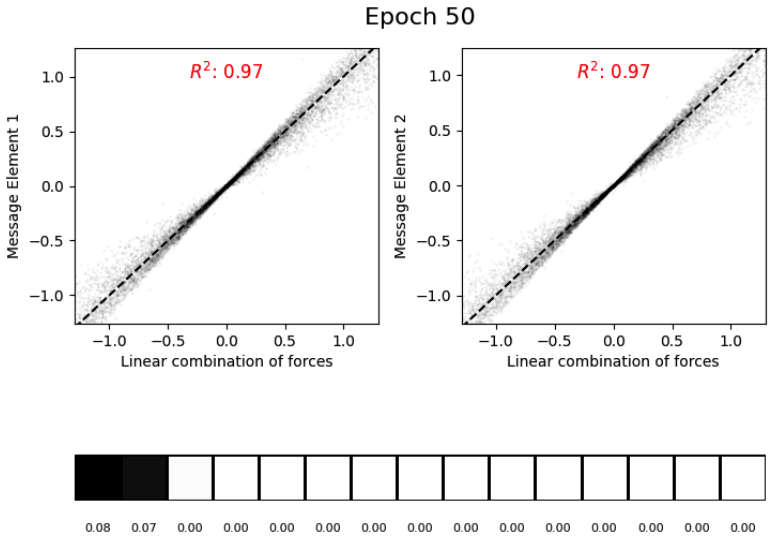

Figure 4: Improved correlation of true forces with learned messages under the linear L_1 schedule.

The R^2 fit values of all other models applied to all systems yielded moderate values as shown in Table 2. We note the original paper achieved higher R^2 values for the majority of L1 and Bottleneck models except for the spring system for which we achieved similar results.

Simulation	Standard	Bottleneck	L_1-Linear	KL
Charge-2	0.515	0.605	**0.810**	-0.01
r^{-1}-2	**0.560**	0.549	0.550	-0.04
r^{-2}-2	0.525	**0.545**	0.530	-0.03
Spring-2	0.590	0.875	**0.970**	-0.08

Table 2: The average R^2 value of a fit of a linear combination of true force components to the message components for all models across two dimensions. Numbers close to 1 indicate the messages and true force are strongly correlated.

Our results for the charge system show significantly improved performance across Bottleneck and L1 models, which we attribute to the data preprocessing step that involved removing graphs with extremely high acceleration labels. This hypothesis is supported by the highest R^2 values observed in

the spring system, both in the original study and our current analysis, which consistently exhibited a narrow range of acceleration values between -300 and 300. In contrast, the charge system, with acceleration values spanning from -150,000 to 100,000, displayed the widest range and the lowest R^2 values in the original study, without the implementation of data pruning. Consequently, we opted to discard graphs with outlier acceleration values based on the Quantile Range to prevent potential skewing of the loss and training process.

Following data preprocessing, we applied symbolic regression separately to the edge and node models of the GNs. The modular nature of GNs, with distinct node and edge structures, is particularly suited to Newtonian physics problems. This involves sequentially calculating pairwise forces (edge model), calculating the net force (aggregation step), and applying Newton's second law (node model). This modularity also simplifies the symbolic regression tasks, thereby reducing the computational complexity required to explore the solution space. The effectiveness of this approach was demonstrated by the successful recovery of known classical mechanical equations through symbolic regression applied to the edge models, using node features as inputs. This was solely possible in models employing L1 regularization with a linear schedule and Bottleneck models. Examples of equations that were correctly recovered, representing linear transformations of the true forces, include:

- Spring, 2D, L_1-Linear (expect $\phi_1^e \approx (a \cdot (\Delta x, \Delta y))\frac{(r-1)}{r} + b$).

$$\phi_1^e \approx \frac{(\Delta x + \Delta y) \cdot (r - 0.9916974)}{r \cdot (m_1 + 6.990457)}$$

- Charge, 2D, Bottleneck (expect expect $\phi_1^e \approx (a \cdot (\Delta x, \Delta y))\frac{(q_1 q_2)}{r^3} + b$).

$$\phi_1^e \approx 26.9520032830977 \cdot \frac{q_1 q_2 (\Delta x + \Delta y)}{r^3 m_1}$$

- $1/r^2$, 2D, Bottleneck (expect $\phi_1^e \approx a \cdot (\Delta x, \Delta y)\frac{m_1 m_2}{r^3} + b$).

$$\phi_1^e \approx 4.306725 \cdot m_2 \cdot \frac{(\Delta x + 4.12451991619306 \cdot \Delta y)}{r^3} + 0.16935113$$

We conclude that the effectiveness of symbolic distillation in L1 and Bottleneck models is closely linked to achieving the correct sparsity level, which accurately reflects the true dimensionality of 2. Additionally, we observed that the node model equations are learned to be normalized by the mass of the target particle, m_1. Following this, the aggregation step sums the normalized pairwise forces exerted on the target particle by others, thereby computing the linearly transformed instantaneous acceleration. Subsequently, the node model deduces the mapping from these linearly transformed instantaneous accelerations back to the true instantaneous accelerations, which are then validated against the actual acceleration data. Ultimately, we integrate the symbolic equations from both the edge and node models to formulate a comprehensive expression for the GN. This expression is evaluated against the GN's performance on novel test data, where the GN demonstrates superior accuracy with lower test loss compared to the distilled symbolic equation. This outcome challenges the principle of Occam's Razor, which suggests a preference for simpler explanations, indicating its inapplicability to classical mechanical systems in our analysis.

4 Key Findings

- **Successful Symbolic Distillation:** Despite some lower R^2 values in Newtonian systems compared to the original study, our approach successfully derived the correct physical laws for Bottleneck and L1 models trained on all systems.
- **Sparse Latent Representations:** L1 regularization with linear schedule and the Bottleneck Model facilitated sparse latent representations that accurately reflect the true dimensional structure.
- **Limitations and Potential for Future Research:** Limited training epochs (50 due to time and resource constraints) suggest further optimization in future studies could enhance the capture of linear transformations of true forces with higher R^2 fit values. Moreover, alternative priors for the KL model could be explored such as the Laplacian distribution to promote better sparsity.

5 Methodological Enhancements

- **Data Pruning:** Removing outliers in the charge system data stabilized the training process and improved learning effectiveness.
- **L1 Regularization Scheduling:** Implementing a scheduler allowed broader exploration of potential solutions early in training, preventing premature convergence to suboptimal configurations.

6 Importance of the Study:

- **Advancing Interpretability in Deep Learning for Science:** Combining deep learning with symbolic regression has significantly improved model interpretability by providing analytical expression, revealing underlying physical principles learned by the Graph Network.

References

[1] M. Cranmer, A. Sanchez-Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel, and S. Ho. Discovering symbolic models from deep learning with inductive biases, 2020.