DISCOVERING GENERALIZABLE GOVERNING EQUA-TIONS FOR GRAPH DYNAMICAL SYSTEMS WITH INTERPRETABLE NEURAL NETWORKS

Anonymous authorsPaper under double-blind review

000

001

002

004

006

008 009 010

011 012 013

014

015

016

018

019

020

021

022

024

025

026

027

028

029

031

033

035

037

038

040

041

042 043

044

046

047

048

051

052

ABSTRACT

The discovery of symbolic governing equations is a central goal in science, yet it remains a formidable challenge, in particular for graph dynamical systems where the network topology further shapes the system behavior. While artificial intelligence offers powerful tools for modeling these dynamics, the field lacks a rigorous, comparative benchmark to assess the true scientific utility of the discovered laws. This work establishes the first rigorous benchmark for this task, moving beyond simple fitting metrics to evaluate discovered laws on their long-term stability and, critically, their out-of-distribution generalization to unseen graph topologies. We introduce the Graph Kolmogorov-Arnold Network (GKAN-ODE), an architecture tailored for this domain, and propose a structure-aware symbolic regression method to leverage its inherent interpretability. Across a suite of synthetic and real-world graph dynamical systems, we demonstrate that symbolic models extracted from neural architectures, particularly our GKAN-ODE, achieve stateof-the-art performance and generalize to unseen networks, significantly surpassing existing baselines. This work presents the first systematic benchmark in this domain, clarifying the expressivity-interpretability trade-offs and offering a pathway from observational data to fundamental physical understanding, providing a critical new tool for data-driven discovery in network science.

1 Introduction

The pursuit of scientific knowledge is undergoing a profound transformation, driven by the confluence of vast datasets and sophisticated computational tools. In this "Fourth Paradigm" of science (Hey et al., 2009), Artificial Intelligence (AI) promises not only to accelerate discovery but to fundamentally change its nature (Wang et al., 2023). The vision extends beyond creating models with high predictive accuracy; the true frontier lies in developing AI that can help us understand the world, unveiling the underlying principles and causal mechanisms that govern complex phenomena (Camps-Valls et al., 2023). This ambition, however, is often hindered by the "black-box" nature of deep learning models, whose internal workings are largely opaque, creating a barrier between computational power and human understanding (Rudin, 2019).

This is challenging especially in the study of *graph dynamical systems* (Barrat et al., 2008). These systems, where entities interact with each other according to a network structure, are ubiquitous in science, from gene regulatory networks and neural circuits to the spread of epidemics and social dynamics (Barabási, 2013). While we can often observe their evolution, the fundamental laws governing their behavior frequently remain unknown and are heavily dependent on the specific graph instance. Our central objective is to move beyond mere simulation by discovering the symbolic, governing *Ordinary Differential Equations* (ODEs) that dictate their evolution directly from observational data.

Symbolic Regression (SR) (Makke & Chawla, 2024) emerges as the natural instrument for this task. While traditional evolutionary algorithms and modern sparsity-based frameworks have laid crucial groundwork, the advent of deep learning has opened new possibilities. Neural Networks (NNs), with their ability to approximate arbitrary nonlinear functions, can learn the underlying dynamics with

high fidelity. However, this expressivity typically comes at the cost of interpretability, requiring a separate, post-hoc SR step to distill symbolic knowledge from the opaque models (Cranmer, 2023).

Despite these advances, a critical gap persists in the literature. The landscape of neural-based equation discovery for graph dynamics is fragmented, with various approaches proposed but no systematic, comparative assessment of their performance in different conditions. Researchers seeking to apply these powerful tools lack a clear reference on which architecture to choose, how to implement it, and how to evaluate the scientific plausibility of the discovered equations. Furthermore, the potential of a novel and interpretable-by-design architecture like *Kolmogorov-Arnold Networks* (KANs) by Liu et al. (2025) remains unexplored in this field, despite their demonstrated potential for scientific discovery in other domains (Liu et al., 2024; Koenig et al., 2024).

This paper aims to fill this gap. We present a rigorous, comparative study designed to unveil the actual performance of neural-based models for equation discovery on graph dynamical systems. Our contributions are fourfold:

- 1. We provide a rigorous and reproducible benchmark of state-of-the-art methods, including a leading sparse regression algorithm and *Multilayer Perceptron*-based architectures (MLPs). By making our code and experimental setup publicly available, we establish a firm baseline for future research ¹.
- 2. We introduce the Graph KAN-ODE (GKAN-ODE), a novel adaptation of Kolmogorov-Arnold Networks for graph dynamics. We enhance the standard architecture with hyperparameter-free multiplicative nodes to better capture physical interactions and propose a principled, structure-aware *Spline-Wise* symbolic regression algorithm to distill faithful formulas directly from KAN architectures.
- 3. We conduct extensive experiments on both synthetic systems with known ground truths, and challenging real-world epidemic data. Our evaluation hinges on a stringent long-term trajectory rollout metric, which assesses the stability of the discovered laws that goes beyond simple one-step prediction accuracy. Moreover, we demonstrate that the learned symbolic models generalize effectively to out-of-distribution settings on unseen scenarios, highlighting their robustness and scientific plausibility.
- 4. We offer a critical analysis of the expressivity-interpretability trade-off. By comparing the symbolic equations extracted from different architectures, we provide practical observations for researchers, clarifying how model choice impacts the complexity and scientific plausibility of the discovered laws.

This work, therefore, serves as both a methodological contribution and a comprehensive benchmark guide, aiming to equip the scientific community with the tools and insights needed to turn observational data into a fundamental understanding of complex systems.

2 RELATED WORKS

2.1 Symbolic Regression for Scientific Discovery

Symbolic regression is a methodology for discovering mathematical expressions from data. Unlike standard regression, which fits parameters to a predefined model, SR searches the space of possible expressions $f_{SYM} \in \mathcal{F}$ to find one that optimally balances predictive accuracy and simplicity. Formally, a SR method takes a dataset of input—output pairs $\{(x,y) \mid y=f(x)\}$ and gives a symbolic approximation of f, i.e., $SR: \{(x,y)\} \mapsto \hat{f}_{SR} \approx f$.

Historically, this field was dominated by evolutionary methods like *Genetic Programming* (GP) (Schmidt & Lipson, 2009; Cranmer, 2023), which, while powerful, often faces scalability challenges. A prominent alternative is the *Sparse Identification of Nonlinear Dynamics* (SINDy) framework (Brunton et al., 2016), which recasts equation discovery as a sparse regression problem over a library of candidate functions. For network systems, TPSINDy extends this by modeling the system's dynamics as a two-part sparse regression problem, finding separate expressions for the self-dynamics and interaction components (Gao & Yan, 2022).

¹Anonymized code available at https://anonymous.4open.science/r/Kan-for-Interpretable-Graph-Dynamics-4499/README.md

2.2 DEEP LEARNING FOR EQUATION DISCOVERY ON GRAPHS

One of the first attempts to leverage NNs to learn analytical expressions was the development of equation learner (EQL) networks (Martius & Lampert, 2017), in which non-linear activation functions are replaced by primitive functions, analogously to SR. Another remarkable work is AI-Feynman (Udrescu et al., 2020), an algorithm that combines SR and NN fitting with a suite of physics-inspired techniques, that outperformed previous benchmarks. A pivotal contribution by Cranmer et al. (2020) showed that Graph Neural Networks (GNNs) can effectively learn the dynamics of physical systems of particles, and their learned latent representations can then be distilled into symbolic expressions via post-hoc SR. The recent Learning Law of Changes (LLC) framework (Hu et al., 2025) advances this approach for graph dynamical systems. It employs separate MLPs to model the self-dynamics and interaction terms (with an explicit multiplication bias) and then parses them into symbolic form using a pre-trained transformer. Their results demonstrate significant performance gains over prior SR techniques for network dynamics, establishing a key state-of-the-art contribution. However, these methods rely on standard MLPs, whose opaque nature complicates the extraction of interpretable symbolic forms, necessitating a model-agnostic, post-hoc SR step.

2.3 KOLMOGOROV-ARNOLD NETWORKS: A PATH TOWARDS INTERPRETABILITY

KANs (Liu et al., 2025) have a fundamentally different architecture than MLPs: they place learnable, univariate activation functions, parametrized as splines ϕ , on the network's edges, while nodes simply perform summation. This design shifts complexity from matrix multiplications and nonlinear activations to a set of univariate functions that can be individually visualized, analyzed, and symbolically regressed. Further technical details can be found in the original paper or in the Appx. A.1. The potential of KANs for scientific discovery has been demonstrated in learning PDE solutions (Liu et al., 2024) and discovering physical laws in dynamical systems without explicit interaction structure (Koenig et al., 2024). However, to our knowledge, KANs have not yet been applied to discover the governing equations of graph dynamical systems, where network topology drives the evolution of node states over time. Their use has been limited to other graph-based tasks (Bresson et al., 2025), not the specific challenge of discovering underlying temporal dynamics.

3 METHODS

This section details our proposed framework for equation discovery. We first establish the formal context for our work, defining graph dynamical systems. Next, we describe the general neural training pipeline, then introduce our Graph KAN-ODE (GKAN-ODE) architecture, and finally outline the symbolic regression procedures and evaluation protocol.

3.1 MATHEMATICAL FORMULATION AND NOTATION

The systems under investigation are graph dynamical systems, or dynamical processes on complex networks. Such a system is defined by a graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$, where \mathcal{V} is a set of N nodes (or components) and \mathcal{E} a set of edges representing their interactions. The state of each node $i\in\mathcal{V}$ at time $t\in\{0,\ldots,T\}$ is described by a vector $\mathbf{x}_i(t)\in\mathbb{R}^d$, while the whole system state is defined as $\mathbf{X}(t)\in\mathbb{R}^{N\times d}$. The graph topological structure can be represented by the adjacency matrix $A\in\mathbb{R}^{N\times N}$, where each entry denotes the connection strength between nodes i and j, and $A_{ij}=0\iff e_{ij}\notin\mathcal{E}$. As in related works, we focus on graphs with *static* topology, where $\forall t\ A(t)=A$, and in a time-invariant context in which the temporal dynamics of a node $\mathbf{x}_i(t)$ are described by an autonomous ODE:

$$\frac{d\mathbf{x}_i}{dt} = f\left(\mathbf{x}_i, \{\mathbf{x}_j\}_{j \in \mathcal{N}(i)}\right) = \dot{\mathbf{x}}_i \quad \forall t,$$
(1)

where $\mathcal{N}(i)$ denotes the neighborhood of node i. For clarity, we will omit the explicit time dependence of $\mathbf{x}_i(t)$ hereafter, unless when denoting data points. Following the principle of universality in network dynamics (Barzel & Barabási, 2013) for pairwise interactions, the governing function f can be decomposed into two fundamental components: an intrinsic, self-dynamics function $H: \mathbb{R}^d \to \mathbb{R}^d$, and an interaction function $G: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ that aggregates effects from

neighboring nodes. The dynamics of any node i can thus be expressed as:

$$\dot{\mathbf{x}}_i = H(\mathbf{x}_i) + \sum_{j=1}^N A_{ij} G(\mathbf{x}_i, \mathbf{x}_j). \tag{2}$$

The primary objective of this work is to discover the symbolic forms of both H and G from discrete-time observations $\{\mathbf{X}(t)\}_{t=0}^T$. Models and estimated quantities are denoted with a hat, e.g., \hat{H} , $\hat{\mathbf{x}}_i$.

3.2 LEARNING DYNAMICS ON GRAPHS WITH NEURAL MODELS

Our primary data consists of time series of graph states $\{\mathbf{X}(t)\}_{t=0}^T$, representing discrete measurements of an underlying continuous process. As a prerequisite for learning, we require an estimate of the instantaneous rate of change, the time derivative $\dot{\mathbf{X}}(t)$. We compute a numerical value of the time derivative for each node \mathbf{x}_i using the five-point stencil method (Gao & Yan, 2022), a choice that balances accuracy with robustness to noise in the observational data. This yields a corresponding sequence of derivative evaluations $\{\dot{\mathbf{X}}(t)\}_{t=0}^T$. We then train a neural framework to learn the mapping from the system's state $\mathbf{X}(t)$ to its derivative $\dot{\mathbf{X}}(t)$. Following the decoupled formulation in Eq. 2, we parameterize the self-dynamics H and interaction dynamics G with two distinct neural networks, \hat{H} and \hat{G} respectively. The models are trained via gradient descent to minimize the Mean Absolute Error (MAE) loss function between the numerically estimated derivatives $\dot{\mathbf{X}}(t)$ and the model's predictions $\dot{\mathbf{X}}(t)$ over the entire training set.

3.3 GRAPH KOLMOGOROV-ARNOLD NETWORKS FOR ODE DISCOVERY

We propose and assess a novel approach, the GKAN-ODE framework, where functions \hat{H} and \hat{G} are parameterized by distinct KANs. In line with the principle that physical laws are often sparse (Brunton et al., 2016), we include the KAN-specific L^1 sparsity penalty (Liu et al. (2025) and Appx. A.1) to encourage both \hat{H} and \hat{G} networks to prune inactive splines.

To better capture the multiplicative relationships common in physical dynamics, we further enhance the standard KAN architecture. While prior work has introduced dedicated multiplication layers between KAN layers (Liu et al., 2024), this adds structural hyperparameters requiring prior knowledge or extensive tuning. To circumvent this, we propose a more integrated extension where multiplication occurs within each KAN layer. Specifically, for a KAN layer with d_{out} output neurons, we designate half $\lceil d_{out}/2 \rceil$ as standard additive nodes and the remaining $\lfloor d_{out}/2 \rfloor$ as multiplicative nodes. This design allows the model itself, guided by data and sparsity, to learn the appropriate functional form (additive, multiplicative, or a combination) without additional hyperparameters. Our empirical findings, detailed in Appx. A.10, confirm that sparse training effectively prunes multiplicative nodes when the dynamics are purely additive and retains them when they are essential, leading to improved performance over the original architecture.

3.4 Symbolic Regression Procedures

Once a neural model is trained, we extract symbolic formulas using two distinct strategies: a model-agnostic, black-box approach and a structure-aware, white-box approach exclusive to KANs.

3.4.1 BLACK-BOX SYMBOLIC REGRESSION

A black-box SR method takes data and a model as input, and produces symbolic expressions approximating the model predictions. Notably, this procedure treats the models as opaque functions, making it applicable to any machine learning method. In our case, given the trained neural networks \hat{H} and \hat{G} , we first generate input-output pairs by performing a forward pass over the training data: $\{\mathbf{x}_i(t), \hat{H}(\mathbf{x}_i(t))\}$, and $\{(\mathbf{x}_i(t), \mathbf{x}_j(t)), \hat{G}(\mathbf{x}_i(t), \mathbf{x}_j(t))\}$ for all interacting pairs. We then fit a separate SR model to each set to obtain symbolic expressions \hat{H}_{SR} and \hat{G}_{SR} :

$$SR(\{\mathbf{x}, \hat{H}(\mathbf{x})\}) = \hat{H}_{SR}, \quad SR(\{(\mathbf{x}_i, \mathbf{x}_j), \hat{G}(\mathbf{x}_i, \mathbf{x}_j)\}) = \hat{G}_{SR}.$$
 (3)

The final symbolic model of the full ODE, $f_{SR} \approx \dot{\mathbf{x}}_i$, is constructed by composing these two discovered expressions according to the governing structure of Eq. 2.

3.4.2 Spline-Wise Symbolic Regression for KANs

The architecture of KANs enables a more granular and transparent approach: instead of regressing on the network's aggregate output, we can distill expressions from its elementary components, i.e., the univariate spline activations ϕ . To fully leverage the transparent structure of KANs, we propose a novel *Spline-Wise* (SW) symbolic regression algorithm for KAN-based models, that systematically converts a trained KAN into a fully symbolic equation. While drawing inspiration from prior work (Liu et al., 2025), our procedure incorporates a principled trade-off between expression complexity and accuracy. The procedure is as follows:

- 1. Affine Function Fitting. Given a trained KAN, let \mathcal{S} be the set of all its spline activations after pruning. For each spline $\phi \in \mathcal{S}$, we test its fit against a library \mathcal{F} of candidate univariate symbolic functions. For each candidate function $f \in \mathcal{F}$, we find the optimal affine transformation parameters $\theta_{f,\phi}^* = (a,b,c,d)$ by non-linear least squares, that minimize the squared error between the spline's output and the transformed candidate function, $f_{\phi}(x;\theta) = a \cdot f(b \cdot x + c) + d$, over a training set.
- 2. Complexity-Penalized Function Selection. For each spline, we must now select the best symbolic representation from the fitted candidates. We search for the function $f_{\phi}(x;\theta_{f,\phi}^*)$ that minimizes a penalized error, balancing approximation accuracy with structural complexity. Specifically, let Γ be a range of regularization hyperparameters. For each $\phi \in \mathcal{S}$ and $\gamma \in \Gamma$, we search for the function $f \in \mathcal{F}$ that minimizes:

$$f_{\phi,\gamma}^* = \arg\min_{f \in \mathcal{F}} \left[\text{MSE}\left(\phi(x), f_{\phi}(x; \theta_{f,\phi}^*)\right) + \gamma \cdot \text{Complexity}(f, \theta_{f,\phi}^*) \right]$$
(4)

where MSE is the Mean Squared Error, and Complexity $(f, \theta_{f,\phi}^*)$ denotes the structural complexity of f, defined as the amount of its operators.

- 3. Pareto-Optimal Formula Selection. The previous step yields a set of $|\Gamma|$ candidate symbolic functions for each spline, representing a Pareto front of accuracy versus complexity. We automatically select the optimal function for each spline f_{ϕ}^* by identifying the expression with the highest performance-complexity score, defined as the negative gradient of the log-MSE with respect to complexity (Cranmer, 2023). Its maximum isolates the point at the Pareto curve where the gain in accuracy for an increase in complexity is the highest.
- 4. **Symbolic Model Reconstruction.** Finally, we replace each spline ϕ in the trained KANs \hat{H} and \hat{G} with its selected symbolic counterpart f_{ϕ}^* . By composing these elementary functions according to the KANs' architectures, we reconstruct the complete symbolic formula f_{SW} , following the structure of Eq. 2.

The pseudo-code algorithm of the above procedure can be found in Appx. A.6, while in Appx. A.7 we show that this approach achieves a more favorable trade-off between accuracy and formula complexity than the SR method proposed by KAN's authors.

3.5 EVALUATION METRIC

The ultimate test of a discovered dynamical law is its ability to forecast the system's evolution. Our primary performance measure is therefore the MAE between ground-truth trajectories and predictions obtained by numerically integrating the learned symbolic dynamics. Formally, given a sequence of observations $\{\mathbf{X}(t)\}_{t=0}^T$, let \hat{H}_{SR} and \hat{G}_{SR} be the extracted symbolic formulas. Since they describe the structure of an ODE, we can integrate them over any time interval $[t_0, t_m] \subseteq [0, T]$:

$$\hat{\mathbf{x}}_{i}(t_{m}) = \mathbf{x}_{i}(t_{0}) + \int_{t_{0}}^{t_{m}} \left[\hat{H}_{SR}(\hat{\mathbf{x}}_{i}(t)) + \sum_{j=1}^{N} A_{ij} \, \hat{G}_{SR}(\hat{\mathbf{x}}_{i}(t), \hat{\mathbf{x}}_{j}(t)) \right] dt.$$
 (5)

Our assessment begins from a given set of initial conditions $\mathbf{X}(t_0)$ from a test trajectory, which are then used to integrate the symbolic model via Eq. 5 for all subsequent time steps, resulting in a predicted trajectory $\{\hat{\mathbf{X}}(t)\}_{t=t_0+1}^{t_m}$. We then compute the trajectory mean absolute error, MAE_{traj}, between the ground-truth observations and predictions:

$$MAE_{traj} = \frac{\sum_{i=1}^{N} \sum_{t=t_0}^{t_m} |\mathbf{x}_i(t) - \hat{\mathbf{x}}_i(t)|}{N(t_m - t_0 - 1)}.$$
 (6)

This integration is autoregressive, meaning prediction errors at one step are propagated into the next. Consequently, even minor inaccuracies in the discovered equations can compound over time, making the MAE_{traj} a stringent and comprehensive test of a model's long-term accuracy and stability. Furthermore, this metric does not rely on prior knowledge of the true governing equations, thereby making it more suitable for real-world scenarios.

4 EXPERIMENTAL DESIGN

This section outlines the empirical framework for assessing equation discovery methods in graph dynamical systems, detailing the models, datasets, and evaluation metrics for performance, symbolic accuracy, and generalization. The Appendices and source code offer further information on dataset generation, model implementation, optimization, SR algorithms, and hyperparameters, ensuring scientific reproducibility and fairness.

4.1 Models under Assessment

We rigorously and fairly assess a set of distinct state-of-the-art methodologies for inferring the governing equations of dynamical systems on graphs. In addition to the proposed GKAN-ODE model, we test three other approaches: our own implementation of a GMLP-ODE model, the neural architecture of LLC, and the TPSINDy algorithm. The GMLP-ODE serves as the direct MLP-based counterpart of GKAN-ODE, where the two KANs are replaced by MLPs and allow for a controlled comparison between the two architectures. LLC is included as a state-of-the-art neural baseline, notable for an MLP-based architecture that explicitly introduces multiplication in the network's structure for \hat{G} in a manner conceptually similar to GKAN-ODE. Unlike these neural approaches, TPSINDy directly learns sparse symbolic expressions for \hat{H} and \hat{G} from data and represents the leading non-neural approach. For the neural architectures, we utilize SR procedures to extract interpretable equations. As black-box SR, the GP-based tool PySR (Cranmer, 2023) is employed, and the resulting symbolic models are labeled with the suffix "+GP". Similarly, SW fitting applied to our proposed model is referred to as GKAN-ODE+SW.

4.2 INFERENCE ON SYNTHETIC DYNAMICAL SYSTEMS

We first evaluate the models' capacity to recover the precise symbolic form of known dynamics. To this end, we utilize four canonical network dynamical systems, chosen to represent a diverse range of nonlinearities common in scientific models (Barzon et al., 2024): Kuramoto oscillators (KUR), epidemic spreading (EPID), biochemical (BIO) and population (POP) dynamics. We generate these synthetic datasets by integrating the models on a fixed Barabási–Albert (Barabási & Albert, 1999) network, chosen for its scale-free topology representative of many real-world systems. To evaluate robustness against measurement uncertainty, we also create noisy variants of these systems by adding white noise to node states at each time step, under different signal-to-noise ratio (SNR) levels. For all experiments with this setting, models are trained on the first 80% of the temporal observations, with the remaining 20% reserved for validation and hyperparameter tuning.

A crucial component of our methodology is the rigorous selection of the final symbolic model. Recognizing that models may overfit to a specific network instance, and that both the GP-based and the SW fitting procedures are sensitive to hyperparameters, we design a robust validation framework. We generate an additional validation set by simulating the same dynamics on a *new* graph with a different topology and initial conditions. For each candidate symbolic formula produced by the SR algorithms, we compute the trajectories rollout error (Eq. 6) on this out-of-distribution (OOD) validation set. The symbolic model achieving the lowest MAE $_{\text{traj}}$ is selected as the definitive expression representing the underlying ODE. Extending the related works, we aim at assessing the generalization of trained models and extracted equations in a novel context: a final *test set* that includes three unique simulations, each with distinct graph topologies and random initial conditions. We report the MAE $_{\text{traj}}$ averaged over these three test trajectories for both models and formulas, indicating their generalization beyond the training domain.

4.3 INFERENCE ON REAL-WORLD EMPIRICAL DATA

To assess performance on a task with unknown ground truth, we utilize the empirical dataset of epidemic dynamics from Gao & Yan, which captures the early, pre-intervention spreading of the H1N1, SARS, and COVID-19 outbreaks across the global airline network. We train the neural models on the COVID-19 dataset and extract symbolic representations. As a true OOD validation set is unavailable, we select the symbolic expression that yields the lowest MAE $_{traj}$ on the training data itself. This procedure discovers a single, homogeneous equation describing the global average dynamics. To account for country-specific variations, we then fine-tune the coefficients of this discovered symbolic structure for each node, following the idea proposed in prior works (Gao & Yan, 2022; Hu et al., 2025) and detailed in Appx. A.13. Our evaluation focuses on the generalizability of the discovered laws. We investigate whether the symbolic structures learned from COVID-19, with only coefficient fine-tuning, can effectively model H1N1 and SARS outbreaks. For final model assessment, we utilize the long-term trajectory rollout metric, MAE $_{traj}$, and compare it with previous studies using a short-term, single-step prediction metric, MAE $_{traj}$, which relies on the Euler method with ground-truth data rather than prior model predictions, emphasizing short-term accuracy and reducing long-term instability impacts.

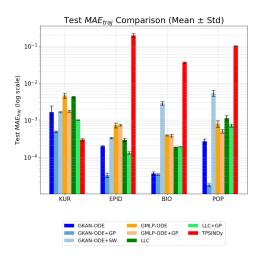
5 RESULTS AND DISCUSSION

5.1 Comparative Performance on Synthetic Systems

Our first key finding, illustrated in Fig. 1 (left), is the superior performance of neural-based architectures over the sparse regression baseline, TPSINDy. The neural models, both before and after symbolic distillation, consistently yield more accurate and stable long-term trajectory rollouts, as measured by the MAEtraj. TPSINDy correctly identifies the KUR dynamics, arguably due to its expressiveness of periodic functions, yet it fails on the other cases. This discrepancy arises because the predefined function libraries do not contain the necessary terms, and since the method is not compositional, it cannot learn them. Conversely, neural-based approaches combined with SR can overcome these limitations by composing complex, nested equations starting from a restricted library of univariate functions and simple binary operators. Among the neural approaches, the models derived from our GKAN-ODE architecture demonstrate remarkable efficacy, and its black-box (GKAN-ODE+GP) symbolic model is consistently among the top performers, achieving the lowest rollout error. While the LLC architecture also performs well, in particular versus its GMLP-ODE counterpart, the GKAN-based models frequently exhibit lower mean error and smaller variance across all time steps (Appx. A.8). Beyond raw performance, GKAN-ODE models are also more parameterefficient than the baselines: Fig. 1 (right) provides a clear visualization of the trade-off between performance (MAE_{trai}) and the number of parameters. The figure promotes once more GKAN-ODE as the most promising choice for efficient equation discovery on graph dynamical systems.

5.2 Symbolic Discovery and Interpretability

As shown in Tab. 1, the black-box GKAN-ODE+GP procedure demonstrates exceptional capability in recovering the ground-truth dynamics, successfully extracting (up to algebraic transformations) the exact symbolic form of the governing equations for all four synthetic systems. The discovered structures are identical to the ground truth, and the fitted coefficients are remarkably precise, validating the entire pipeline from neural training to symbolic distillation. The structure-aware GKAN-ODE+SW approach offers a more direct window into the model's inner workings. Detailed in Tab. 2, this method also successfully identifies the correct underlying dynamics. For the BIO system, it retrieves slightly different coefficients, while for the KUR system it discovers a phase-shifted sine function which is mathematically equivalent to the ground truth. For more complex dynamics like EPID or POP, the SW method sometimes yields expressions with additional, small-coefficient terms. This reflects a trade-off: while the GP approach imposes a stronger global simplicity prior to find the most compact formula, the SW approach provides a more faithful, granular representation of what the individual splines have learned from the data, including minor, data-driven nuances. Both outcomes are valuable forms of interpretation. The equations extracted by TPSINDy, GMLP-ODE+GP and LLC+GP, along with their discussion, can be found in Appx. A.9.



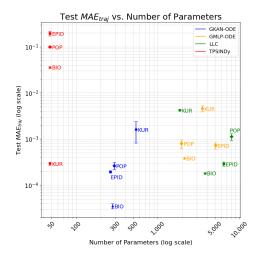


Figure 1: Performance comparison on synthetic dynamics. (Left) Comparison of test MAE_{traj} for both models and the inferred equations. (Right) Test MAE_{traj} and number of parameters of the trained neural-based models and TPSINDy (whose parameters are defined by its symbolic function library). Values are averaged on three test graphs and the standard deviation is reported as errors bars.

When dealing with dynamics with observation noise, neural models are still able to recover competitive expressions under low noise conditions (SNR = 70 dB). However, at higher noise levels (SNR $\leq 50 \text{ dB}$), all models tend to degenerate. More details are available in Appx. A.11.

Table 1: Ground-truth and discovered symbolic equations f_{SR} for the four synthetic dynamical systems learned by the best-validated GKAN-ODE+GP model and rounded to four decimal places. The structural complexity matches between ground-truth and learned models: it is 5 for KUR and POP, and 6 for EPID and BIO.

Dataset	Ground-Truth Equation	GKAN-ODE+GP Discovered Symbolic Expression
KUR	$2 + \frac{1}{2} \sum_{i} A_{ij} \sin(x_j - x_i)$	$1.9992 + \sum_{i} A_{ij} (-0.5005 \cdot \sin(x_i - x_j))$
EPID	$-\frac{1}{2}x_i + \frac{1}{2}\sum_{j}^{j} A_{ij}(1-x_i)x_j$	$-0.4997 \cdot x_i + \sum_j A_{ij} (x_j \cdot (0.5001 - 0.5002 \cdot x_i))$
BIO	$1 - \frac{1}{2}x_i - \frac{1}{2}\sum_j A_{ij}x_ix_j$	$-0.5006 \cdot x_i + 1.0002 + \sum_j A_{ij} (-0.4998 \cdot x_i x_j)$
POP	$-\frac{1}{2}x_i + \sum_j A_{ij} \frac{x_j^3}{5}$	$-0.4999 \cdot x_i + \sum_j A_{ij} (0.2000 \cdot x_j^3)$

Table 2: Best-validated Spline-wise symbolic formulas f_{SW} and their structural complexity for the GKAN-ODE+SW model on the four synthetic dynamical systems.

Dataset	GKAN-ODE+SW Discovered Symbolic Expressions	Complexity
KUR	$1.9991 + \sum_{j} A_{ij} (-0.5005 \cdot \sin(-0.9992 \cdot x_i + 0.9995 \cdot x_j + 3.1373))$	8
EPID	$-0.4988 \cdot x_i + \sum_j A_{ij} (-0.4961 \cdot x_i x_j + 0.4970 \cdot x_j - 0.0022 \cdot x_i + 0.0018)$	10
BIO	$-0.5000 \cdot x_i + 1.0001 + \sum_i A_{ij} (-0.4899 \cdot x_i x_j)$	6
POP	$-0.2862x_i - 0.1744 \tanh(1.4270x_i - 0.0779) - 0.0122 + \sum_j A_{ij}(0.1474x_j^3 + 0.0066x_j^2 + 0.0204x_j)$	16

5.3 DISCOVERY IN REAL-WORLD EPIDEMIC DYNAMICS

In this scenario, the symbolic equations derived for the global average dynamics are reported in Appx A.12. The models yield diverse functional forms, with TPSINDy favoring a logistic-like interaction, while neural architectures learn more complex nonlinearities. This scenario further highlights the critical trade-off between model expressivity and interpretability. Notably, the GKAN-ODE+GP model distills a particularly simple and plausible law, suggesting a linear self-term with an

exponential growth interaction from neighbors. In contrast, the GKAN-ODE+SW method produces a significantly more complex, but fully transparent, expression by directly translating the KAN's internal splines. This presents a choice for domain experts: pursuing the simplest explanatory model (via GP) or analyzing a more faithful, albeit complex, representation of the neural-learned dynamics (via SW). Both are valid pathways to scientific discovery, serving different analytical goals.

Key evaluation rests on the models' stability over time and generalization to unseen data. Fig. 2 contrasts the discovered equations' performance on the COVID-19 test set against their adaptability to H1N1 and SARS dynamics after tuning coefficients. While TPSINDy is competitive in single-step forecasting (MAE $_{eul}$), it leads to catastrophic error accumulation in long-term trajectory rollouts (MAE $_{traj}$). Conversely, all neural-derived laws, and in particular GKAN-ODE models, exhibit strong long-term stability and generalization, crucial for identifying scientific models in complex systems.

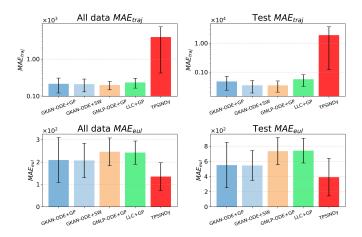


Figure 2: Performance comparison of the symbolic formulas of Tab. 13 averaged on the COVID, H1N1, SARS datasets. Both MAE_{traj} (top) and MAE_{eul} (bottom) are computed on the complete (left) and test (right) datasets.

6 CONCLUSION

This paper rigorously assesses the most prominent AI methods of equation discovery for graph dynamical systems to reveal their true performance. Our findings establish that our proposed GKAN-ODE models significantly outperform sparse regression and MLP-based baselines. KAN-based models further demonstrate a superior balance of predictive accuracy, parameter efficiency, and an architecture inherently amenable to interpretation.

A key contribution of this work lies in the distillation of symbolic knowledge from these models. We have shown how a model-agnostic symbolic regression can effectively recover the ground-truth equations. In parallel, our novel Spline-Wise fitting algorithm provides a transparent and truthful, albeit more granular, symbolic representation of KANs'internal logic. This presents a valuable choice for researchers: a pragmatic path to the most parsimonious symbolic law, or a more detailed, faithful representation of what the model has actually learned from the data.

By establishing a reproducible benchmark and advocating for evaluation based on long-term, out-of-distribution generalization, this work aims to serve as a practical reference and open-source contribution for the interdisciplinary scientific community working on complex systems. It clarifies the state-of-the-art and promotes a human-in-the-loop paradigm where AI acts as a powerful collaborator that generates plausible, testable hypotheses, thereby augmenting human intuition and understanding. Future research should focus on adapting such models for time-dependent systems and with evolving topology, as well as developing tools for dealing with noisy and more complex real-world systems. Ultimately, this study confirms the viability of interpretable neural architectures as powerful tools for the scientific community in the quest to understand the fundamental laws governing complex systems.

REFERENCES

- Albert-László Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.
- Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286 (5439):509–512, 1999.
 - Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge university press, 2008.
 - Baruch Barzel and Albert-László Barabási. Universality in network dynamics. *Nature physics*, 9 (10):673–681, 2013.
 - Giacomo Barzon, Oriol Artime, Samir Suweis, and Manlio De Domenico. Unraveling the mesoscale organization induced by network-driven processes. *Proceedings of the National Academy of Sciences*, 121(28):e2317608121, 2024.
 - Jürgen Braun and Michael Griebel. On a constructive proof of kolmogorov's superposition theorem. *Constructive approximation*, 30:653–675, 2009.
 - Roman Bresson, Giannis Nikolentzos, George Panagopoulos, Michail Chatzianastasis, Jun Pang, and Michalis Vazirgiannis. Kagnns: Kolmogorov-arnold networks meet graph learning. *Trans. Mach. Learn. Res.*, 2025, 2025. URL https://openreview.net/forum?id=03UB1MCAMr.
 - Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.
 - Gustau Camps-Valls, Andreas Gerhardus, Urmi Ninad, Gherardo Varando, Georg Martius, Emili Balaguer-Ballester, Ricardo Vinuesa, Emiliano Diaz, Laure Zanna, and Jakob Runge. Discovering causal relations and equations from data. *Physics Reports*, 1044:1–68, 2023.
 - Miles Cranmer. Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl, May 2023. URL http://arxiv.org/abs/2305.01582. arXiv:2305.01582 [astro-ph, physics:physics].
 - Miles D. Cranmer, Alvaro Sanchez-Gonzalez, Peter W. Battaglia, Rui Xu, Kyle Cranmer, David N. Spergel, and Shirley Ho. Discovering symbolic models from deep learning with inductive biases. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/c9f2f917078bd2db12f23c3b413d9cba-Abstract.html.
 - Ting-Ting Gao and Gang Yan. Autonomous inference of complex network dynamics from incomplete and noisy data. *Nature Computational Science*, 2(3):160–168, 2022.
 - Tony Hey, Stewart Tansley, Kristin Tolle, and Jim Gray. *The Fourth Paradigm: Data-intensive Scientific Discovery*. Microsoft Research, 2009. ISBN 978-0-9825442-0-4. URL https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/.
 - Jiao Hu, Jiaxu Cui, and Bo Yang. Learning interpretable network dynamics via universal neural symbolic regression. *Nature Communications*, 16(1):6226, 2025.
 - Benjamin C Koenig, Suyong Kim, and Sili Deng. Kan-odes: Kolmogorov–arnold network ordinary differential equations for learning dynamical systems and hidden physics. *Computer Methods in Applied Mechanics and Engineering*, 2024.
 - Andrei Nikolaevich Kolmogorov. On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. American Mathematical Society, 1961.

- Ziming Liu, Pingchuan Ma, Yixuan Wang, Wojciech Matusik, and Max Tegmark. Kan 2.0: Kolmogorov-arnold networks meet science. *arXiv preprint arXiv:2408.10205*, 2024.
- Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljacic, Thomas Y. Hou, and Max Tegmark. KAN: kolmogorov-arnold networks. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025.*OpenReview.net, 2025. URL https://openreview.net/forum?id=Ozo7qJ5vZi.
- Nour Makke and Sanjay Chawla. Interpretable scientific discovery with symbolic regression: a review. *Artif. Intell. Rev.*, 57(1):2, 2024. doi: 10.1007/S10462-023-10622-0. URL https://doi.org/10.1007/s10462-023-10622-0.
- Georg Martius and Christoph H. Lampert. Extrapolation and learning equations. In 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Workshop Track Proceedings, 2017. URL https://openreview.net/forum?id=BkqRpOFYe.
- Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.*, 1(5):206–215, 2019. doi: 10.1038/S42256-019-0048-X. URL https://doi.org/10.1038/s42256-019-0048-x.
- Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *science*, 324(5923):81–85, 2009.
- Silviu-Marian Udrescu, Andrew K. Tan, Jiahai Feng, Orisvaldo Neto, Tailin Wu, and Max Tegmark. AI feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/33a854e247155d590883b93bca53848a-Abstract.html.
- Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, Anima Anandkumar, Karianne Bergen, Carla P. Gomes, Shirley Ho, Pushmeet Kohli, Joan Lasenby, Jure Leskovec, Tie-Yan Liu, Arjun Manrai, Debora S. Marks, Bharath Ramsundar, Le Song, Jimeng Sun, Jian Tang, Petar Velickovic, Max Welling, Linfeng Zhang, Connor W. Coley, Yoshua Bengio, and Marinka Zitnik. Scientific discovery in the age of artificial intelligence. *Nat.*, 620(7972):47–60, 2023. doi: 10.1038/S41586-023-06221-2.