

# From GNNs to Symbolic Surrogates via Kolmogorov–Arnold Networks for Delay Prediction

Sami Marouani<sup>1</sup>, Kamal Singh<sup>1</sup>, Baptiste Jeudy<sup>1</sup>, Amaury Habrard<sup>1,3,4</sup>

<sup>1</sup>Université Jean Monnet Saint-Étienne, CNRS, Inst. d’Optique Graduate School, Lab. Hubert Curien, F-42023 Saint-Étienne, France

<sup>3</sup>Institut Universitaire de France (IUF) <sup>4</sup>Inria

Email: <sup>1</sup>{sami.marouani, kamal.singh, baptiste.jeudy, amaury.habrard}@univ-st-etienne.fr

**Abstract**—Accurate prediction of flow delay is essential for optimizing and managing modern communication networks. We investigate three levels of modeling for this task. First, we implement a heterogeneous GNN with attention-based message passing, establishing a strong neural baseline. Second, we propose FlowKANet in which Kolmogorov–Arnold Networks replace standard MLP layers, reducing trainable parameters while maintaining competitive predictive performance. FlowKANet integrates KAMP-Attn (Kolmogorov–Arnold Message Passing with Attention), embedding KAN operators directly into message-passing and attention computation. Finally, we distill the model into symbolic surrogate models using block-wise regression, producing closed-form equations that eliminate trainable weights while preserving graph-structured dependencies. The results show that KAN layers provide a favorable trade-off between efficiency and accuracy and that symbolic surrogates emphasize the potential for lightweight deployment and enhanced transparency.

**Index Terms**—GNNs, KANs, Attention, Message Passing, Symbolic Regression, Delay Prediction, Network Modeling.

## I. INTRODUCTION

Flow delay is a key performance metric in communication networks, influencing congestion control, Traffic Engineering (TE), and Quality of Service (QoS). Network performance prediction has traditionally relied on analytical models and Discrete Event Simulation (DES). Queueing models make strong assumptions, while simulations face scalability and runtime limitations. In recent years, data-driven methods have emerged as an alternative, with Graph Neural Networks (GNNs) showing strong ability to capture dependencies between flows, links, and topologies.

Despite their success, GNN-based models face two main challenges. First, they are often parameter-heavy with tens of thousands of weights, hindering deployment in resource-constrained environments. Second, they remain largely black-box models: while accurate, they provide little transparency about how input features combine to yield predicted delays, limiting trust and interpretability in operational use.

Recently, Kolmogorov–Arnold Networks (KANs) [1] have emerged as a promising alternative to conventional multi-layer perceptrons (MLPs). Grounded in the Kolmogorov–Arnold representation theorem, KANs replace fixed activations with learnable spline functions, enabling smooth and interpretable functional modeling. Their transparent structure makes them particularly suitable for analyzing relationships in network performance data. In this work, we investigate a spectrum of models that balance precision, compactness, and interpretability. Our contributions are:

- A heterogeneous GNN with attention-based message passing as a strong baseline for flow delay prediction.
- **FlowKANet**, a fully KAN-based GNN architecture in which all components are implemented with spline-based operators for greater functional consistency, efficiency, and interpretability. The model incorporates **KAMP-Attn**, a mechanism that leverages KAN operators to compute both feature transformations and attention coefficients within the message-passing process.
- A symbolic distillation of FlowKANet through block-wise regression, yielding compact analytical surrogates that preserve graph dependencies and enable lightweight, interpretable deployment.

## II. RELATED WORKS

### A. Traditional Network Modeling

Analytical queuing models, such as M/M/1 and M/M/k systems, provide closed-form expressions for metrics like average delay. While those models are mathematically elegant, they depend on strong assumptions (e.g., Poisson arrivals, exponential service times). In contrast, machine learning (ML) approaches learn directly from data, allowing more flexible and adaptive modeling of complex network behaviors. DES approaches, on the other hand, can capture detailed protocol dynamics and queueing interactions with high fidelity. They are widely used in academia and industry, but their computational complexity is prohibitive: DES is inherently difficult to parallelize, scales poorly with network size, and remains unsuitable for real-time applications. These limitations motivate the shift toward data-driven ML approaches, which learn performance models directly from traffic traces without restrictive assumptions.

### B. Machine Learning for Networking

Machine learning has been used for various networking tasks such as traffic classification, anomaly detection, routing, and resource allocation [2], [3]. However, these approaches typically treat flows as independent samples, failing to capture the inherent graph structure of communication networks. This motivates the use of GNNs, which represent networks as graphs and capture node, link, and flow dependencies.

1) *GNN-based Models*: GNNs process graph-structured data by iteratively propagating and aggregating information between neighboring nodes through message passing [4], [5]. In general, a GNN layer can be expressed as:

$$h_v^{(k+1)} = \phi^{(k)} \left( h_v^{(k)}, \text{AGG}_{u \in \mathcal{N}(v)} \psi^{(k)}(h_v^{(k)}, h_u^{(k)}, e_{uv}) \right) \quad (1)$$

where  $e_{uv}$  are edge features, AGG is a permutation-invariant aggregation operator (e.g., sum, mean, max),  $h_v^{(k)}$  is the feature of node  $v$  at layer  $k$ ,  $\psi^{(k)}$  is the message function, and  $\phi^{(k)}$  is the update function. This formulation suits communication networks, where the interaction between flows and links can naturally be represented as a graph.

In traffic engineering (TE), GNNs have shown strong potential for resource allocation [2], [6]. TEAL [6] combined a GNN with reinforcement learning and ADMM for WAN optimization, achieving near-optimal results. Building on this direction, FlowAttune [7] applied graph attention to dynamically weight neighboring nodes during message passing. Formally, in a Graph Attention Network (GAT) [8], the message from a neighbor  $u \in \mathcal{N}(v)$  to node  $v$  is weighted by an attention coefficient

$$\alpha_{vu} = \frac{\exp(\text{LeakyReLU}(a^\top [Wh_v \parallel Wh_u]))}{\sum_{k \in \mathcal{N}(v)} \exp(\text{LeakyReLU}(a^\top [Wh_v \parallel Wh_k]))}, \quad (2)$$

where  $h_v$  and  $h_u$  are node features,  $W$  is a learnable weight matrix,  $a$  is the attention vector, and  $\parallel$  denotes concatenation. The updated node representation is then obtained as

$$h'_v = \sigma \left( \sum_{u \in \mathcal{N}(v)} \alpha_{vu} Wh_u \right), \quad (3)$$

where  $\sigma$  is a non-linear activation function.

Attention mechanisms enable GNNs to capture structural dependencies and adapt to dynamic traffic, offering a flexible alternative to fixed aggregation functions. Both methods rely on modeling the network as a flow-link bipartite graph, which facilitates message passing between flows and their associated links. This representation has inspired our own work, where we adopt a similar bipartite modeling strategy for flow delay prediction. GNNs have also been applied to performance prediction. RouteNet [9], [10] and its extensions have estimated end-to-end metrics such as delay and jitter with high accuracy across unseen topologies, while the GNNet Challenge [11] further validated GNN-based performance prediction using real traffic traces, adopting RouteNet-Fermi as its baseline. These works highlight the versatility of GNNs in networking, spanning from traffic allocation to performance prediction. However, existing GNNs still face (i) large parameter counts, leading to high training and inference costs; (ii) limited scalability on large graphs; and (iii) lack of transparency, which constrains their use in operational and real-time network environments.

2) *Kolmogorov–Arnold Networks (KANs)*: KANs [1] are inspired by the Kolmogorov–Arnold representation theorem, which expresses any multivariate function as a composition of univariate ones. KANs replace fixed activations with trainable spline operators. KAN layer computes:

$$y = W \phi(x), \quad (4)$$

where  $\phi(\cdot)$  is a learnable spline function defined on a grid. This design enables smoother function approximation, improved parameter efficiency, and greater transparency of learned transformations. KANs have shown promising results in scientific

machine learning and physics-informed tasks [12]–[14], but remain unexplored in networking or combined with GNNs. This motivates one of the main axes of our study: exploring the integration of KAN layers within GNN architectures for flow delay prediction.

### III. FRAMEWORK FOR FLOW DELAY PREDICTION

We introduce a unified framework that provides a single pipeline from raw network data to graph-based models. Unified here means that the same data representation, preprocessing, and normalization steps are shared across architectures, ensuring that improvements can be attributed to the model design itself. The framework starts with two important steps: constructing a heterogeneous bipartite graph that captures flow-link relationships, and selecting a compact set of relevant features that improves efficiency and generalization. On this foundation, we implement two predictive models: a baseline GNN and a KAN-augmented GNN.

#### A. Graph Construction

The network is represented as a heterogeneous bipartite graph  $\mathcal{G} = (\mathcal{V}_f \cup \mathcal{V}_l, \mathcal{E})$ , where  $\mathcal{V}_f$  denotes the set of flow nodes, each corresponding to a unidirectional flow characterized by features describing its traffic profile (e.g., rate, packet size, burstiness, loss), and  $\mathcal{V}_l$  denotes the set of link nodes, each representing a physical link annotated with its capacity and load. The edge set  $\mathcal{E}$  connects flows to the links they traverse, as determined by routing, and includes both  $(f \rightarrow l)$  and  $(l \rightarrow f)$  directions to enable bidirectional message passing. This construction captures dependency of flows on the sequence of links along their path and vice versa.

*Data extraction and feature engineering*: The raw dataset provides per-flow, per-link, and topology-level information. From this, we build feature vectors for each node type. Flow nodes include basic attributes such as average traffic, number of packets, mean packet size, flow type, and path length, augmented with distributional descriptors that capture burstiness and variability, including inter-packet gap (IPG) statistics (mean, variance, and selected percentiles), packet-size percentiles, packet loss ratio, variance of packet sizes, inter-burst gap (IBG), rate, per-burst bitrate, and type of service (ToS). Each link node is annotated with its capacity and a normalized load, computed as

$$L_\ell = \frac{\sum_{f \in \mathcal{F}(\ell)} \text{traffic}(f)}{C_\ell \cdot 10^9 + \varepsilon}, \quad (5)$$

where  $\mathcal{F}(\ell)$  is the set of flows traversing link  $\ell$  and  $C_\ell$  is the link capacity (Gbps). The concatenation  $[C_\ell, L_\ell]$  forms the feature vector for link  $\ell$ . This enriched representation incorporates not only average values but also distributional characteristics of packet timing and size, which are crucial for accurately modeling flow delay.

#### B. Feature Selection

The flow feature set exceeds one hundred dimensions, many of which are correlated or redundant. To reduce complexity and improve generalization, we apply Sequential Forward

Selection (SFS) with a linear regression proxy and three-fold cross-validation, using mean squared error (MSE) for stable optimization on small delay values. SFS incrementally adds features that maximize performance gain until convergence, yielding a compact subset of 16 flow features (Table I) that balance expressiveness and efficiency.

TABLE I: Selected flow features after Sequential Forward Selection

Feature name	Description
flow_traffic	Average flow bitrate
flow_packets	Number of packets in the flow
flow_packet_size	Mean packet size
flow_type	CBR or MB
flow_length	Path length (hop count)
flow_p10PktSize	10th percentile of packet size
flow_tos	Type of Service (ToS) field
flow_packet_loss	Packet loss ratio (%)
ibg	Inter-burst gap
rate	Flow generation rate
flow_bitrate_per_burst	Average bitrate per burst
flow_ipg_mean	Mean inter-packet gap
flow_ipg_var	Variance of inter-packet gap
IPG percentile P11, P99, P100	11th, 99th, 100th percentile of IPG distribution

*Feature normalization:* To stabilize training and ensure comparable scaling across heterogeneous features, we apply min–max normalization to all flow attributes:

$$\tilde{x}_f = (x_f - \mathbf{m}_{\min}) \odot (\mathbf{m}_{\max} - \mathbf{m}_{\min})^{-1}, \quad (6)$$

where  $\mathbf{m}_{\min}$  and  $\mathbf{m}_{\max}$  are the per-feature minima and maxima computed on the training set. These statistics are stored in model buffers (`min_feat`, `inv_range`) and reused during inference, ensuring consistent scaling across datasets.

### C. Baseline GNN Model Architecture

The baseline architecture is a heterogeneous GNN designed for flow delay prediction. Its workflow is given in Algorithm 1, while the main architectural blocks are shown in Figure 1.

#### Algorithm 1 Forward Pass of the GNN Baseline

```

1: Encode flows:  $h_f^{(0)} \leftarrow \text{MLP}_f(x_f)$ , and links:  $h_\ell^{(0)} \leftarrow \text{MLP}_\ell(x_\ell)$ 
2: for  $k = 1$  to  $K$  do
3:   for each edge  $(f, \ell) \in \mathcal{E}$  do
4:     Compute attention weight  $\alpha_{f\ell}^{(k)}$ 
5:     Compute message  $m_{f\ell}^{(k)} \leftarrow \psi(h_f^{(k)}, h_\ell^{(k)}, \alpha_{f\ell}^{(k)})$ 
6:   end for
7:   for each node  $v \in \mathcal{V}_f \cup \mathcal{V}_\ell$  do
8:     Aggregate messages  $M_v^{(k)} \leftarrow \sum_{u \in \mathcal{N}(v)} m_{uv}^{(k)}$ 
9:     Update node state  $h_v^{(k+1)} \leftarrow \phi(h_v^{(k)}, M_v^{(k)})$ 
10:  end for
11: end for
12: Refine flow embeddings using GRU
13: Fuse  $[h_f^{(K)} \parallel \text{agg}(h_\ell^{(K)})]$ 
14: Predict delay  $\hat{d}_f \leftarrow \text{MLP}_{\text{readout}}(\cdot)$ 

```

- **Flow and Link Encoders:** Raw flow and link features are projected into latent embeddings of dimension  $d_h$ .
- **Message Passing:** Each link aggregates messages from incident flows and each flow from its traversed links. Attention mechanisms compute flow-to-link and link-to-

flow scores, weighting contributions by traffic intensity and congestion.

- **Readout and Prediction:** A gated recurrent unit (GRU) refines flow embeddings across iterations, followed by fusion of flow and aggregated link embeddings. A Softplus layer outputs the final delay prediction.

### D. FlowKANet Model Architecture

To reduce complexity while preserving expressivity, we extend the baseline by replacing all MLPs with KANs layers. The overall message passing structure remains identical, but every transformation block is spline-based. The workflow is summarized in Algorithm 2.

- **Flow and Link Encoders:** Flow and link features are projected into latent embeddings by KAN layers. These initial encoders provide compact hidden representations tailored to each node type.
- **KAMP-Attn (Kolmogorov–Arnold Message Passing with Attention):** Messages are exchanged between flows and links using KAN operators for both transformation and attention, ensuring bidirectional propagation of information across the bipartite graph.
- **Readout and Prediction:** After message passing, flow embeddings are fused with their aggregated link representations and passed through a final KAN block with Softplus activation to predict the per-flow delay.

*KAN-based message passing:* Let  $\mathbf{h}_f^{(k)}$  and  $\mathbf{h}_\ell^{(k)}$  denote the embeddings of flow  $f$  and link  $\ell$  at iteration  $k$ . For each edge  $(f, \ell) \in \mathcal{E}$ , the message is computed using two *shared* spline-based operators: (i) a *transformation operator*  $\mathcal{T}_{f \rightarrow l}^{\text{KAN}}$  that maps flow embeddings into the link space, and (ii) an *attention operator*  $\mathcal{A}_{f \rightarrow l}^{\text{KAN}}$  that produces edge-specific importance weights. These operators are shared across all edges in the same direction. Formally,

$$\tilde{\mathbf{h}}_{f\ell}^{(k)} = \mathcal{T}_{f \rightarrow l}^{\text{KAN}}(\mathbf{h}_f^{(k)}), \quad (7)$$

$$s_{f\ell}^{(k)} = \mathcal{A}_{f \rightarrow l}^{\text{KAN}}(\text{LeakyReLU}(\mathbf{h}_\ell^{(k)} + \tilde{\mathbf{h}}_{f\ell}^{(k)})), \quad (8)$$

$$\alpha_{f\ell}^{(k)} = \frac{\exp(s_{f\ell}^{(k)})}{\sum_{f' \in \mathcal{N}(\ell)} \exp(s_{f'\ell}^{(k)})}. \quad (9)$$

The aggregated message at a node  $v$  is then

$$\mathbf{M}_v^{(k)} = \sum_{u \in \mathcal{N}(v)} \alpha_{uv}^{(k)} \tilde{\mathbf{h}}_{uv}^{(k)}, \quad (10)$$

and the node embedding is updated with a residual connection:

$$\mathbf{h}_v^{(k+1)} = \mathbf{h}_v^{(k)} + \mathbf{M}_v^{(k)}. \quad (11)$$

This mechanism operates in both directions: flows send messages to links via  $\mathcal{T}_{f \rightarrow l}^{\text{KAN}}$  and  $\mathcal{A}_{f \rightarrow l}^{\text{KAN}}$ , while links send messages back to flows through distinct operators  $\mathcal{T}_{l \rightarrow f}^{\text{KAN}}$  and  $\mathcal{A}_{l \rightarrow f}^{\text{KAN}}$ . Each operator is shared across all edges of its respective direction, ensuring consistency and avoiding edge-specific parameterization.

*Concise single-step composition:* Combining transformation, aggregation, and fusion, the per-flow output after  $K$

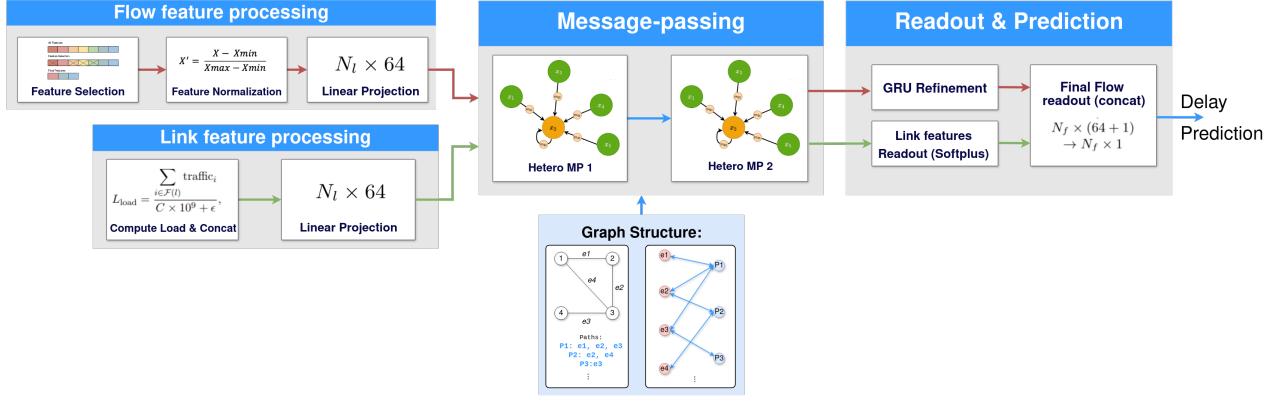


Fig. 1: Baseline GNN architecture with heterogeneous message passing, attention, and GRU refinement.

## Algorithm 2 Forward Pass of the FlowKANet

```

1: Encode flows and links:  $\mathbf{h}_f^{(0)} \leftarrow \text{KAN}_f(x_f)$ ,  $\mathbf{h}_\ell^{(0)} \leftarrow \text{KAN}_\ell(x_\ell)$ 
2: for  $k = 1$  to  $K$  do
3:   for each edge  $(f, \ell) \in \mathcal{E}$  do
4:      $\tilde{\mathbf{h}}_{f\ell}^{(k)} \leftarrow \mathcal{T}_{f \rightarrow \ell}^{\text{KAN}}(\mathbf{h}_f^{(k)})$ 
5:      $\alpha_{f\ell}^{(k)} \leftarrow \mathcal{A}_{f \rightarrow \ell}^{\text{KAN}}(\text{LeakyReLU}(\mathbf{h}_\ell^{(k)} + \tilde{\mathbf{h}}_{f\ell}^{(k)}))$ 
6:   end for
7:   for each node  $v \in \mathcal{V}_f \cup \mathcal{V}_\ell$  do
8:      $\mathbf{M}_v^{(k)} \leftarrow \sum_{u \in \mathcal{N}(v)} \alpha_{uv}^{(k)} \tilde{\mathbf{h}}_{uv}^{(k)}$ 
9:      $\mathbf{h}_v^{(k+1)} \leftarrow \mathbf{h}_v^{(k)} + \mathbf{M}_v^{(k)}$ 
10:  end for
11: end for
12:  $\mathbf{c}_f^{(K)} \leftarrow \sum_{\ell \in \mathcal{N}(f)} \mathbf{h}_\ell^{(K)}$ 
13:  $\mathbf{z}_f^{(K)} \leftarrow g_{\text{fuse}}[\mathbf{h}_f^{(K)}; \mathbf{c}_f^{(K)}]$ 
14: Predict delay  $\hat{d}_f \leftarrow g_{\text{final}}(\mathbf{h}_f^{(K)} + \mathbf{z}_f^{(K)})$ 

```

rounds of bidirectional message passing can be expressed as

$$\hat{d}_f = g_{\text{final}}\left(\mathbf{h}_f^{(K)} + g_{\text{fuse}}[\mathbf{h}_f^{(K)}; \mathbf{c}_f^{(K)}]\right), \quad (12)$$

where  $\mathbf{h}_f^{(K)}$  and  $\mathbf{h}_\ell^{(K)}$  are the final flow and link embeddings, and  $\mathbf{c}_f^{(K)} = \sum_{\ell \in \mathcal{N}(f)} \mathbf{h}_\ell^{(K)}$  is the aggregated link context. Here,  $g_{\text{fuse}}(\cdot)$  and  $g_{\text{final}}(\cdot)$  denote the KAN fusion and readout blocks, respectively, with the latter ending in a Softplus activation to ensure non-negative delay predictions.

## E. Symbolic Surrogate Models

Although the KAN-augmented GNN is lighter than the MLP baseline, it still contains many trainable parameters. To further reduce deployment overhead, we distill the trained FlowKANet into symbolic surrogate models, replacing each KAN block with an analytical expression that approximates its learned mapping, yielding a fully symbolic pipeline from input features to predicted delay. We employ PySR [15], [16], combined with Optuna-based hyperparameter optimization, to discover compact analytical expressions that closely match the outputs of the corresponding KAN operators.

*a) Sequential block-wise search:* The symbolic distillation is performed progressively, one block at a time, following the network structure. At each step, previously symbolized equations are frozen while downstream components remain neural. This ensures consistency of symbolic dependencies

and yields a coherent chain of analytical transformations. The procedure can be summarized as follows:

- 1) For each block  $b$ , freeze all previously symbolized blocks and keep downstream blocks neural.
- 2) Fit a PySR regressor to approximate the KAN output of  $b$ , while Optuna tunes PySR hyperparameters (population size, mutation rate, parsimony).
- 3) Evaluate each candidate expression inside the full model on a validation subset, fix the best formula, and proceed to the next block  $b+1$ .

*b) Final surrogate pipeline.:* Once all blocks have been symbolized, we obtain a fully analytical model that respects the underlying graph structure. Formally, the surrogate prediction takes the form:

$$\hat{d}_f = \mathcal{E}_{\text{symbolic}}(x_f, [C_\ell, L_\ell], \{\mathcal{N}(f), \mathcal{N}(\ell)\}), \quad (13)$$

where  $\mathcal{E}_{\text{symbolic}}$  denotes the composed surrogate equations,  $x_f$  are the selected flow features,  $[C_\ell, L_\ell]$  are link descriptors, and  $\{\mathcal{N}(f), \mathcal{N}(\ell)\}$  encodes the neighborhood relations in the bipartite flow-link graph. In other words, the symbolic surrogate maintains the same message-passing dependencies as the neural FlowKANet: flow delay predictions depend not only on local features but also on the aggregated symbolic contributions of neighboring links and flows. This yields a fully analytical yet graph-aware predictor that mirrors the inductive bias of the original architecture while eliminating the need for neural inference.

## IV. PERFORMANCE EVALUATION

### A. Experimental Setup

We use the GNN Challenge dataset [11], which provides realistic topologies and flow-level traces for graph-based performance prediction. All experiments employ the heterogeneous bipartite graph representation described in Section III. The dataset is accessed through an API that exposes per-flow, per-link, and topology-level features, and is split into 3,511 training graphs (80%) and 878 test graphs (20%).

### B. Hyperparameter Search with Optuna

We employ Optuna to automatically select the main architectural hyperparameters of the FlowKANet. The search space

includes: the hidden dimensions of flow and link embeddings, the number of message-passing layers  $K$ , KAN parameters (grid size  $G$ , spline order  $k$ , and scaling  $\sigma$ ), as well as dropout rate, learning rate, and activation configuration.

TABLE II: Best KAN parameters per block (Optuna).

Block	Grid $G$	Order $k$	Scale $\sigma$
flow_init	9	3	0.93
link_init	7	5	1.66
flow → link (i=0)	5	3	0.55
flow → link (i=1)	6	4	0.70
flow → link (i=2)	8	4	0.82
link → flow (i=0)	7	3	0.73
link → flow (i=1)	7	5	0.77
link → flow (i=2)	10	3	0.33
fuse	6	5	1.15
final	10	5	2.28

We tested several activation functions (ReLU, SiLU, Softplus, Tanh) and four placement strategies: **final\_only** (after the fusion block), **except\_mp** (all blocks except message passing), **all** (every KAN block), and **no\_activation** (none applied). Regardless of the chosen strategy, the last readout block always applies Softplus to guarantee non-negative flow delay predictions. Each Optuna trial was trained for up to 150 epochs with early stopping on the validation set. The Tree-structured Parzen Estimator (TPE) sampler was used for efficient exploration of the large search space. Table III summarizes the best global hyperparameters, while Table II details the KAN-specific settings for each block.

TABLE III: Best global FlowKANet hyperparameters (Optuna).

Parameter	Best Value	Description
Flow hidden dim.	8	Size of flow embedding
Link hidden dim.	2	Size of link embedding
MP layers $K$	3	Number of heterogeneous MP layers
Dropout	0.1	Regularization between layers
Learning rate	0.002	Optimizer step size
Activation type	Tanh	Activation applied to KAN outputs
Activation mode	except_mp	Applied to all blocks except MP

### C. Symbolic Surrogate Search

TABLE IV: Symbolic surrogate search space and constraints.

Component	Options / Constraints
Binary operators	$\{+, -, \times\}, \{+, -, \times, \div\}, \{+, -, \times, \div, \wedge, \vee\}$ with exponent range $[-1, 2]$
Unary operators	$\{\exp, \log,  \cdot \}, \{\exp, \log, \tanh,  \cdot \}, \{\exp, \log, \tan, \tanh,  \cdot \}, \{\exp, \log, \sin, \cos, \tan, \tanh,  \cdot \}$
Expression size	$\text{maxsize} \in \{7, 14, 21, 28, 35\}$
Numerical guards	$\log(\max(x, \varepsilon)), \varepsilon = 10^{-8}$ ; exp clipped to $[-50, 50]$ ; NaN/ $\pm\infty$ replaced

We apply the symbolic distillation procedure described in Section III-E to the trained FlowKANet model. For each KAN block, PySR performs symbolic regression guided by Optuna-based hyperparameter optimization, jointly tuning the

operator sets, tree complexity (`maxsize`), population size, iteration count, and parsimony coefficient to minimize the validation MSE. Expressions exceeding the desired complexity are penalized via the parsimony term, and model selection favors the most accurate symbolic representations. Numerical robustness is enforced through safe log / exp operators and replacement of non-finite values. The search runs in parallel through a shared Optuna RDB and is limited to 250 trials per block. FlowKANet weights remain frozen during distillation. For each block, input–output activations are sampled from the training graphs, with a fraction  $\gamma$ , set to 0.5 in our runs used for symbolic fitting and the remainder for validation within the hybrid model. The best expression per block is validated in context and fixed, progressively replacing all KAN modules to obtain a fully symbolic surrogate of the network.

### D. Predictive Accuracy

We compare the baseline GNN, the KAN-augmented GNN, and the symbolic surrogate distilled from FlowKANet, reporting MSE and  $R^2$  on the test set.

TABLE V: Test-set predictive accuracy.

Model	MSE (lower is better)	$R^2$ (higher is better)
Baseline GNN	38.6358	0.8113
FlowKANet	40.8094	0.8727
Symbolic surrogate	54.8562	0.8290

Both models achieve strong predictive power on the GNN dataset. FlowKANet attains a higher  $R^2$  with slightly higher MSE, showing that spline operators capture small-delay variance and remain stable via the Softplus readout. The symbolic surrogate, though less accurate, provides transparent closed-form equations suited for interpretable or lightweight deployment. Figure 2 shows predicted vs. true delays; FlowKANet points align closer to the diagonal, confirming improved variance capture. During symbolic distillation, we tracked the model MSE as each KAN block was replaced (“progressive hybrid”). Figure 3 shows that replacing early encoders and mid-level message-passing blocks causes mild degradation, while symbolizing the final layers increases error more sharply. Thus, late components are most critical for accuracy, suggesting hybrid deployments, symbolic early/mid blocks with neural readout, offer the best trade-off between interpretability and performance. Overall, KANs bridge conventional GNNs and symbolic models, reducing parameters while retaining accuracy and enabling transparent surrogates. Future work will address accuracy loss in final blocks through symbolic operators for message passing and adaptive hybrid designs.

### E. Parameter Efficiency

Model compactness is critical for deployment in practical settings. Table VI compares the number of parameter across the three evaluated models. FlowKANet reduces the trainable parameter count by nearly 5× compared to the GNN (20k vs. 98k), and the symbolic surrogate eliminates all trainable weights, leaving only 267 fixed constants in the final equations. This progression illustrates a efficiency-interpretability

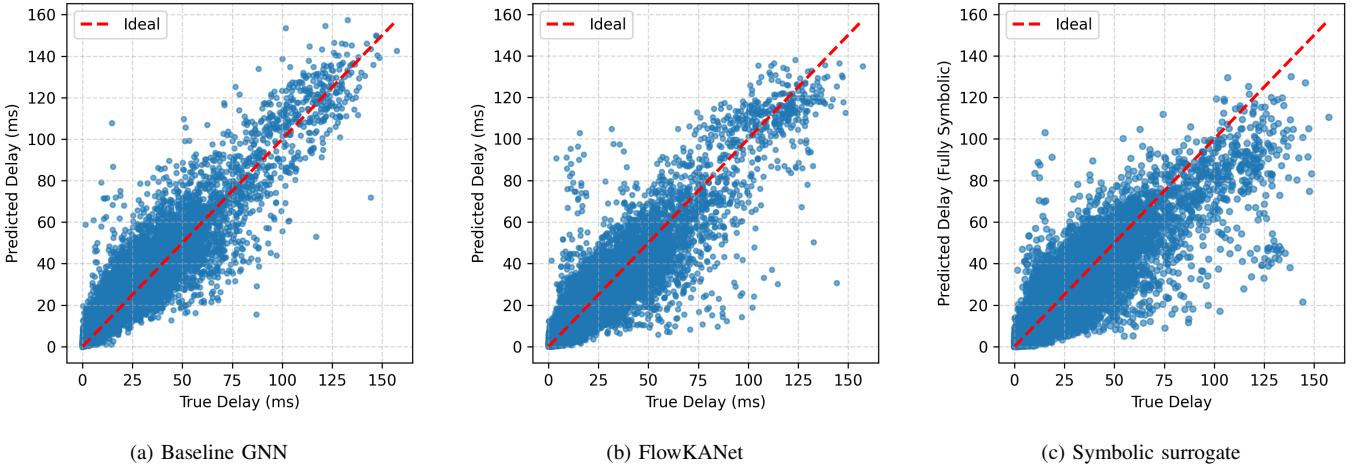


Fig. 2: Predicted vs. true per-flow delay on the test set (878 graphs, containing 13,704 flows).

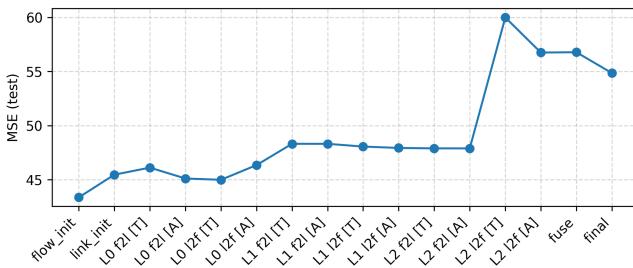


Fig. 3: Progressive-hybrid full-dataset MSE as blocks are symbolized in sequence (lower is better).  $L0-L2$  denote message-passing layers;  $f2f$  and  $l2f$  refer to flow-to-link and link-to-flow directions.

spectrum: from large but flexible GNNs, through compact KAN-augmented models, to purely analytical surrogates suitable for constrained or safety-critical environments.

TABLE VI: Trainable parameter counts of baseline GNN, FlowKANet, and symbolic surrogate.

Model	Model Parameters
Baseline GNN	98,210
FlowKANet	20,094
Symbolic surrogate	267 (Constants in Equations)

## V. CONCLUSION

We have presented a unified framework for flow delay prediction, covering three levels of model design: a heterogeneous GNN with attention-based message passing, FlowKANet model with spline-based transformations, and fully symbolic surrogates distilled from the KAN model. Our experiments show that the FlowKANet achieves comparable accuracy to the GNN while reducing parameter count nearly five-fold. The symbolic surrogates, although less accurate, eliminate trainable parameters entirely and produce transparent closed-form equations that respect the original graph structure. This progression highlights a clear spectrum of trade-offs: from accuracy-focused neural models, to compact spline-based architectures, to symbolic predictors suitable for deployment in resource-constrained or safety-critical environments. In future work, we will focus on deeper interpretation of the learned

transformations, both in the KAN-based model and in the distilled symbolic equations, to provide further insights into how graph-structured dependencies drive flow delay.

## ACKNOWLEDGMENT

This work has been supported by grant ANR-21-CE25-0005 from the Agence Nationale de la Recherche, France for the SAFE project.

## REFERENCES

- [1] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljacic *et al.*, “KAN: Kolmogorov-arnold networks,” in *ICLR*, 2025.
- [2] P. Almasan, J. Suárez-Varela, K. Rusek *et al.*, “Deep reinforcement learning meets graph neural networks: Exploring a routing optimization use case,” *Computer Communications*, vol. 196, pp. 184–194, 2022.
- [3] M. A. Ridwan, N. A. M. Radzi, F. Abdullah, and Y. E. Jalil, “Applications of machine learning in networking: A survey of current issues and future challenges,” *IEEE access*, vol. 9, pp. 52523–52556, 2021.
- [4] J. Zhou, G. Cui, S. Hu, Z. Zhang *et al.*, “Graph neural networks: A review of methods and applications,” *AI open*, vol. 1, pp. 57–81, 2020.
- [5] J. Gilmer, S. S. Schoenholz, P. F. Riley *et al.*, “Neural message passing for quantum chemistry,” in *ICML*, 2017, pp. 1263–1272.
- [6] Z. Xu, F. Y. Yan, R. Singh, J. T. Chiu *et al.*, “Teal: Learning-accelerated optimization of wan traffic engineering,” in *ACM SIGCOMM*, 2023.
- [7] S. Marouani, K. Singh, B. Jeudy, A. Bradai, and A. Habrard, “Advanced traffic engineering in wan using graph attention networks,” in *WiMob*. IEEE, 2024, pp. 657–662.
- [8] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph attention networks,” in *ICLR*, 2018.
- [9] K. Rusek, J. Suárez-Varela, P. Almasan, P. Barlet-Ros, and A. Cabellos-Aparicio, “Routenet: Leveraging graph neural networks for network modeling and optimization in sdn,” *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 10, pp. 2260–2270, 2020.
- [10] M. Ferriol-Galmés, J. Paillas, J. Suárez-Varela *et al.*, “Routenet-fermi: Network modeling with graph neural networks,” *IEEE/ACM transactions on networking*, vol. 31, no. 6, pp. 3080–3095, 2023.
- [11] C. Güemes-Palau, M. F. Galmés, A. Cabellos-Aparicio, and P. Barlet-Ros, “Building a graph-based deep learning network model from captured traffic traces,” *arXiv preprint arXiv:2310.11889*, 2023.
- [12] T. Ji, Y. Hou, and D. Zhang, “A comprehensive survey on kolmogorov arnold networks (kan),” *arXiv preprint arXiv:2407.11075*, 2024.
- [13] S. Rigas, M. Papachristou, T. Papadopoulos, F. Anagnostopoulos, and G. Alexandridis, “Adaptive training of grid-dependent physics-informed kolmogorov-arnold networks,” *IEEE Access*, 2024.
- [14] S. Somvanshi, S. A. Javed, M. M. Islam, D. Pandit *et al.*, “A survey on kolmogorov-arnold network,” *ACM Comp. Surv.*, vol. 58, no. 2, 2025.
- [15] M. Cranmer, “Pysr: High-performance symbolic regression in python and julia,” *Astrophysics Source Code Library*, pp. ascl–2409, 2024.
- [16] ———, “Interpretable machine learning for science with pysr and symbolicregression.jl,” 2023.