

REVIEW

Graph-Mamba: A Survey of Selective State Space Models for Graph Learning

Guangyu Xu^{1,2,*}

¹Future Tech Institute, Guangzhou Huashang University, Guangzhou, China

²School of the Environment, The University of Queensland, Brisbane, QLD, Australia

*Corresponding Author: Guangyu Xu. Email: guangyu.xu@student.uq.edu.au

Received: 27 April 2026; Accepted: 28 July 2026; Published: 28 August 2026

ABSTRACT: The fusion of GNNs and SSMs is creating a new era in the realm of dynamical learning with structures. Graph-Mamba is one of the most promising works in this fast-moving field. With its unique capability to incorporate graph topology and selective dynamics in a stable manner, Graph-Mamba has demonstrated its ability to model intricate relationships. However, the existing research on Graph-Mamba has not been compiled into a coherent form; the theoretical basis and practical implementation of the framework have not been synthesized systematically across different domains. First, we demonstrate the theoretical connection between graph propagation and state space evolution and articulate Graph-Mamba as a hybrid dynamical system defined through selective propagation. We further discuss the particular developments made possible within each application area to which the model is relevant. Regardless of whether spatio-temporal and physical systems, biomedical networks, or heterogeneous graphs are in question, we articulate how selective propagation can be adapted to address specific structural and temporal requirements. In light of this discussion, we formulate the overarching theoretical framework for understanding how Graph-Mamba facilitates selective evolution of graph connectivity and state space evolution through the mediation of contextual selectivity. This survey goes beyond summarizing empirical results and focuses on a more profound methodological change. Specifically, we make a case for viewing Graph-Mamba as a promising paradigm that transforms graph-based learning from a static topological approach to a controllable dynamic one. Essentially, this indicates the convergence of graph theory and dynamical systems that will most likely inspire generalized sequence-structure architectures. Ultimately, Graph-Mamba offers not merely a model design, but a foundational direction for understanding the interplay between selective sequence scanning and structured relational dynamics.

KEYWORDS: Graph neural networks; state-space models; Graph-Mamba

1 Introduction

Since they natively represent the dependency relationships between objects, graphs become a basic building block for modeling in contemporary computer science. Applications span across several different fields that include social behavior analysis, biology, and physics, among others [1–5]. This has led to GNNs becoming the prevailing approach for tackling structured data and achieving state-of-the-art results on problems like node classification, link prediction, molecular property prediction, and spatiotemporal forecasting [6–10]. However, in spite of such progress, the recursion in message passing algorithms used in graph neural networks (GNNs) still faces restrictions from local propagation, which restricts the ability to learn long-term dependencies and increases the probability of over-smoothing when adding depth to neural layers [11–15]. The selective state-space approach used by Mamba has recently been applied to reduce over-smoothing issues in graph neural networks (GNNs) [16]. New entropy-based approaches also shed light

on this problem [17]. Systematic reviews further reveal that over-squashing is a key challenge that can be addressed by Mamba's model of long-term dependency [18]. Attention-based graph transformers also face scalability limits, as quadratic complexity becomes prohibitive in massive graphs [19,20].

Meanwhile, state-space models (SSMs) are changing the landscape of sequence learning and spatial learning [21–23]. Recent research has focused primarily on the Mamba architecture [24]. Mamba abandons traditional attention mechanisms and instead uses input-dependent gating and selective scanning to parameterize continuous-time state evolution. This design uniquely achieves linear time complexity while fully capturing global dependencies. It has been rapidly applied in multiple fields, including natural language processing, computer vision, time series analysis, and structure learning [25–29].

Based on these advancements, recent research has begun to combine graph learning with state-space modeling, resulting in models now known as Graph-Mamba [30,31]. This paper defines Graph-Mamba as any learning framework that explicitly combines graph structure representations (e.g., nodes, edges, adjacency matrices, or relational graphs) with Mamba-based state-space mechanisms (for selective or cyclic state propagation). Essentially, Graph-Mamba unifies the topological expressiveness of graph structures with the dynamic modeling capabilities of state-space systems. This definition distinguishes Graph-Mamba from broader “Mamba-inspired” models that lack explicit graph representations, such as Vision Mamba or Mamba-based image encoders [32–34].

The concept of Graph-Mamba first emerged in 2024, when early studies such as GraphMamba [31] and STG-Mamba [34] introduced state-space scanning into graph neural networks to enhance long-range relational reasoning. These studies further demonstrated that continuous-time state dynamics have potential as an alternative or even replacement for attentional processing in modeling graph dependencies. This Graph-Mamba paradigm has thus quickly gained popularity in applications ranging from geometric reasoning, connectomics, spatiotemporal forecasting, to molecular learning [35–38]. Despite its rapid development, current studies remain at an early stage and face several important limitations. The theoretical behavior of selective state-space propagation on irregular graph topologies is not yet completely understood, especially under dynamic gating and non-stationary graph structures. Empirical evidence is also fragmented across domains, as existing works use heterogeneous datasets, evaluation protocols, graph construction strategies, and serialization mechanisms. Moreover, many Graph-Mamba models rely on domain-specific preprocessing or topology-aware scanning designs, which may affect scalability, reproducibility, and generalization. Therefore, the advantages of Graph-Mamba should be interpreted within these methodological and empirical constraints, and this survey revisits these limitations when discussing broad claims about its theoretical and practical potential.

Research to integrate Graph-Mamba-related papers is currently in an infant stage. To date, only one preprint paper is available entitled, Exploring Graph Mamba: A Comprehensive Survey [39]. However, the usefulness of this paper is limited by the fact that it has not been reviewed, and it uses more taxonomies than theoretical integration in its analysis. Although it successfully categorizes design variations such as hybrid attention-state network and multiresolution embedding, it overlooks the underlying theory connecting the topology of a graph to state space evolution. Additionally, some key subdomains that have seen substantial growth in research have been left out entirely by the preprint paper, namely those involving biomedical graphs, heterogeneity, 3D geometric structures, and structural vision problems.

Therefore, gaps remain in the existing literature. Researchers have not yet attempted to elucidate the relationship between selective state propagation and graph structure, and the role of dynamism in traditional message passing has also been largely overlooked, resulting in significant ambiguity in the theoretical transformation of graph learning techniques.

To fill this gap, this survey employs a theory-driven approach to study Graph-Mamba and its extensions. It outlines the conceptual boundaries and presents graph learning and state-space dynamics in a unified manner. Furthermore, it delves into how the Mamba intelligently scans different types of graphs, such as spatiotemporal graphs, heterogeneous graphs, molecular graphs, geometric graphs, and visual graphs. Therefore, this paper does not rely on comparative analyses of different architectures but focuses on the fundamental theory of graph-state relationships. As shown in the [Fig. 1](#).

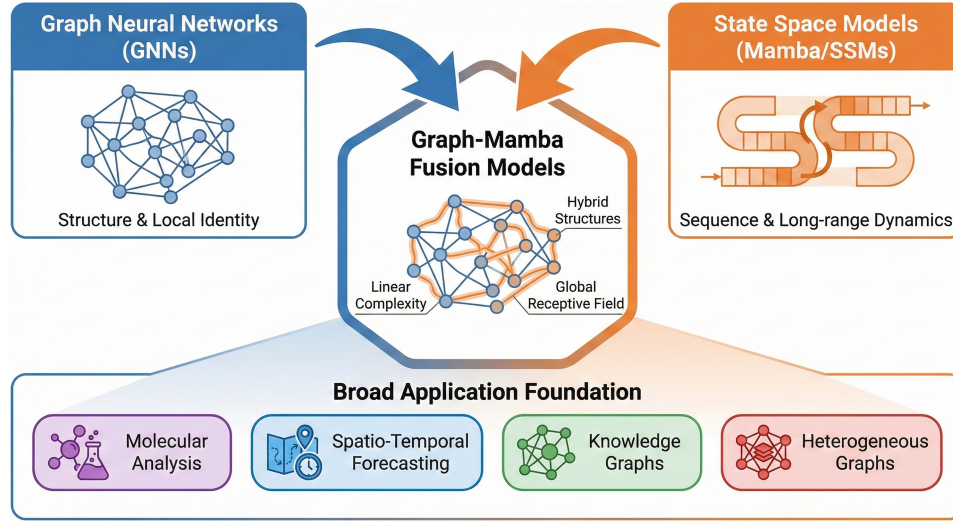


Figure 1: Graph-Mamba models: a unified paradigm for structure and sequence learning.

Compared with existing preliminary surveys on Graph-Mamba, this review aims to provide a more integrated and practically useful overview of the field. Rather than merely listing model variants, we organize Graph-Mamba research from four complementary perspectives. First, we revisit the methodological foundation of Graph-Mamba by explaining how graph message passing, state-space evolution, and Mamba-style selective scanning can be connected within a unified propagation perspective. Second, we examine how Graph-Mamba mechanisms are adapted to different application scenarios, including spatio-temporal systems, biomedical and bioinformatics graphs, heterogeneous graphs, geometric structures, and visual data. Third, we summarize the concrete datasets and empirical settings used in representative studies, thereby helping readers identify the benchmark environments and data structures associated with each domain. Fourth, we discuss representative algorithmic frameworks, including graph construction strategies, serialization mechanisms, selective state propagation, and graph–sequence interaction designs. Through this organization, the survey is intended to help readers understand not only the methodological basis of Graph-Mamba, but also how to locate relevant application domains, datasets, and architectural choices for future research.

This review is systematically written from theory to practice. [Section 2](#) describes the theory behind state-space representation and introduces the criteria used to select the research subjects for this review. Building on this theoretical foundation, [Section 3](#) provides a detailed overview of the classification of Graph-Mamba architectures by explaining the transformation from traditional message passing processes to state-space updates. [Section 4](#) explores the integration of these models with spatiotemporal data, biomedical data, geometric 3D data, and visual data. Finally, [Section 5](#) discusses optimization problems. [Section 6](#) provides a summary and outlook.

2 Literature Search and Screening Process

2.1 Search Strategy and Database Selection

All stages of literature retrieval and research screening strictly followed the PRISMA 2020 standard [40]. This review focuses on academic research that integrates state-space models into graph structures, and the time span of the study is from the date the Graph-Mamba architecture was first published in 2024 to 27 October 2025. Meanwhile this survey searched the Web of Science (WoS) Core Collection, which comprehensively covers literature in the fields of computer science and artificial intelligence, specifically using the SCI-EXPANDED and CPCI-S indexes.

The Boolean query used in the search was: (“Graph” [All Fields]) AND (“Mamba” [All Fields]). Core metadata were harvest-ed for all retrieved records, restricting the language strictly to English. Although the use of the WoS Core Collection may not exhaustively retrieve all emerging studies related to Graph-Mamba, especially because this topic is developing rapidly and some foundational or adjacent works first appear as preprints or conference papers. However, this survey does not aim to construct an unrestricted inventory of all Mamba-related graph studies. Instead, its main analytical structure is organized around representative application domains, data structures, graph construction strategies, and reproducible model settings. For this reason, the formal screening corpus maintains a controlled and reproducible basis for discussing methods, datasets, and empirical settings. Influential preprints or conference papers are cited where necessary for back-ground, conceptual framing, or comparison, but the domain-level qualitative synthesis and representative model discussion are based on the screened corpus that satisfies the inclusion criteria.

2.2 Screening and Selection Process

The selection process was conducted based on the PRISMA protocol for systematic reviews [41], following several steps. After conducting the first search, two reviewers independently reviewed the titles and abstracts of all citations to filter out those articles that had no relevance to the Graph-Mamba system. At this step, those publications that used the term “Mamba” in a non-specialized manner or which addressed topics other than graph-based data were filtered out.

Full-text screening was then carried out on the remaining articles by two independent reviewers. The inclusion criteria required that the articles use graph structural representation as well as Mamba or state space modeling approaches. Any disagreement between the two reviewers during this process was sorted out through consultation to reach a consensus.

Fig. 2 provides a detailed breakdown of the steps followed during the screening process. The process began with the collection of 117 documents from the Web of Science Core Collection. After removing the duplicates, the remaining articles were scrutinized based on the above-mentioned criteria, resulting in a final total of 81 papers.

2.3 Inclusion and Exclusion Criteria

The inclusion and exclusion criteria were based on the methodological and theoretical relevance. Only articles that dealt with the connection between graph topologies and Mamba architectures, which are the two basic concepts in the Graph-Mamba theory, were selected.

Inclusion Criteria:

The final corpus was generated under extremely restrictive inclusion criteria. From a methodological point of view, papers had to make a fundamental use of graph structures, such as relational graphs or adjacency matrices, while using the state-space approach of Mamba. In terms of meta-data, only English-language papers published in SCI-EXPANDED and/or CPCI-S journals within the period of January

2024–October 2025 have been considered. As new architectures for the graph representation of information are developed rapidly nowadays, conference papers with rigorous approaches have been included.

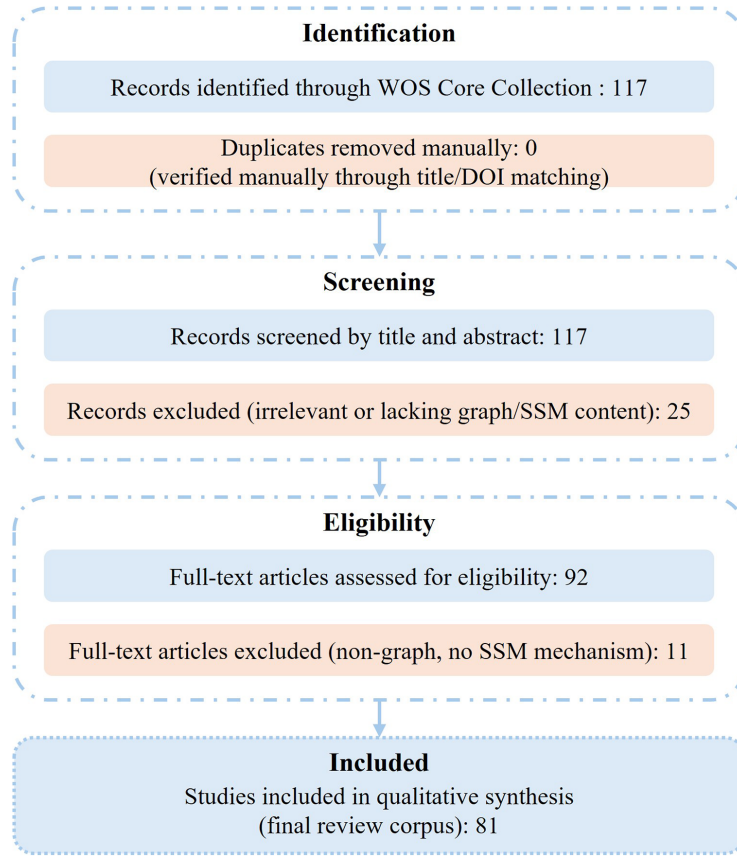


Figure 2: Literature screening process based on PRISMA.

Exclusion criteria:

We also had some exclusion criteria in order not to be misled into including false positive papers. Papers that refer to Mamba without having a state space implementation mechanism were discarded. The same criterion applies to papers that did not use any graph topology or adjacency calculation, regardless of any architectural connections with Mamba. Last, but not least, all the non-empirical works, like editorials or workshop abstracts, as well as articles whose full text was unavailable or had inconsistent metadata have been excluded from the analysis.

To make the inclusion boundary more explicit, we further define what does and does not count as Graph-Mamba in this survey. A model was included only if it simultaneously used an explicit graph representation and a Mamba-style selective state-space mechanism. The graph representation could appear as an adjacency matrix, graph Laplacian, molecular graph, skeleton graph, traffic sensor graph, patch graph, or other relational topology. The Mamba component had to play an active role in propagation, scanning, temporal modeling, or graph–sequence interaction.

Models were excluded when either of these two components was absent. For example, pure Mamba-based language, vision, or time-series models without explicit graph topology were not counted as Graph-Mamba. Similarly, conventional GNNs, graph transformers, or graph convolutional models without Mamba or selective state-space modules were excluded. Borderline cases were judged by whether graph

structure had a computational role in the architecture. If graph information was used only for preprocessing, visualization, or descriptive analysis, the model was excluded. If graph connectivity guided node serialization, topology-aware scanning, adjacency-based propagation, or graph-level feature interaction within a Mamba-based module, the model was included.

Finally, a total of 81 relevant papers have made it through the filtering process and became our source data for the review. These papers will be studied further on depending on their application and architecture types. According to PRISMA protocol, the final corpus is fully replicable for future researches.

3 Theoretical Foundations of Graph-Mamba

3.1 Overview of State-Space Models and the Mamba Mechanism

In fact, the use of state-space models (SSMs), originally introduced into control theory and signal processing before their usage in the realm of deep learning, is much more profound than that. The standard representation of the model implies the following Eq. (1):

$$h_t = Ah_{t-1} + Bx_t, \quad y_t = Ch_t \quad (1)$$

where the state vector at time step t is denoted as $h_t \in R^d$, x_t represents the input vector, y_t the output of interest, and A , B , and C are the learnable transition matrix, input matrix, and output matrix, respectively.

As can be seen from the above formula, this is actually a Markov process, where there is only a linear dependency between elements, and each state depends only on the previous state and its corresponding input. The key characteristic of the state-space model is that its evolution is achieved through recursive linear transformations of variables, thus conforming to the behavior of cyclic states [42,43]. Fig. 3 shows the architecture of the SSM, with its components marked by dashed lines.

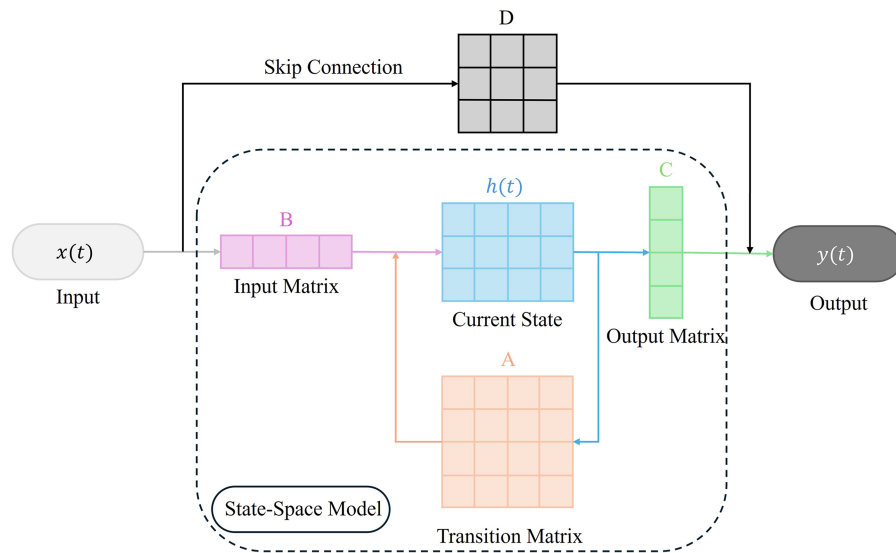


Figure 3: Graphical representation of state-space models.

In continuous-time form, the dynamics of an SSM are often represented by the differential equation given in Eq. (2):

$$\frac{dh(t)}{dt} = Ah(t) + Bx(t) \quad (2)$$

where A governs the temporal evolution of the hidden state. The solution to this differential equation involves matrix exponentials and convolutional integration over input signals, yielding Eq. (3):

$$h(t) = e^{At}h(0) + \int_0^t e^{A(t-\tau)}Bx(\tau)d\tau \quad (3)$$

which reflects how the input sequence drives state transitions in continuous time.

Modern developments in deep learning have led to the reformulation of the traditional approach based on system theory as the state space layer using neural networks. In this case, the matrices A , B , and C are obtained using data as opposed to being pre-specified. It is possible to model long-range dependencies in sequences using a neural state-space layer in a more efficient way than attention mechanisms. In particular, the Structured State-Space Model (S4) [44] introduces a transformation of matrix A to its diagonal representation allowing for an efficient computation in parallel using frequency domain convolutions. This enables large-scale deployment of state-space layers in deep learning. The structure of S4 is shown in Fig. 4.

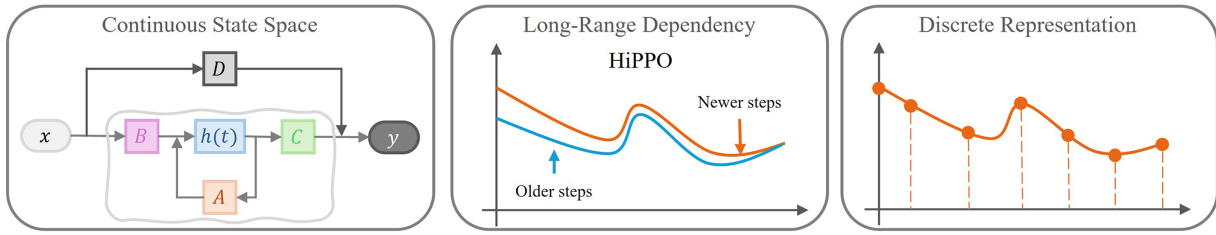


Figure 4: The structure of structured state-space model.

Following up on S4, Mamba [24] proposes a selective, input-conditioned variant of state-space models. In particular, Mamba employs a gating scheme that controls how much new input contributes to the updating of the hidden state at every time step. The formulation for selective-scan is defined as follows Eq. (4):

$$h_t = (1 - g_t) \odot h_{t-1} + g_t \odot f(x_t, h_{t-1}) \quad (4)$$

In the above formula, the gate $g_t = \sigma(W_g x_t)$, which defines the update rate, is learned, where $f(\cdot)$ is a learned state-update function that is typically linear or low-rank, and $\odot$ represents the Hadamard product (element-wise multiplication).

Through input-conditioned gating, the network filters relevant information by retaining relevant memory and adapting to newly acquired information. Since these conditioning operations occur via continuous state-space dynamics instead of token-to-token relations, Mamba processes input linearly, effectively avoiding the quadratic computational complexity of typical self-attention mechanisms.

An interesting characteristic of Mamba's design is that it allows selective scans to mimic the context modeling done by self-attention at an extremely reduced cost, as the explicit pairing of the attention score matrix is avoided through the transformation of state-space dynamics into a convolutional kernel, as shown in Eq. (5):

$$y_t = \sum_{k=0}^t K_{t-k} x_k$$

$$K_{t-k} = C e^{A(t-k)} B \quad (5)$$

In the selective scan approach, attention-based expressiveness is achieved at lower costs using input-based kernel parameterization. Such a structure can easily keep signal flow over trajectories having several thousand steps, thus overcoming the difficulties associated with vanishing gradients in classical recurrent neural networks. To apply sequential processing techniques on non-Euclidean structures, Graph-Mamba approaches use dedicated traversal approaches. In fact, techniques like permutation scanning [45] and order-based methods [46] make the model pay attention to the complex graph connectivity structure. Coupling these methods with graph differential equations [47] yields a powerful continuous dynamical perspective, definitively resolving the tension between discrete node updates and continuous selective flows. Fig. 5 visualizes the core architecture driving these operations.

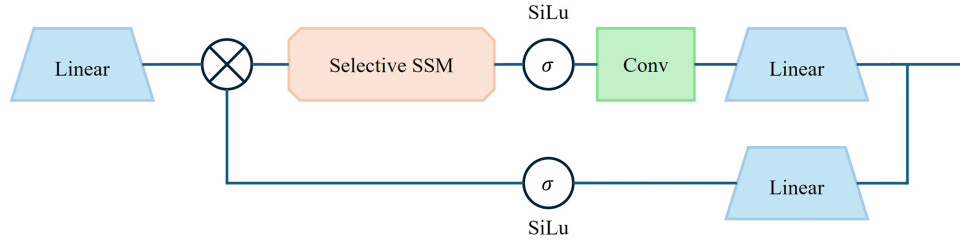


Figure 5: Architecture of the Mamba block.

Mamba, with its low computational overhead from Transformers [48,49], demonstrates superior modeling capabilities and is rapidly being applied to non-sequence domains such as geometric graphs and relational graphs. Its core Selective State-Space (SSM) mechanism reconstructs long sequence modeling through physically inspired mathematical methods. Based on this, the Graph-Mamba framework was developed: it selectively introduces input-awareness into complex topologies, achieving a deep fusion of graph network structure awareness and Mamba's linear-time propagation.

3.2 Graph Representation and Message Propagation

Graph-structured data offers a powerful and flexible framework for modeling entities and their pairwise or higher-order relations [50]. Formally, a graph is defined in Eq. (6):

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, X) \quad (6)$$

where $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ denotes the set of N nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ the set of edges, and $X \in \mathbb{R}^{N \times d}$ the feature matrix containing d -dimensional attributes for each node. The topology of $\mathcal{G}$ is typically represented by an adjacency matrix $A \in \mathbb{R}^{N \times N}$, where $A_{ij} = 1$, if $(v_i, v_j) \in \mathcal{E}$, and $A_{ij} = 0$ otherwise.

Within the scope of weighted or directed graphs, the matrix entries naturally go beyond simple binary labels and take on continuous or asymmetric forms that properly reflect the weight and direction of each edge.

The central problem of graph learning lies in the mapping of topological components, from nodes to full structures, into continuous latent spaces in a way that retains their connectivity. GNNs made progress on this front through the concept of spatial message passing, which allows for the processing of data in its native topological form [51].

The node-level evolution process of the standard graph neural network architecture is formally described in Formula (7):

$$H^{(l+1)} = \sigma(\tilde{A}H^{(l)}W^{(l)}) \quad (7)$$

where $H^{(l)} \in \mathbb{R}^{N \times d_l}$ represents the embeddings of nodes at the l -th layer, $W^{(l)}$ stands for the weight matrix, $\tilde{A} = D^{-\frac{1}{2}} (A + I) D^{-\frac{1}{2}}$ is the adjacency matrix, and $\sigma(\cdot)$ denotes the nonlinearity function.

This equation essentially captures the graph convolution operation itself, whereby each node fuses information from its local neighbors weighted by the network's connectivity before passing it to the next layer using a learnable weight matrix.

More generally, the message-passing process in a GNN can be expressed as Eq. (8) [52]:

$$h_i^{(l+1)} = \phi \left(h_i^{(l)}, \oplus_{j \in N(i)} \psi \left(h_i^{(l)}, h_j^{(l)}, e_{ij} \right) \right) \quad (8)$$

where $h_i^{(l)}$ is the hidden representation of node v_i at layer l , e_{ij} denotes the edge features between nodes i and j , ψ is a message function, $\oplus$ is a permutation-invariant aggregation operator, and ϕ is the node update function.

This theoretical umbrella covers several foundational models, including the Graph Convolutional Network (GCN), GraphSAGE, and Graph Attention Network (GAT) [53–55]. Mechanically, this setup drives information diffusion across local neighborhoods. As we stack deeper layers, the receptive field for any given node naturally pushes outward.

However, conventional GNNs remain heavily constrained by two inherent flaws. The first is the well-documented over-smoothing phenomenon, where endless aggregation dilutes node representations into identical vectors, neutralizing the benefits of deeper networks [56]. The second flaw stems from strict operational locality. Because a single layer only pulls data from one-hop neighbors, modeling long-range dependencies demands deep layer stacking or patchwork solutions like hierarchical pooling and skip connections. Together, these constraints create a relentless tension between capturing complex global structures and maintaining computational viability in graph learning.

Formula (9) formalizes this layer-by-layer propagation as a discrete-time linear system controlled by the topology of the underlying graph.

$$H^{(l+1)} = \mathcal{A} H^{(l)} W^{(l)} \quad (9)$$

where $\mathcal{A}$ represents the normalized adjacency operator.

$$h_t = A h_{t-1} + B x_t \quad (10)$$

From a mathematical perspective, the message passing in GNNs and the discrete state update formula of SSMs (such as Eq. (10)) share a high degree of structural similarity: both generate new states by linearly combining previous states through propagation operators [57]. This structural analogy reveals that the two paradigms differ only in dimension—SSM propagates signals along the time axis, while GNNs disperse signals along the topological structure. From this perspective, we can view graph message passing as a kind of “structured” state transition. This paradigm fusion not only demonstrates the rationale for applying state-space dynamics in graph environments but also provides theoretical support for novel architectures such as Graph-Mamba.

$$H^{(l+1)} = U \Lambda U^T H^{(l)} W^{(l)} \quad (11)$$

where U and Λ be eigenvectors and eigenvalues, respectively, of the graph Laplacian. Eq. (11) shows that transformation follows structurally identical principles as those behind the Fourier transform used in state space models (SSM). The latter uses kernels based on eigenmodes of the transition matrix A . In the end, the

two approaches use linear operators to filter the signals through their own domains—in time in SSMs and in the topological domain in graph approaches.

Graph-Mamba implements precisely such analogy in its structure and, thus, marries the fluid transition state concept inherent to state space models with the inductive bias of graph models. Mechanism of such a marriage is rather straightforward and effective—the standard temporal transition matrix A becomes graph-based operator $\mathcal{A}$. In conjunction with Mamba gating principles, it generalizes the ideas from one-dimensional sequences typical of state space models into relation-aware data processing. Following the idea of such a transformation, in the next section we consider mathematics behind Graph-Mamba, treating the model purely as a continuous dynamic system operating on graph topology, not on time.

3.3 Unifying Graph Propagation and State-Space Dynamics

The theoretical link between graph message passing and state-space modeling is that they both describe processes wherein information is passed iteratively across some kind of structure. State-space models involve iterative information passing through time while graph neural networks involve iterative information passing through nodes on a graph. The Graph-Mamba paradigm integrates these two propagation dimensions—temporal and topological—within a single dynamic formulation in which hidden states evolve not only based on previous states and new inputs, but also in accordance with the structural constraints of the graph topology.

3.3.1 Extending State-Space Models to Graph Domains

To generalize the discrete-time state-space formulation to graph-structured data, we begin with the standard SSM recurrence relation, given in Eq. (12).

$$h_t = Ah_{t-1} + Bx_t \quad (12)$$

where A captures the internal transition dynamics and B encodes the influence of the input signal. For graph data, the latent representation at time t is replaced by a matrix of node-level hidden states, $H_t \in \mathbb{R}^{N \times d}$, where each row corresponds to the hidden representation of a node.

The internal transition matrix A was then generalized to the graph-induced propagation operator $\mathcal{A}$, which controls how the state spreads within the graph's topology. Analogous to the normalized adjacency matrix in GNN, $\mathcal{A}$ can be defined as Eq. (13):

$$\mathcal{A} = \alpha \tilde{A} + (1 - \alpha) I \quad (13)$$

where $\tilde{A}$ is the symmetrically normalized adjacency matrix and $\alpha \in [0, 1]$ balances neighbor aggregation with self-retention.

The resulting graph state-space system is given by Eq. (14):

$$H_t = \mathcal{A}H_{t-1} + \mathcal{B}X_t, \quad (14)$$

where $\mathcal{B}$ maps the node input features X_t into the hidden space. This formulation explicitly couples temporal dynamics (through the recurrence across t) with structural propagation (through $\mathcal{A}$).

The continuous-time variant of this equation, which takes the form of a dynamical system on a graph, is given by Eq. (15):

$$\frac{dH}{dt}(t) = \mathcal{A}H(t) + \mathcal{B}X(t) \quad (15)$$

where Matrix $\mathcal{A}$, as a diffusion operator based on the Laplacian operator, describes the geometric structure of the underlying data manifold, while the exponential term $e^{\mathcal{A}t}$ controls the information flow, making it related to graph topological dynamics. Thus, we can clearly see the connection between the graph theory behind GNNs and the continuous-time perspective of SSMs.

3.3.2 Selective State Update on Graphs

While the linear state-space equation provides a general model for structured propagation, it is poorly suited to heterogeneous or dynamic graph settings. The Mamba mechanism addresses this limitation by introducing selective gating, which allows each node to control the extent to which its hidden state is updated based on current inputs and neighborhood context.

In graph settings, this yields the Graph-Mamba update rule shown in Eq. (16):

$$H_t = (1 - G_t) \odot H_{t-1} + G_t \odot f_{graph}(X_t, H_{t-1}) \quad (16)$$

$G_t = \sigma(W_g X_t)$ is a node-wise gating matrix, $\odot$ denotes the Hadamard product, and $f_{graph}(\cdot)$ is a graph-aware transformation function. This function can be instantiated as Eq. (17):

$$f_{graph}(X_t, H_{t-1}) = \phi(\mathcal{A}H_{t-1} + \mathcal{B}X_t) \quad (17)$$

where $\phi(\cdot)$ is a non-linear function. The gating factor G_t controls the rate at which each node moves into its next state, thereby helping the network maintain a tradeoff between storing information from the past and acquiring information from the present neighborhoods.

In other words, through this gatekeeping approach, Graph Mamba can act as a context-dependent information diffuser, where those nodes carrying redundant information keep their old state and others receiving new information get updated. This gating mechanism replaces the uniform time step in standard SSMs with node-adaptive dynamics, allowing Graph-Mamba to model non-stationary or spatially varying behavior on graphs.

3.3.3 Unified Graph-Mamba Formulation

Combining the structural diffusion operator and the selective update rule results in the unified Graph-Mamba formulation given in Eq. (18):

$$H_t = (1 - G_t) \odot H_{t-1} + G_t \odot \phi(\mathcal{A}H_{t-1} + \mathcal{B}X_t) \quad (18)$$

where both $\mathcal{A}$ and G_t can be either learnable or data-dependent. This equation represents a hybrid propagation mechanism combining the global dynamical modeling of Mamba with the local topological constraints of graphs.

To illustrate how historical states and distant neighbors are aggregated, we can unroll the recursive process. However, due to the input-dependent dynamic gating and the nested non-linear activation ϕ , a strict algebraic parallelization is highly complex. If we assume a first-order linearized approximation for the activation pathway (to conceptually demonstrate the multi-hop receptive field), the expanded dynamic process can be approximated as in Eq. (19):

$$H_t \approx \sum_{k=0}^t \left(\prod_{i=1}^k (1 - G_{t-i+1}) \right) \odot G_{t-k} \odot \phi(\mathcal{A}^k X_{t-k}) \quad (19)$$

the product term $\prod_{i=1}^k (1 - G_{t-i+1})$ demonstrates Mamba's input-dependent, time-varying selective scanning, where the decay rate dynamically adapts at each historical step rather than remaining a static constant. It should be explicitly noted that "flattening" the highly nested non-linear operations ϕ into this summation is a conceptual approximation. In real deployments, the recurrent nested non-linearities are preserved to maintain the model's expressivity.

From an operator-theoretic perspective, Graph-Mamba defines a non-stationary graph filter whose kernel dynamically evolves in response to node-dependent gating. The spectral response of such a filter is given by Eq. (20):

$$\mathcal{H}(\lambda, t) = \sum_{k=0}^t G_{t-k} e^{\lambda k} \quad (20)$$

where λ denotes the eigenvalues of A . This suggests that the gating mechanism dynamically reshapes the model's spectral filtering behavior, enabling selective emphasis or suppression of frequency components across nodes and over time—a capability absent in traditional graph convolutions. Fig. 6 illustrates the core processing flow of the Graph-Mamba Block (GMB) from [30]; further details are provided in the original paper.

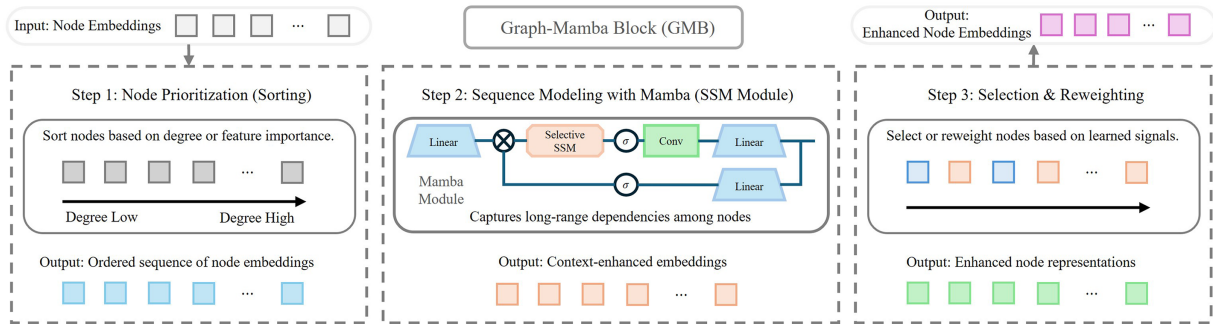


Figure 6: Core processing flow of Graph-Mamba Block (GMB).

Fig. 6 illustrates the core processing flow of the Graph-Mamba Block (GMB). In general, the block contains three major stages. First, graph-structured data are transformed into an ordered one-dimensional sequence through node prioritization or topology-aware serialization. Second, a Mamba-based state-space module performs selective scanning over the serialized sequence to capture both short-range and long-range dependencies. Third, the resulting representations are reweighted, aggregated, or projected back to the graph domain to produce updated node, patch, region, or graph-level embeddings.

Among these stages, node serialization is particularly important because Mamba was originally designed for one-dimensional sequences, whereas graph data are unordered and non-Euclidean. Therefore, before Mamba can be applied to a graph, the graph must be converted into an ordered sequence of nodes, joints, atoms, patches, superpixels, or other graph entities. This step is not merely a formatting operation. It directly affects the effective receptive field, the causal direction of information flow, and the way in which graph locality is exposed to the selective state-space mechanism.

Formally, let $\pi: V \rightarrow \{1, 2, \dots, N\}$ denote a permutation function that maps the node set V of graph G into a one-dimensional scan order. After serialization, the graph feature matrix X is converted into an ordered sequence $X_\pi = [x_\pi(1), x_\pi(2), \dots, x_\pi(N)]$. Under causal selective scanning, the hidden state at position k is mainly conditioned on previously scanned entities $\{\pi(1), \pi(2), \dots, \pi(k-1)\}$. As a result, different

permutations may induce different computational neighborhoods even when they are applied to the same graph. Nodes that are adjacent in the original topology may become distant in the scan sequence, while weakly related nodes may become adjacent after ordering. Thus, serialization acts as an implicit structural prior that determines which graph relations are treated as local and which relations must be maintained through long-range recurrent memory.

Different serialization strategies impose different inductive biases. Traversal-based ordering follows graph connectivity and tends to preserve local topological continuity. Breadth-oriented traversal emphasizes hop-wise neighborhood expansion and is close to diffusion-like message passing, whereas depth-oriented traversal emphasizes path continuity and long-chain dependencies. Centrality-based ordering prioritizes structurally important nodes, allowing hubs or influential entities to enter the state memory earlier, but it may distort local neighborhoods. Spectral ordering constructs a scan path from the eigenstructure of graph operators and can preserve smooth global organization, although it is more sensitive to graph perturbations and computational cost. Geometry-aware ordering is useful for spatial graphs, visual patches, point clouds, or medical regions, because it attempts to preserve physical proximity when mapping two- or three-dimensional structures into a one-dimensional sequence. Adaptive or learned ordering further treats the scan path as data-dependent, potentially improving flexibility but increasing optimization complexity and reducing transparency.

This perspective highlights an important distinction between Graph-Mamba and conventional GNNs. In standard message-passing GNNs, the receptive field is mainly determined by the adjacency matrix and the number of layers. In Graph-Mamba, the effective receptive field is jointly determined by the graph operator $\mathcal{A}_G$, the serialization function π , and the selective gate G_t . The graph operator defines which topological relations are available, the serialization function determines how these relations are exposed to the one-dimensional scan, and the gate controls how much information is retained or updated during recurrent propagation.

Therefore, node serialization in Graph-Mamba plays a role comparable to adjacency construction in GNNs or positional encoding in Transformers. A poorly aligned ordering may introduce artificial causal dependencies or weaken graph locality, whereas an appropriate ordering can align the Mamba scan with meaningful graph neighborhoods. Since different application domains impose different structural constraints, no universally optimal serialization strategy exists. Traffic networks, skeleton graphs, molecular graphs, medical images, visual patch graphs, and heterogeneous relational graphs may each require different forms of topology-aware ordering. For this reason, this section focuses on the theoretical role of serialization, while concrete mechanisms and model-specific implementations are discussed later in the corresponding application sections.

3.3.4 Interpretation and Theoretical Implications

The unified formulation of Graph-Mamba provides numerous theoretical insights. First, it proposes that Mamba's selective scanning may be viewed as a generalised graph diffusion process, wherein gating induces node-specific temporal heterogeneity. Second, the model gives a low-rank approximation of multi-hop propagation without saying so directly, since each step combines linear diffusion with selective updates. This makes it possible to model long-range dependencies well without having to calculate high-order adjacency directly. Finally, the fact that the state updates happen over and over again gives the model implicit memory, which lets it remember topological information from the past, which feed-forward GNNs can't do.

From a dynamical-systems perspective, the stability of Graph-Mamba can be understood by examining whether repeated state propagation remains bounded and gradually approaches a steady representation.

Under simplified settings with fixed propagation parameters, classical stability analysis suggests that such convergence may occur when the state-transition process is sufficiently contractive. This provides useful intuition for interpreting how selective diffusion can avoid unbounded accumulation over long propagation ranges. However, practical Graph-Mamba models typically involve input-dependent, time-varying, and dynamically gated transitions. Therefore, this interpretation should not be regarded as a rigorous convergence guarantee for the full model. Establishing formal stability under dynamic gating and arbitrary graph structures remains an open theoretical challenge.

3.4 Theoretical Implications and Model Interpretability

The unified theory presented in the above section provides a useful perspective for connecting graph propagation with state-space dynamics and for understanding how Graph-Mamba captures dependencies, stability, and structure-driven reasoning. Under simplified linear or locally stable conditions, this perspective allows us to analyze expressiveness, convergence behavior, and information flow by relating selective state updates to dynamic diffusion on graphs. At the same time, some interpretations in the following discussion should be understood as conceptual or operator-theoretic explanations rather than fully established general theorems. In particular, when nonlinear transformations, arbitrary graph topologies, and input-dependent dynamic gates are considered together, rigorous theoretical guarantees remain an open problem. Therefore, this section aims to clarify the analytical intuition behind Graph-Mamba while avoiding the assumption that all observed behaviors have already been formally proven.

3.4.1 Stability and Convergence Analysis

One of the essential characteristics of a dynamical system is stability, which relates to whether the hidden state converges or diverges. As described by Eq. (16), Graph-Mamba uses a selection-based, topology-specific recurrence for evolving the representations of nodes.

In order to analyze stability, we can linearize the system at some equilibrium point H^* such that $H_t \approx H_{t-1}$. Linearizing the function $\phi(\cdot)$ through the Jacobian J_ϕ and assuming G_t to be constant around the equilibrium point, we obtain the linear dynamics of the system defined by Eq. (21).

$$\Delta H_t \approx [(1 - G_t)I + G_t \mathcal{A} J_\phi] \Delta H_{t-1} \quad (21)$$

The eigenvalues of the linearized transition operator provide a local stability criterion; under this simplified setting, the system is locally stable when the spectral radius $\rho((1 - G_t)I + G_t \mathcal{A} J_\phi) < 1$. Since $\mathcal{A}$ is derived from the normalized adjacency matrix, its eigenvalues are typically located in the interval $[-1, 1]$. Therefore, if the effects of gating and nonlinear transformations remain contractive around the equilibrium point, Graph-Mamba may support bounded or locally stable state evolution under the linearized approximation.

This stability condition is analogous to that of continuous-time dynamical systems, where the Lyapunov function $V(H) = \|H - H^*\|^2$ satisfies Eq. (22):

$$\frac{dV}{dt} \leq -\lambda_{\min}(I - \mathcal{A}) \|H - H^*\|^2 \quad (22)$$

In this case, exponential convergence to equilibrium can be expected only under contractive conditions, for example when $\mathcal{A}$ is contractive. The function G_t acts as a gate that modulates the update magnitude of each node. In this sense, Graph-Mamba can inherit useful stability intuitions from SSMs while improving

robustness in some complex graph settings, although such behavior depends on the specific gating dynamics and graph topology.

It is crucial to acknowledge that the aforementioned local stability analysis relies on linearizing the system around an equilibrium point and assuming a quasi-static gating matrix ($G_t \approx G^*$). While this Linear Time-Invariant (LTI) simplification provides an intuitive baseline for understanding the bounds of Graph-Mamba, it inherently bypasses the core defining characteristic of Mamba: its highly dynamic, input-conditioned selectivity ($G_t = \sigma(W_g X_t)$). Providing a rigorous Lyapunov stability proof for such a non-autonomous, data-dependent dynamical system on arbitrary graph topologies remains an open challenge in the field. Thus, although existing empirical results indicate stable behavior of Graph-Mamba in several domains, strict theoretical guarantees under fully dynamic gating represent a critical direction for future theoretical research.

3.4.2 Expressiveness and Spectral Characteristics

In spectral terms, Graph-Mamba can be considered a temporal-topological graph filter. In GNNs, propagation may be seen as filtering of the node signals using a polynomial defined by the graph Laplacian $L = D - A$, Eq. (23) represents its modeling process:

$$H = \sum_{k=0}^K \theta_k \tilde{A}^k X \quad (23)$$

where $\tilde{A}$ represents the normalized adjacency matrix, and θ_k are the learnable coefficients. However, this static filtering mechanism limits the receptive field to within K hops, restricting its adaptability to different graph frequencies.

In contrast, the Graph-Mamba kernel evolves dynamically as a function of both time and the gating variable, as described in Eq. (24):

$$K_t(\lambda) = \sum_{k=0}^t g_{t-k} e^{-\lambda k} \quad (24)$$

here, λ strictly represents the eigenvalues derived from the spatial normalized graph Laplacian matrix, rather than the temporal transition eigenvalues found in standard 1D continuous SSMs. To avoid any conceptual conflation between temporal scaling and spatial topology, we clarify that $e^{-\lambda k}$ functions as a spatial spectral filter scaled across a k -hop neighborhood. This spatial propagation is then dynamically modulated and weighted by the time-varying temporal gain g_{t-k} generated by Mamba's selective mechanism. This formulation provides a useful theoretical mapping for interpreting how state-space mechanisms may perform input-dependent spectral filtering on non-Euclidean structures.

3.4.3 Information Flow and Interpretability

One useful way to interpret Graph-Mamba is as a structured information flow system. Many traditional message-passing GNNs aggregate information from neighboring nodes using layer-wise spatial aggregation, often without an explicit recurrent memory mechanism for controlling temporal retention. Contrast, the selective gating mechanism in Graph-Mamba enables each node to modulate its information intake rate, thereby establishing an implicit temporal hierarchy of influence among nodes, as reflected in Eq. (25):

$$\Delta H_t = G_t \odot (\phi(\mathcal{A}H_{t-1} + \mathcal{B}X_t) - H_{t-1}) \quad (25)$$

The value of G_t can be interpreted as an indicator of update intensity or information retention at time t for the corresponding node. The nodes with large values of gates have undergone many state changes, which means that they are responding to the messages that they have received. Conversely, nodes with small gate values act as a storage device, ensuring that they hold on to old knowledge within the network structure. Therefore, visualizing gate activations may provide a diagnostic view of information diffusion and update patterns, loosely analogous to attention visualization in Transformer models, but not equivalent to a causal explanation of influence.

Also, due to the structural nature of the propagation operator $\mathcal{A}$ used in Graph-Mamba's learning algorithm, the learned dynamics always have roots within the topology of the network structure. Furthermore, in the spectral domain of this network, the eigenvalues represent stable diffusion patterns on the graph. Consequently, nodes associated with similar spectral components may exhibit related dynamic patterns, which can support spectral or structural analysis of graph substructures.

3.4.4 Theoretical Bridge to Existing Paradigms

The Graph-Mamba framework can also be viewed as a bridge among several existing paradigms in machine learning. Within the framework of differential equations, Graph-Mamba can be interpreted, under suitable assumptions, as a discretized neural differential equation on graphs, where the hidden state trajectory evolves under a topology-aware vector field, as shown in Eq. (26):

$$\frac{dH(t)}{dt} = \mathcal{F}(H(t), X(t), \mathcal{A}) \quad (26)$$

Define $\mathcal{F}$ as the selective dynamic function using Mamba gates, which explicitly ties the model with Neural ODEs [58] and graph differential equations [59], while maintaining the efficiency of discrete recurrence.

From the point of view of signal processing, Graph-Mamba is the development of graph convolution filtering into a causal process. Compared with ordinary graph convolutions, Graph-Mamba can be interpreted as forming an autoregressive graph filter through successive state propagations. Such an interpretation suggests a way to incorporate sequential dependency or temporal ordering into relational data analysis, separating it from spatial GNN approaches and static state space models.

Graph-Mamba occupies an intermediate position between attention-based architectures and recurrent state-space models. In addition to the ability to focus on certain elements of the graph provided by selective scanning, it may capture long-range nonlinear dependencies in a manner partly analogous to attention, while retaining the linear-time recurrence of state-space models. Moreover, its gating unit can be understood as a learnable time-topology modulation mechanism that links attention-like selectivity with recurrent state transitions.

4 Graph-Mamba across Domains: Data Structures, Graph Constructions, and Model Interactions

The theory behind Graph-Mamba presented above shows that Graph-Mamba is a holistic model framework that combines time-state dynamics with the dynamics of graph structures. The main advantage of this model framework consists of its versatility regarding a wide range of domains and types of data structures. Real-life graphs are not only purely mathematical constructs but represent real-world entities and geometries in particular problem domains. Therefore, the use of Graph-Mamba in a given application domain depends on the peculiarities of data representation and relations between entities.

The graphs that can be found in science and engineering may vary significantly from one another. Molecular science deals with atomic nodes and chemical bonds represented as edges. In social networks, agents are nodes whereas edges stand for connections between these agents. Physics and biology may involve spatial or functional relations among the studied entities. This is where Mamba comes in handy because the presented framework offers a way to represent dynamics in time, interdependence, or long-range relations in the described frameworks. Together, the two form an extensive model class.

This chapter is dedicated to exploring Graph-Mamba models from a multidisciplinary perspective, focusing particularly on how the data representation affects the formulation and performance of the proposed model type. Each section of the chapter will cover one application domain and discuss the following subjects:

1. Graph Construction and Representation: how data entities are mapped to nodes, edges, and features, including strategies for adjacency generation and topology encoding.
2. Model-Data Interaction: how Mamba's selective state propagation interacts with structural or temporal dependencies in the graph, and what modeling behaviors emerge from this interaction.
3. Comparative Insights: how Graph-Mamba differs from conventional GNNs or Mamba-based models in the same domain, and what its advantages and limitations are.

The theoretical model of Graph-Mamba is based on the unified propagation [Eq. \(18\)](#), whereas its practical implementation is strongly domain-specific. In biological data, the graph operator $\mathcal{A}$ may represent protein-protein, molecular, or biochemical interactions, while the selective state G_t reflects context-dependent biological transitions. In spatio-temporal systems, $\mathcal{A}$ usually encodes geographic, physical, or functional connectivity, and the gating mechanism captures dynamic dependencies such as traffic flow, sensor variation, or environmental change. In geometric graphs and 3D point clouds, $\mathcal{A}$ is often constructed from spatial neighborhoods or mesh connectivity, while the state dynamics describe structural deformation or shape evolution. These examples show that Graph-Mamba provides a general propagation framework, but its actual design depends on domain-specific graph semantics and data structures.

Accordingly, this section organizes Graph-Mamba studies primarily by representative application domains, because the field is still emerging and current model development is largely driven by practical scenarios rather than by a fully stabilized architectural taxonomy. [Section 4.1](#) discusses spatio-temporal and physical systems, including traffic, sensor, communication, and motion graphs, where selective propagation supports long-range and stable forecasting. [Section 4.2](#) reviews biomedical and medical applications, including molecular interaction graphs, brain networks, and graph-driven medical image analysis, where Graph-Mamba combines biological topology with selective state modeling and interpretability. [Section 4.3](#) focuses on heterogeneous and multi-relational graphs, such as social, knowledge, and citation networks.

Although the main organization follows application domains, methodological dimensions such as graph type, serialization or scanning strategy, Mamba integration point, gating mechanism, task setting, and evaluation condition are also considered throughout the discussion. The main tables summarize domain-specific data structures and tasks. Access links for the datasets discussed throughout this section are provided in [Appendix Table A1](#), whereas [Appendix Table A2](#) presents a cross-model comparison of Mamba-specific architectural characteristics. In this sense, the survey adopts an application-oriented taxonomy complemented by method-level comparison, showing how Graph-Mamba bridges graph learning and selective state-space modeling across diverse domains.

4.1 Spatio-Temporal and Physical Graphs

Graph-Mamba models are widely applicable in a range of fields, with spatio-temporal graph systems proving increasingly popular. Such systems seek to discover state dependencies evolving with time and space,

making it possible to apply Mamba's selective state-space propagation to the realm of graph-based relational learning. In dynamic contexts, Graph-Mamba demonstrates exceptional efficacy in forecasting multivariate time series [45] and refining pose sequences [60]. Furthermore, extensions such as SourceDetMamba [61] for sequential hypergraphs and evolutionary methods for predicting remaining useful life [62] underscore the architectural flexibility of Mamba in capturing high-order temporal dependencies.

Recent literature has explicitly applied this paradigm to fields such as traffic prediction, motion recognition, and dynamic physical systems. Table 1 briefly summarizes the six studies discussed in this section. Currently, almost all of these studies use the Mamba library within the spatiotemporal graph learning framework.

Table 1: Representative Graph-Mamba studies in spatio-temporal and physical systems.

Model	Primary Task	Key Mechanism	Core Contribution
STMAGRNN [63]	Traffic prediction	Mamba encoder + spatio-temporal memory module + dynamic graph convolutional recurrent decoder	Generates dynamic graph embeddings from node-specific traffic patterns and improves traffic prediction efficiency and accuracy
STGNN-M [64]	Cellular traffic prediction	GNN + Mamba hybrid encoder	Integrates selective state-space modeling into ST-GNN
MGCN [65]	Long-term traffic forecasting	Chebyshev GCN + Bidirectional Mamba	Enhances stability and horizon length in forecasting
Transfer-Mamba [66]	Cross-city few-shot prediction	Knowledge transfer + Mamba recurrence	Learns transferable Mamba priors across spatial domains
SkeMamba [67]	Skeleton-based action recognition	Dual-path (spatial & temporal) Mamba	Models parallel state updates for body-joint graphs
ActionMamba [68]	Skeleton-based action recognition	GCN + Mamba aggregation	Hierarchical graph-sequence interaction mechanism

4.1.1 Data Structures and Graph Construction

This review identifies three main types of data in six representative studies namely, traffic and mobility systems, telecommunications networks, and skeletal-based human motion datasets. These domains differ in graph structure, node semantics, and temporal sampling characteristics.

(1) Traffic and Mobility Datasets

Traffic prediction constitutes another area in which Graph-Mamba networks with spatiotemporal components have been widely used. In this case, the nodes refer to static sensors or road sections, while the edges indicate proximity or similarity; the temporal component accounts for changes in traffic patterns. STMAGRNN is evaluated on six public traffic prediction datasets, including two traffic speed datasets, METRLA and PEMS-BAY, and four traffic flow datasets, PEMS03, PEMS04, PEMS07, and PEMS08. In these datasets, traffic networks are represented as sensor graphs, where nodes correspond to traffic sensors and edges encode dynamically generated traffic dependencies. The model uses the past one hour of observations to predict traffic conditions in the next hour. Table 2 summarizes the key statistics of these datasets.

In these datasets, the graph construction process differs from earlier static traffic graph models. Traditional traffic GNN often rely on predefined adjacency matrices derived from road distance or connectivity. Some later models use adaptive matrices to learn hidden dependencies, but these learned adjacency

structures are usually static after training. STMAGRN addresses this limitation by generating dynamic graph embeddings through the spatio-temporal memory module. Specifically, the model uses a memory bank to store representative node-level traffic patterns and retrieves the patterns most relevant to the current input. These retrieved patterns are then used to dynamically construct graph embeddings, allowing the model to better reflect time-varying traffic dependencies.

Table 2: Common traffic and mobility datasets information.

Dataset	Nodes	Timesteps	Time Interval	Time Range
METR-LA [69]	207	34,272	5 min	01 March 2012–30 June 2012
PEMS-BAY [69]	325	52,116	5 min	01 January 2017–30 June 2017
PEMS03 [70]	358	26,208	5 min	01 September 2018–30 November 2018
PEMS04 [70]	307	16,992	5 min	01 January 2018–28 February 2018
PEMS07 [70]	883	28,224	5 min	01 May 2017–31 August 2017
PEMS08 [70]	170	17,856	5 min	01 July 2016–31 August 2016

MGCN also belongs to the traffic prediction category, but its graph construction and modeling emphasis differ from STMAGRN. MGCN uses a weighted undirected road sensor graph and applies Chebyshev graph convolution to extract spatial dependencies. It then uses a bidirectional Mamba architecture to capture long-range temporal dependencies. In contrast, STMAGRN places stronger emphasis on dynamic graph generation through the STM module, using memory-based traffic pattern matching to update graph representations for the GCRN decoder. Therefore, STMAGRN is more directly aligned with the problem of dynamically changing traffic topology and time-varying road dependencies.

(2) Telecommunication Network Data

In the telecommunications industry, STGNN-M uses the Milan Cellular Network Traffic Dataset [62] when applying the Graph-Mamba framework. This dataset records the real mobile network usage of thousands of base stations in the Milan metropolitan area of Italy.

Table 3 summarizes the dataset characteristics used in STGNN-M. The original dataset comprises a 100×100 grid of nodes, and a 30×30 subregion was selected for this study.

Table 3: Milan cellular traffic dataset information.

Dataset	Nodes (Sensors)	Edges	Total Timestamps (Per Cell)	Time Interval	Spatial Resolution
Milan [71]	900 (30×30)	3240	8640	10 min	$235 \text{ m} \times 235 \text{ m}$

In this model, connectivity between base stations (nodes) is not determined by physical distance but is instead built based on the correlation of historical traffic, thus revealing cross-regional network load dependencies. Simultaneously, the STGNN-M architecture embeds Mamba as a temporal module into the GNN to simulate long-term periodic trends and short-term bursts of cellular network demand. Experiments on the Milan dataset demonstrate that this model can overcome physical topology limitations and efficiently solve the spatiotemporal prediction challenges in high-dimensional communication networks.

(3) Skeleton-Based Motion and Physical Dynamics Datasets

In tasks that involve understanding how people move, the nodes of the graph stand for anatomical joints, and the edges stand for skeletal connectivity (for example, bones that connect joints). Temporal sequences show how poses change from one video frame to the next, creating a natural spatio-temporal skeleton graph.

Both SkeMamba and ActionMamba utilize the widely used NTU RGB+D 60 and NTU RGB+D 120 datasets, captured with Kinect depth sensors at 30 frames per second, along with smaller datasets, including NW-UCLA and UAV-Human.

The NTU datasets include tens of thousands of 3D motion sequences, with each frame providing the spatial coordinates of 25 skeletal joints, and actions ranging from basic gestures to complex multi-person interactions. NW-UCLA contains three camera views and 10 subjects performing 10 action classes, while UAV-Human extends motion capture to aerial viewpoints. Table 4 shows the basic information for the commonly used Skeleton-Based Motion and Physical Dynamics Datasets.

Table 4: Skeleton-based motion and physical dynamics dataset information.

Dataset	Subjects	Actions/Classes	Samples	Views/Cameras	Modalities	Environment
NTU RGB+D 60 [72]	40	60	56,880	3 views	RGB, Depth, Skeleton, IR	Indoor (Kinect v2)
NTU RGB+D 120 [72]	106	120	114,480	3 views	RGB, Depth, Skeleton, IR	Indoor (Kinect v2)
UAV-Human [73]	119	155	67,428	Multi-view UAV cameras	RGB, Skeleton, Thermal, Depth, IMU	Outdoor (Drone-based)
NW-UCLA [74]	10	10 activities (multi-person)	~1470 videos	3 Kinect cameras	RGB, Depth, Skeleton	Indoor (Multi-angle video)

When processing the above data, although the graph construction is based on the real skeletal structure, the introduction of an “attention mechanism” allows Graph-Mamba to learn to find patterns on its own, discovering joints that are not physically adjacent but are highly interconnected in movement (such as the opposite hand and foot during running). The different processing methods for different models are as follows:

SkeMamba: Specifically focuses on spatial structure (between joints), and another approach focuses on the transitions between frames. In both approaches, Mamba acts as a “smart valve,” ensuring that the recognized movements are continuous and the transitions are natural.

ActionMamba: It allows GCN to focus on building the skeleton and capturing space, while Mamba is dedicated to managing the timeline. This division of labor ensures that the model doesn’t miss subtle local movements and can also control the macroscopic evolution of the overall posture.

Regardless of the domain, the way in which the Graph-Mamba framework makes sense of the data lies in its structure. Specifically, in traffic and communication systems, the nodes stand for the geographically or functionally spaced sensors that communicate through dynamic weights, and in human motion systems, the nodes denote physically restricted joints with continuity in their temporal states. Nevertheless, one thing remains the same across all applications—the selective gate G_t ensures that update strength is modulated by the variation in the states of the nodes themselves, dampening redundant updates for static nodes and amplifying them for dynamic areas.

4.1.2 Model-Data Interaction: Integrating Graph and Mamba Dynamics

The Graph-Mamba model framework combines graph structure with state space dynamics. For the spatiotemporal and physical systems considered here, this combination allows the models to jointly account for spatial interactions through graph convolutions or message passing with attention mechanisms and temporal interactions via selective recurrence within the Mamba framework.

(1) Traffic and Urban Mobility Systems

In the traffic prediction models STMAGRNN and MGCN both graph and Mamba modules are organized in a sequence-based hybrid framework, in which the graph encoder encodes spatial characteristics for each snapshot in time, and the Mamba module encodes the sequential relationship between the embeddings. In this way, it becomes possible to treat the temporal signal of each node as a state trajectory governed by interactions under spatiotemporal constraints. Fig. 7 illustrates the overall architecture of STMAGRNN, which is designed for traffic prediction in dynamic road sensor networks. Unlike earlier traffic forecasting models that rely on predefined or static adaptive adjacency matrices, STMAGRNN explicitly addresses the time-varying nature of traffic dependencies by combining Mamba-based sequence modeling with memory-enhanced dynamic graph construction. The model consists of four major components: a data embedding layer, a Mamba encoder, a spatio-temporal memory module, and a GCRN decoder. Together, these modules enable STMAGRNN to capture long-range temporal dependencies, node-specific traffic patterns, and dynamically evolving spatial relationships in traffic networks.

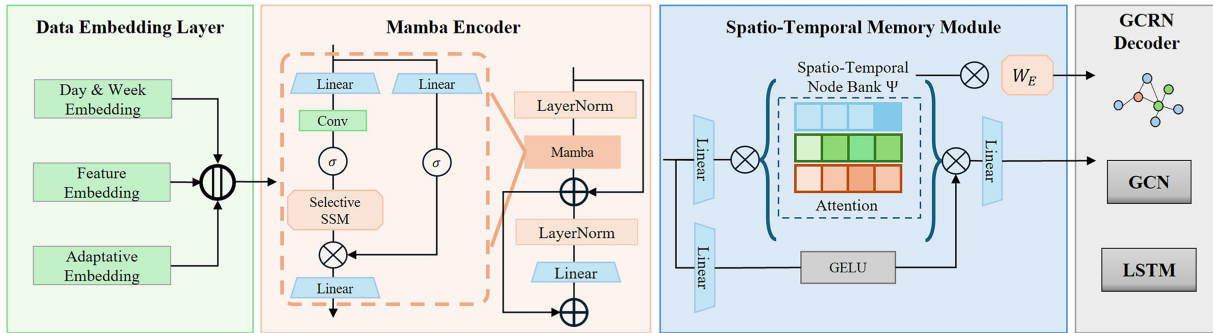


Figure 7: STMAGRNN overall framework (redrawn from [63]).

MGCN utilizes a parallel framework for integration. Spatial dependency is modeled with the help of GCN based on Chebyshev polynomials, while global temporal dependency is incorporated using Mamba models. Fig. 8 presents the structure of the MGCN in [65].

Both pathways are then combined using a multi-level feature aggregation process. The model allows for local spatial spreading and global temporal interaction to play an equal role in making accurate predictions over time, while maintaining linear time complexity in the temporal domain.

The Transfer-Mamba model enhances this method by separating the state propagation process of Mamba from any graph-dependent parameters. Here, the Mamba component is fine-tuned as a reusable temporal prior, while the graph encoders are adapted to each city. Fig. 9 shows the architecture of Transfer-Mamba, as outlined in [66].

This decomposition allows selective recurrence to be generalized across different cities, since the gating function G_t is influenced by local variations in flows but is unaware of the actual structure of the graph. Practically speaking, this results in efficient cross-domain generalization even when the available data is limited.

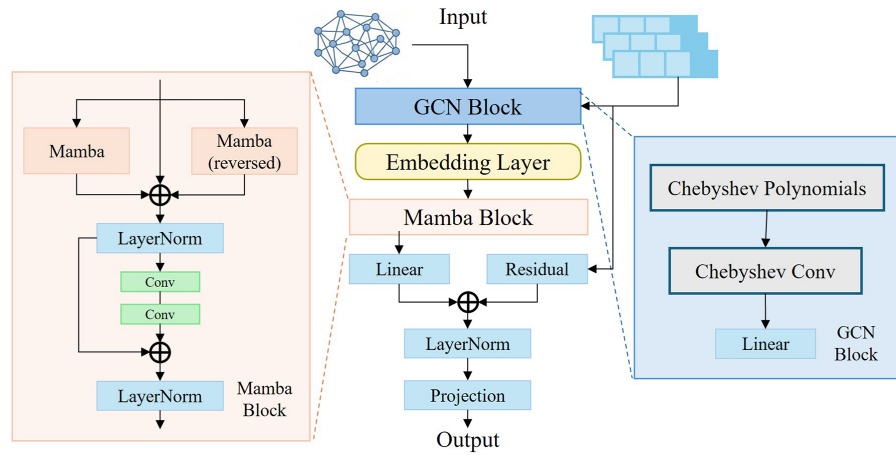


Figure 8: Overall MGCN framework (adapted from [65]).

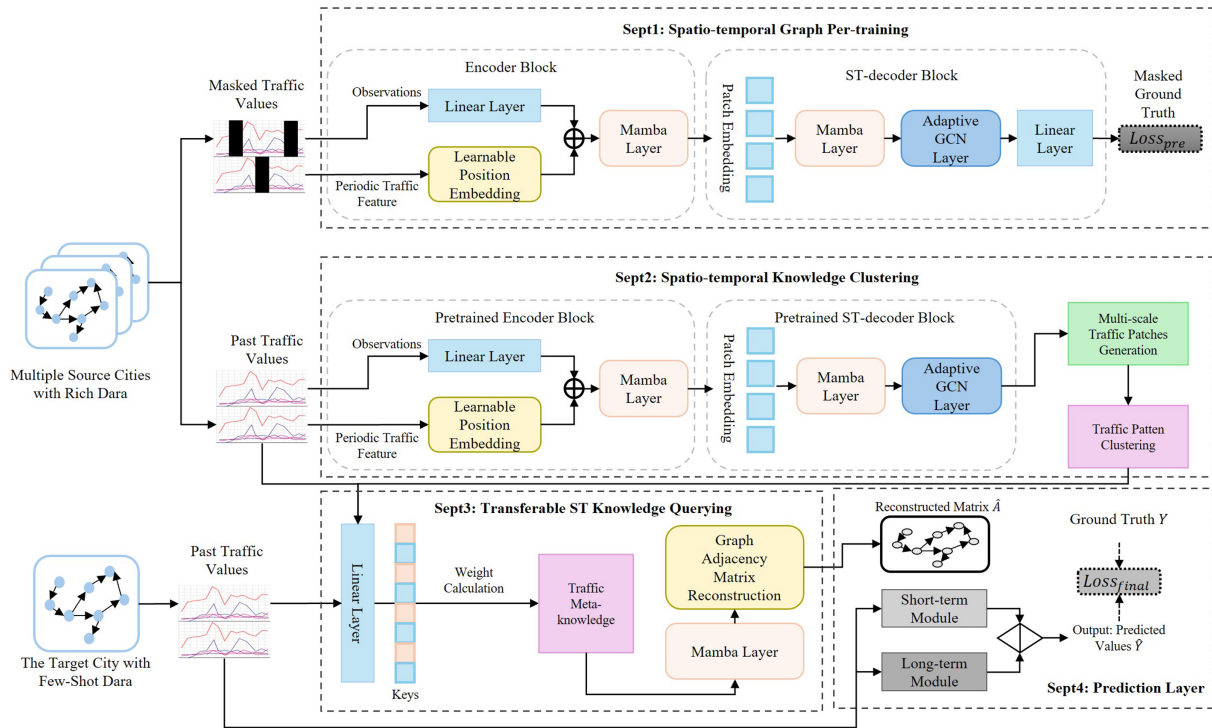


Figure 9: Transfer-Mamba overall framework (redrawn from [66]).

(2) Cellular Network Forecasting

In the STGNN-M model, the selective state-space mechanism (Mamba) is cleverly integrated into a graph neural network designed specifically for telecom traffic prediction. Unlike geographic location-dependent traffic maps, the Milan cellular network graph here is “functional”—the connections between nodes are based entirely on the correlation of signals, rather than physical distance. Fig. 9 shows the overall architecture of STGNN-M proposed in reference [64].

To jointly capture complex spatiotemporal dynamics, the STGNN-M model employs three core components: a bidirectional Mamba encoder, a dual-view graph learning module, and a Mamba-based decoder,

which can model the dynamic and heterogeneous connectivity relationships in cellular networks with extremely high efficiency. The overall framework of STGNN-M is shown in Fig. 10.

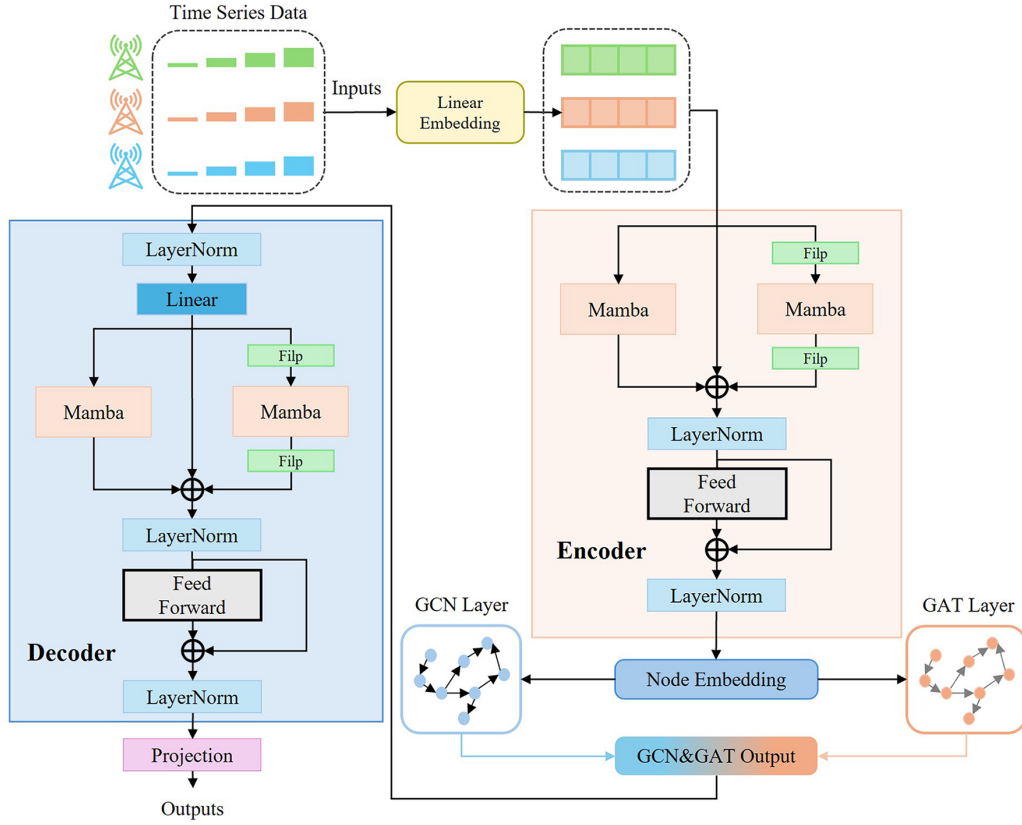


Figure 10: STGNN-M overall framework (redrawn from [64]).

(3) Skeleton-Based Motion and Physical Dynamics

In problems involving human activity recognition, Graph-Mamba architectures exhibit increased coupling between spatial and temporal aspects. The SkeMamba and ActionMamba architectures are designed to synchronize graph processing with Mamba processing, capturing the natural continuity and dependency between the components of human movement dynamics.

SkeMamba builds two independent recursive sequences—one moving along the spatial graph structure (from joint to joint), and another along the temporal sequence (from frame to frame). Each recursive sequence is controlled by an individual gating mechanism, and their combined output results in the unified description of the body movement. Fig. 11 shows the full architecture of SkeMamba, as described in reference [67].

Such a design allows the model to discriminate between structural invariance, e.g., rigid skeleton structure connectivity, and temporally varying components, e.g., body pose dynamics. Through empirical study, it is noted that activations from gates mostly occur near important joints such as the trunk and limb regions, suggesting that Mamba intrinsically focuses on semantically meaningful motion areas.

On the other hand, ActionMamba uses a hierarchy of GCN-Mamba mixed backbone architecture. Local dependency among joints is modeled by the graph convolutional layer first, and then global temporal relation among the skeleton joints is captured by Mamba layers. An Action Perception Module (APM)

unites the output from both pathways, enabling the model to simultaneously examine both local and global motions. Fig. 12 illustrates the overall architecture of ActionMamba [68].

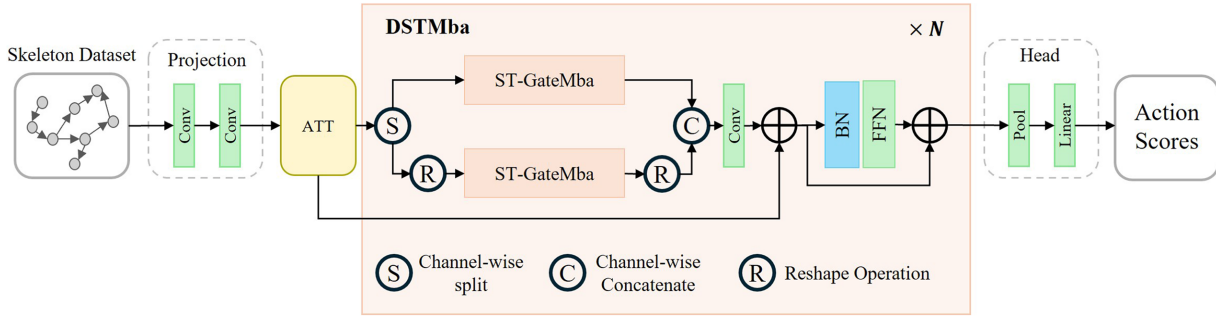


Figure 11: SkeMamba overall framework (redrawn from [67]).

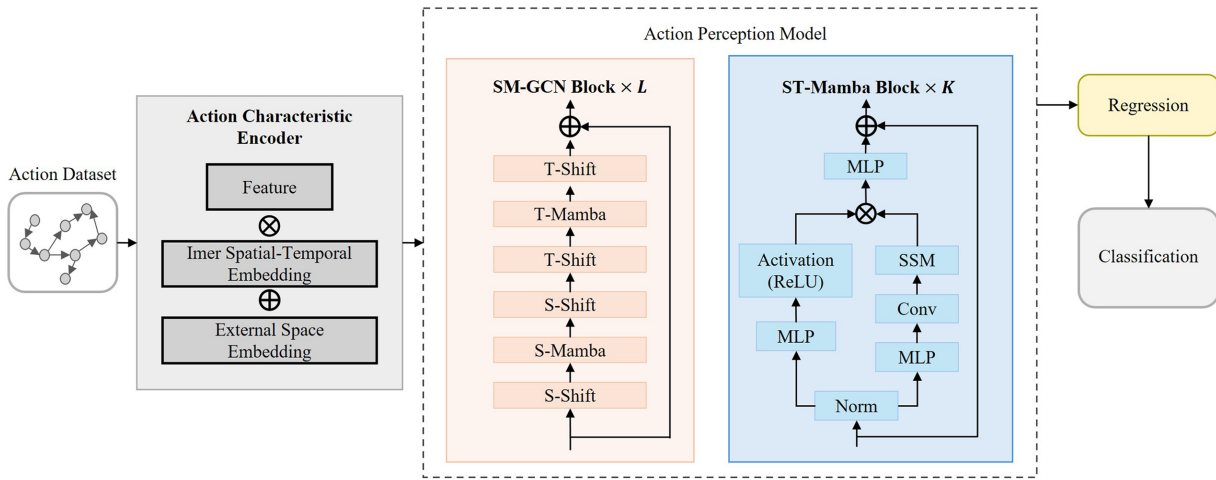


Figure 12: ActionMamba overall framework (redrawn from [68]).

This hierarchical structure facilitates multi-level spatio-temporal reasoning by integrating static pose encoding with dynamic motion aggregation.

(4) Cross-Model Observations

Regardless of different types of problems and architectures, all the Graph-Mamba models follow certain general design principles.

First of all, the entire architecture is built on the basis of selective propagation as opposed to traditional self-attention. In the course of applying the gating mechanism from the Mamba model in the temporal dimension, the Graph-Mamba models enable learning of long-term dependencies in linear time complexity while avoiding scalability issues inherent to self-attention mechanisms. Furthermore, graph structure offers spatial priors for modeling: no matter whether the adjacency matrix represents physical, functional, or anatomical connections, it imposes restrictions on the evolution of the state space.

The process of evolving a state space creates yet another interpretation opportunity since one can track how the hidden state evolves. The evolution pattern corresponds to the way information propagates and stabilizes at nodes in the physical world, as seen when applied to transportation or human coordination

systems. Finally, dual-path coupling creates better noise-resistance properties for the architecture because dual-axis propagation helps maintain temporal consistency.

In the end, these mechanisms show that Graph-Mamba is much more than just a simple mashup of GNNs and Mamba. It works as a framework based on theory that sees graph data as changing dynamic systems.

4.1.3 Theoretical Insights and Empirical Observations

Spatio-temporal and physical perspectives offer useful insights into how Graph-Mamba differs from conventional spatio-temporal graph neural networks, particularly in its treatment of long-range dependencies and state propagation. Rather than relying on explicit global pairwise interactions, as commonly used in attention-based models, Graph-Mamba captures long-term correlations through input-dependent selective state updates. The following subsections illustrate this modeling shift by reviewing representative architectures and the empirical and qualitative findings reported in recent studies.

(1) Stability and Controllable Recurrence

In the reviewed spatio-temporal Graph-Mamba models, temporal stability mainly arises from the selective state-space update inherited from Mamba. Compared with attention-based models, which compute pairwise dependencies across sequence positions, Mamba encodes long-range information through selective state transitions. This design allows the model to retain useful historical information while filtering less relevant inputs, thereby supporting stable long-horizon forecasting.

In MGCN, temporal stability is improved by using a bidirectional Mamba architecture after Chebyshev graph convolution. The Chebyshev GCN first extracts spatial dependencies from the road network, and the bidirectional Mamba then captures long-term temporal dependencies in both forward and reverse directions. This helps the model maintain prediction quality as the forecasting horizon increases.

For STMAGRNN, stability is achieved differently. The model first uses a Mamba encoder to extract long-range spatio-temporal features from embedded traffic sequences. Because Mamba has linear sequence complexity and a selective state-space mechanism, it can model temporal evolution more efficiently than attention-based traffic models. STMAGRNN then introduces a spatio-temporal memory module to retrieve representative traffic patterns from a memory bank. This memory-based pattern matching helps stabilize prediction by grounding the current input in learned historical traffic patterns rather than relying solely on instantaneous node signals. Finally, the GCRN decoder combines GCN and LSTM to jointly model spatial dependencies and temporal recurrence.

(2) Flexibility in the Dynamics of Graph State Coupling

The combination of graph structures and selective state-space modeling gives Graph-Mamba models flexible expressive power. In traffic prediction, a node's future state depends not only on its own historical observations but also on spatial interactions with other sensors or road segments. Classical traffic GNNs often use predefined adjacency matrices based on physical distance or road connectivity. Later models introduce adaptive adjacency matrices, but these learned graphs are usually static after training and may fail to reflect real-time changes in traffic dependencies.

STMAGRNN directly addresses this limitation through its spatio-temporal memory module. Instead of relying on a fixed graph or a static adaptive adjacency matrix, the STM module learns representative traffic patterns from node-related data and dynamically generates graph embeddings. These embeddings are then

used by the GCRN decoder to guide graph convolution. In this way, the graph structure becomes input-sensitive and traffic-pattern-aware, allowing the model to better capture non-stationary spatial correlations caused by congestion, rush hours, incidents, or changing road conditions.

This mechanism differs from MGCN. In MGCN, the graph structure is based on a static road sensor topology, and Chebyshev graph convolution is used to extract multi-hop spatial information before temporal modeling. This design is effective for long-term prediction but still relies on a relatively fixed spatial structure. STMAGRN, by contrast, places greater emphasis on dynamic graph generation. Its STM module retrieves the most relevant historical traffic patterns and uses them to construct dynamic graph embeddings, making the spatial dependency structure more adaptive to current traffic states.

In the STGNN-M architecture, the Mamba mechanism operates as a temporal filter, adaptively responding to fluctuations in network-level activity. This enhances short-term adaptability under irregular time intervals, indicating that the gating function G_t functions similarly to a data-driven, time-sensitive controller. These results confirm that the advantage of Graph-Mamba lies not in architectural complexity, but in the adaptive coupling between local structure (graph) and global evolution (state).

(3) Generalization and Transfer Properties

Evidence from the Transfer-Mamba approach proves that the temporal modeling through Mamba networks is capable of generating generalization. In particular, the temporal dynamics generated from an urban traffic network, such as METR-LA, could be effectively applied to other networks, for example, Chengdu and Shenzhen, with minimal tuning. Thus, the selectively updated states can learn temporal dynamics without any domain-specific information.

The conclusion aligns with the modular nature of Graph-Mamba's design, where the temporal dynamics generated via Mamba layers are independent of the spatial structure. Hence, as $\mathcal{A}_t$ changes according to the new graphs, the recurrent kernel can be reused as a common temporal dynamic generator for different graphs.

(4) Interpretability through Selective Propagation

An observation consistently found in all six experiments is the ability to interpret the propagation mechanism of Graph-Mamba, since Mamba uses explicit hidden representations H_t through time. Therefore, the changes in these representations through time can be plotted or visualized.

In traffic networks, activation values G_t increase when there are traffic jams. On the other hand, in skeleton action recognition problems, high gate activations are typically observed when joints exhibit large motion variations, such as limbs, and low activation levels are seen in relatively static joints, such as the torso. These observations, derived from SkeMamba and ActionMamba visualization results, lend support for interpreting G_t as a selective attention strategy based on temporal continuity.

Moreover, the state-space formulation of Graph-Mamba provides a structured view of temporal information propagation and motivates further analysis using concepts from signal processing and control theory. Although general stability and convergence results remain to be established, the explicit representation of historical states may facilitate model inspection and help trace how past information contributes to current representations.

(5) Empirical Trends across Models

The observed empirical results across the studied spatio-temporal tasks display some consistent findings. The incorporation of Mamba modules in all models leads to decreased MAE and RMSE in prolonged prediction horizons (30–60 min) relative to Transformer-based STGNNs. Since Mamba modules require linear-time computation and hence effectively limit model sizes, they address the typical issues of training

large temporal graphs. Moreover, they yield better generalization abilities as the Transfer-Mamba framework successfully transfers to other cities through identical state parameters, implying that the learned temporal recurrences exhibit physically transferable properties. In addition, in visual motion tasks, activations in gates correspond to important physical motions. Rather than discovering only statistical associations, selective propagation represents physical events. Finally, the Graph-Mamba framework surpasses mere modularity-based aggregation by combining graph structures with robust state space dynamics.

4.2 Biomedical and Bioinformatics Graphs

It is clear that the biomedical field represents an especially promising area of Graph-Mamba implementation because biological systems are essentially structured, multi-level, and dynamic. Indeed, there are many areas in biology and medicine where the problem lies in classifying images based on WSIs [75,76] and segmenting 3D medical images [77]. The mentioned problems can be solved using hybrid architectures such as VGM-UNet [78] and GCNet-Mamba [79], which use adaptive fusion in dealing with high resolution and complex spatial dependencies typical for digital pathology and clinical image processing. Importantly, whether at the molecular level, in brain connectivity, or even medical imaging, each dataset has one thing in common: they all can be described as graphs of interconnected entities evolving through time or space. Hence, Graph-Mamba can serve as a tool to capture the spatiotemporal characteristics of living things by combining topological representation with selective dynamical propagation.

In the biomedical field, various heterogeneous data, from microscopic molecular structures and mesoscopic dynamic brain networks (such as EEG/fMRI) to macroscopic medical images (pathology/radiology), can be uniformly represented as graph structures. However, existing modeling paradigms have inherent limitations: GNNs struggle to handle long-range dependencies and high-dimensional sequences, while sequence models (RNNs or attention mechanisms) lack awareness of explicit graph topology. Graph-Mamba effectively bridges this gap by introducing a selective SSM. It treats multi-scale networks as continuous dynamic systems, achieving context awareness and selective scanning across time and space, demonstrating outstanding performance in complex tasks such as protein-ligand binding prediction, brain network classification, and medical image segmentation. More importantly, compared to “black box” attention models, Graph-Mamba’s explicit hidden state dynamics endow the model with strong interpretability. By tracking signal propagation and gating activation in the network, it can clearly reveal the driving factors of molecular interactions, brain region synchronization, and pathological boundaries, highly meeting the rigorous requirements of transparent artificial intelligence in clinical applications.

The current section will review the use of Graph-Mamba in three fundamental biomedicine scales. First, it will specify datasets, graph models, and their node-edge semantics in each specific sub-area. Next, the coupling of Mamba’s state space algorithm with different biologically based graphs will be analyzed. The aspects of stability, generalization, and interpretation will be covered with quantitative evidence provided.

In these sections, we want to show that Graph-Mamba is not just a new way to build things, but also a way to model structured biological dynamics that connects molecular bioinformatics, systems neuroscience, and medical imaging in a single mathematical and computational framework. This section selects six papers that represent the core of current biomedical Graph-Mamba research (as of October 2025). They cover various levels of biology: the molecular scale-protein-ligand and RNA-disease networks, where graph nodes represent molecular components; the neural scale-dynamic brain connectivity networks derived from EEG/fMRI signals; and the organ/tissue scale-medical imaging tasks (segmentation, classification) represented by graph-structured pixel or patch interactions. Table 5 provides a list of the papers and basic information.

Table 5: Representative Graph-Mamba studies in biomedical and bioinformatics.

Model	Primary Task	Domain	Graph-Mamba Integration
MGDTA Model [80]	Drug-Target Affinity (DTA) Prediction	Molecular Bioinformatics	Combines Graph Transformer with Mamba to jointly encode protein sequences and molecular graphs via selective state propagation.
MambaCAttn GCN+ [81]	piRNA-Disease Interaction Prediction	Genetic Bioinformatics	Uses MambaTextCNN for RNA sequence dynamics, with GCN and attention for cross-modal association.
Brain-Mamba [82]	Brain Functional Network Modeling, Multi-modal EEG/MEG/fMRI Decoding	Computational Neuroscience	Introduces dual-channel (BTMamba + BNMamba) for temporal brain activity and graph-level connectivity.
AGIIBM model [83]	Multi-view Brain Network Classification (ASD, MDD)	Neuroinformatics	Employs Mamba as temporal encoder within a GIB compression scheme for dynamic brain graph features.
DG-SSM-MIL [84]	Whole-Slide Image Classification	Digital Pathology	Combines SSM temporal modeling with graph aggregation of patch-level representations for slide-level prediction.
GraphMamba (Medical Image Segmentation) [85]	Segmentation (Heart MRI, Liver/Organ CT)	Medical Imaging	Proposes graph-driven topology-aware scanning and spatial consistency learning within Mamba propagation.

4.2.1 Data Foundations and Biomedical Graph Structures

Graph-Mamba models have been successful in biomedical fields because they can encode complicated biological data as structured graphs that are made better by dynamic state-space propagation.

In these instances, the graph abstraction functions as a cohesive framework that integrates molecular connectivity, neural circuitry, and spatial anatomy. This part talks about the basic data structures that biomedical Graph-Mamba models are built on and how they work with the selective dynamics of the Mamba mechanism.

(1) Molecular Graphs and Bioinformatic Networks

From a molecular and genetic perspective, structures of items like proteins, ligands, and non-coding RNA have intrinsic structure properties that make graph modeling possible [86–89].

Under this model, each molecule or biomolecular object is expressed in the form of graph $G = (V, E)$, where $V = \{v_1, v_2, \dots, v_n\}$ are atoms, residues, or subsequences, and $E \subseteq V \times V$ represent chemical bonds, interactions, or functional relations.

With the use of atom connectivity matrices, the MGDTA Model [80] converts molecules into graphs based on their SMILES representations [90–92]. It further leverages a transformer encoder to generate protein embeddings, followed by using Mamba state-space propagation to connect them. The selective recurrence feature makes it possible for the model to capture long-range residue dependencies while avoiding excessive memory requirements, which is critical when dealing with big proteins and flexible ligands.

MambaCAttnGCN+ [81] uses graph representations to make connections between piRNA nodes and disease nodes extracted from biomedical databases (MNDR v4.0 [93], piRDisease v1.0 [94]). The MNDR v4.0 benchmark used in Ref. [81] was derived from RNADisease v4.0 [93], from which the authors extracted a piRNA–disease subset for model evaluation. In this approach, a bipartite graph is built, where piRNA nodes equipped with features obtained via MambaTextCNN are connected to ontology-based embeddings carried by disease nodes. Simultaneously, Mamba retains temporal-like sequence characteristics in the nucleotide sequence. From this perspective, the usage of a recurrent network by Mamba means that there exists hierarchy in contextual relations in static biological graphs, which could be considered as a time metaphor of biochemical sequence evolution.

Five key datasets commonly used in DTA and ncRNA-Disease Association studies are presented in Table 6 below. MNDR 4.0 and piRDisease v1.0 dataset details relate to the structures described in [67] rather than the actual dataset sizes.

Table 6: Summary of public datasets for DTA and ncRNA-disease association tasks.

Dataset	Drugs/Molecules	Graphs	Avg. Nodes	Avg. Edges
Davis [95]	68 drugs	25,380	11,776	352,479
KIBA [96]	2116 drugs	88,992	272	9841
BindingDB [97]	>1,000,000 compounds	700	400	6194
MNDR 4.0 [93]	8205 piRNAs	642	31	252
piRDisease v1.0 [94]	4350 piRNAs	204	62	2419

The models in the table above show that although molecular and genome maps have static topologies, they can better represent dynamic intra-sequence coupling and intermolecular communication thanks to the selective propagation of Mamba.

(2) Brain Connectivity Graphs

At the neuroscience level, Graph-Mamba is widely used to analyze the dynamic interaction networks of the brain. In a standard brain network graph, nodes correspond to unique brain regions, which in most cases would be the 116 AALs of interest in fMRI experiments. Node attributes are derived from the time-series associated with EEG or fMRI measurements, whereas edges represent functional associations among such signals, and these associations change over time. Normally, the degree of edge strength is computed using the Pearson correlation. For example, Brain-Mamba [82] cleverly designs two complementary modules: BTMamba is responsible for capturing the local “temporal dynamics” of each brain region, while BNMamba treats the global network connections as a kind of “graph state transition”. The intuition of this design is that the cognitive function of the brain depends on the dynamic connections between brain regions; at the same time, Mamba’s unique “selectivity” mechanism gives the model a strong noise resistance and can easily cope with the test of different time resolutions, which is extremely important for decoding neural signals.

In terms of processing multimodal data, AGIIBM [83] treats different types of data as different “views”. It cleverly eliminates redundant connections by utilizing the information bottleneck principle, retaining only the core dynamic evolution.

Data sources discussed above can be found in Tables 7 and 8 below. More specifically, Table 7 summarizes the graph structures associated with data source [82] and Table 8 presents an overview of data source [83].

Table 7: Summary of topological graph properties and modalities in neuroimaging datasets.

Dataset	Modality	Graphs	Avg. Nodes	Avg. Edges	Description/Task
BVFC [98]	fMRI	25,380	11,776	352,479	Task-based fMRI from 3 subjects viewing 8460 images (9-class visual categorization).
BVFC-MEG [98]	MEG	88,992	272	9841	MEG counterpart of BVFC, same 9-class visual task with 4 subjects.
ADHD-200 [99]	fMRI	700	400	6194	250 ADHD + 450 controls; resting-state fMRI for disorder classification.
TUH-EEG [100]	EEG	642	31	252	EEG recordings (31 channels, 642 subjects) for seizure vs. normal classification.
MPI-EEG [101]	EEG	204	62	2419	EEG data from young and elderly subjects; cognitive aging analysis.
HCP (Mental States) [102]	fMRI	7440	1000	7635	Predicting one of 7 mental states (Emotion, Language, Motor, etc.).
HCP (Age Prediction) [102]	fMRI	1067	1000	8041	Age group prediction (5 classes) based on fMRI connectivity.

Note: The edge statistics are reproduced from Table 5 of BrainMamba [82]. The original study does not explicitly specify whether symmetric functional connections are counted once as unique undirected edges or twice in a bidirectional computational representation. Therefore, this value should be interpreted as the edge count reported by the source rather than as the number of unique undirected node pairs.

Table 8: Summary of electrophysiological and clinical brain network datasets for functional classification.

Dataset	Modality	Subjects/Samples (As Stated in the Paper)	Task Type
CHB-MIT [103]	EEG	23 patients; 16 brain channels	Seizure detection (binary classification)
Siena EEG Dataset [104]	EEG	27 common monopolar electrodes and 16 bipolar electrodes	EEG-based epilepsy classification
ABIDE-I [105]	fMRI	1096 brain network samples (527 ASD + 569 controls)	Brain network classification (ASD vs. control)

For the EEG-based datasets in Table 8, CHB-MIT and Siena, preprocessing followed standardized clinical EEG protocols to ensure cross-subject consistency. In the CHB-MIT dataset, EEG signals from 23 patients were recorded at 256 Hz using the international 10–20 electrode system. Sixteen channels consistently present across all subjects were retained for analysis. The data were segmented into 1-s windows, maintaining a 1:1 ratio of seizure to non-seizure samples to balance classification. Similarly, the Siena epilepsy dataset employed 27 common monopolar and 16 bipolar electrodes. The one-second sliding window

methodology was retained, and balanced sets were formed by matching seizures and non-seizures for every patient.

For the ABIDE-I dataset, preprocessed resting-state fMRI data were obtained from the Preprocessed Connectomes Project (PCP) [106]. The Automated Anatomical Labeling (AAL) atlas [107] was then used to parcellate the brain into 90 regions of interest. Pairwise Pearson correlation coefficients between regional BOLD time series were calculated to construct weighted functional connectivity matrices. To reduce spurious correlations associated with zero BOLD signal vectors, problematic samples were removed during preprocessing. The final dataset contained 1096 functional brain networks, including 527 individuals with autism spectrum disorder and 569 typically developing controls, and was used for cross-patient ASD classification.

(3) Medical Imaging Graphs

Graph-Mamba goes further than just time-dependent modeling and introduces a spatially organized propagation based on the anatomy and imaging. As noted earlier, in image analysis in medicine, one often represents spatial structures such as tissues, organs, or lesions through graph-based regions to show the relationship between them rather than pixels, unlike traditional CNN models.

For instance, slides in pathology are divided into smaller sections called patches and used as the vertices of DG-SSM-MIL [84]. The edges between these patches will either depend on proximity or the similarity of color and textures, whereas the vertices will contain information about histology from the corresponding patches. Within Mamba, the state space model takes care of the time aspect of these relationships and models scanning order for pathology images. Additionally, the Graph Neural Network will gather local information for global prediction tasks. This architecture, where the state space propagation in Mamba is combined with graph-based modeling of spatial relationships, allows the efficient representation of gigapixel images while retaining context consistency.

GraphMamba: Medical Image Segmentation [85] proposes an extension of Mamba's selective propagation to two-dimensional and volumetric cases. GTAS is used by GraphMamba instead of scanning pixels in the order of a simple sequence. In consequence, the order in which Mamba scans spatial entities depends on the graph connectivity obtained either via superpixel segmentations or using anatomical priors. SCL further improves propagation by imposing the anatomical consistency constraints on the boundaries of the generated masks. In experiments on LVCM, LiTS, and BTCV benchmarks, it was proven that this approach is a combination of speed provided by Mamba and accuracy due to the use of graph-based modeling in the spatial domain. Table 9 contains the key information about the above-mentioned datasets.

Table 9: Public and clinical datasets used in DG-SSM-MIL and GraphMamba.

Dataset	Modality/Domain	Samples/Slides/ Volumes	Resolution or Size	Annotation Type
BRACS	Histopathology (Breast Cancer)	547 WSIs (training + validation)	512 × 512 patches extracted at 20× magnification	Slide-level labels (benign, malignant, normal)
TCGA-NSCLC	Whole-Slide Pathology (Lung)	965 WSIs	Up to gigapixel resolution	Slide-level diagnosis from TCGA clinical metadata

(Continued)

Table 9 (continued)

Dataset	Modality/Domain	Samples/Slides/ Volumes	Resolution or Size	Annotation Type
TCGA-RCC	Whole-Slide Pathology (Kidney)	891 WSIs	~1–2 GB per slide	Slide-level diagnosis
CAMELYON16	Histopathology (Breast Lymph Node)	400 WSIs (270 train, 130 test)	50,000 × 50,000 pixels	Pixel-level tumor/normal masks
LVCM	Cine-MRI (Cardiac)	10,400 MRI frames from 105 subjects	256 × 256 × 20 sequences	Pixel-wise myocardium and lesion masks
LiTS	CT (Liver Tumor Segmentation)	131 train + 70 test volumes (≈58,600 slices)	512 × 512 axial resolution	Pixel-level liver and tumor masks
BTCV	CT (Abdominal Multi-Organ)	50 volumes, 80–225 axial slices per volume	512 × 512, 1–6 mm slice thickness	13 organ-level segmentation masks

Both DG-SSM-MIL and GraphMamba formulate medical image analysis tasks as graph-structured learning problems, where image regions are represented as nodes and their spatial and semantic relationships are encoded as edges. In DG-SSM-MIL, each whole-slide image (WSI) is divided into non-overlapping 512×512 pixel patches extracted at $20\times$ magnification. Each patch serves as a node, and edges are defined using k-nearest neighbor connectivity ($k = 8$) based on spatial distance or color-texture similarity between adjacent patches. Node features include 1024-dimensional vectors obtained using pretrained CNN backbone models (ResNet-50 or UNI). The GNN takes care of spatial locality propagation. On the other hand, Mamba state-space model incorporates temporal dependencies between patches to account for scanning dependency. This combination offers a highly efficient gigapixel-level representation without losing any important information.

The Graph-Mamba method constructs graphs based on superpixels or patches. The first step is dividing the medical image (CT or MRI) into smaller segments known as patches. Graphs are created through connectivity based on Euclidean distance and feature similarity. Graph pruning is then performed to get topology-aware graphs. Graph-Mamba utilizes such topology-aware graphs to guide the Graph-Driven Topology-Aware Scanning (GTAS). In GTAS, BFS algorithm decides on the sequence of state-space propagation. The Spatial Consistency Learning (SCL) helps in increasing embedding consistency for better results across various scanning orders.

(4) Cross-Scale Observations

In the molecular, neural, and imaging fields, Graph-Mamba exhibits a highly consistent core logic: it uses adjacency matrices as structural operators, cleverly integrating topological relationships into continuous dynamic equations to guide state changes; simultaneously, its gating mechanism acts like an intelligent regulating valve, flexibly controlling signal propagation intensity based on local environments such as molecular chemical stability, brain region synchronicity, or the spatial smoothness of images. This design brings two very practical advantages. First, extremely high computational efficiency. Since graph nodes

in biomedical fields often number in the tens of thousands, Mamba's linear computational complexity can easily handle this massive scale, directly overcoming the quadratic growth of computational power bottleneck of traditional attention models. Second, process transparency. The selective gating and state-update mechanisms of Graph-Mamba offer a structured way to inspect how information is retained and propagated across biological or anatomical graphs. Although these internal states may sometimes align with meaningful spatial or functional patterns, such alignment does not by itself constitute a causal explanation or guarantee clinical interpretability. Further validation is needed to establish when these representations are biologically meaningful and clinically reliable.

4.2.2 Model Interactions and Cross-Domain Adaptations

In this section, we explore how models work together and how their interactions facilitate adaptation in different biomedical applications.

(1) Sequential-Structural Fusion in Molecular Bioinformatics

In molecular modeling tasks such as protein-ligand binding affinity and piRNA-disease interactions, the Graph-Mamba architecture relies on two-stream fusion techniques. The graph stream encodes the topological relationships between molecular subunits or biomedical entities, while the Mamba stream processes the corresponding sequential or contextual features.

In this context, as seen from Fig. 13, the MGDTA framework [80] utilizes a multimodal late-fusion method that uses graphical representations of chemical molecules together with protein sequences for predictions. While using both modalities separately, the framework aims to fuse their features to better predict the bindings of drugs and targets through leveraging their dependencies structurally and sequentially.

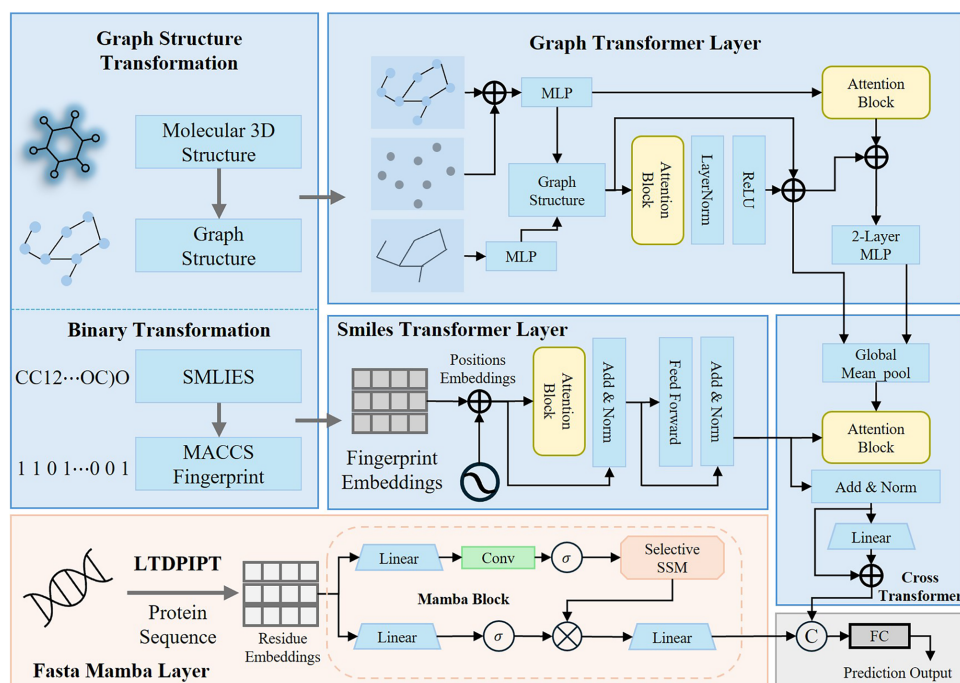


Figure 13: Schematic overview of the MGDTA model for protein-ligand affinity prediction (redrawn from [80]).

Concerning the molecules, the process transforms the three-dimensional chemical structure of a compound into a graph in which nodes correspond to atoms and edges represent covalent bonds. After being

annotated with physicochemical descriptors, the application of Laplacian eigenmaps enables the creation of position embeddings such that the geometrical structure is maintained. After that, the Graph Transformer Layer (GTT) model is used, where the message passing neural networks (MPNNs) are combined with multihead self-attention models. In doing so, the GTT model manages to assign importance to the local chemical bonds and structural relationships without suffering from the atom/structure dilemma.

Meanwhile, the structure of the molecule is converted from the SMILES representation into a 166-bit MACCS fingerprint for capturing the structural keys that are pre-defined. The sequential data, upon undergoing embedding projection and the Transformer encoder operation, are used to supplement the GTT outputs with important information concerning the functional group and the overall structure of the molecule. On the other hand, the protein modality leverages the Mamba architecture. The Mamba model can adjust its internal state transitions using the scanning mechanism to find the precise binding site residues without affecting linear computation.

This architecture integrates all three types of generated data in the cross-domain feature fusion using dot-product cross-attention. This way, dynamic re-weighting of the multi-channel input feature is performed, which leads to a synergy between the representation of the drug structure and the embedding of the protein sequence. Finally, these features will be passed through the fully connected layers for predicting the binding affinity score.

Similarly, the MambaCAttnGCN+ algorithm applies a tri-branch fusion strategy, which integrates: (1) MambaTextCNN for capturing information from the nucleotide sequences, (2) cross-attention for semantics alignment, and (3) graph convolutional network for inferring topology within the piRNA-disease association network. The MambaCAttnGCN+ model architecture is shown in Fig. 14 from reference [81].

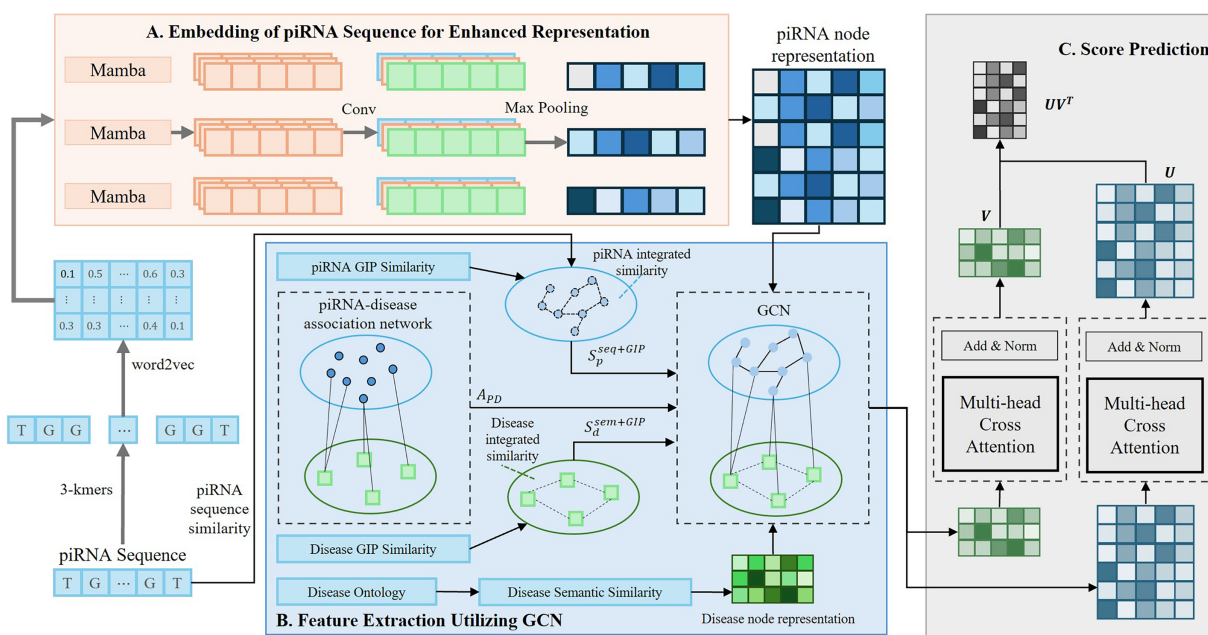


Figure 14: Schematic overview of the MambaCAttnGCN+ model for piRNA-disease association prediction (redrawn from [81]).

The Mamba block is not used to model time per se, but to propagate context along the nucleotide sequence, creating a continuous latent representation that complements the discrete graph structure. Through this mechanism, molecular graphs and biological sequences operate as parallel state spaces

(one defined by structural topology and the other by informational sequence order) which interact via cross-modal attention.

The above architecture demonstrates that Mamba can serve as a dynamic context generator in biochemical graph learning, introducing selective recursion to complement graph convolution or attention-based relational encoding.

(2) Temporal-Topological Integration in Neural Graphs

At the neural level, the Brain-Mamba and Multi-view Brain Network Mamba models showcase Graph-Mamba's ability to combine temporally relevant neurodynamics with structural or functional connectivity graphs. Both models treat brain activity as a collection of temporal series of regions connected by evolving correlations, formalized as a dynamic graph $G_t = (V, E_t)$.

In *Brain-Mamba*, the architecture is explicitly dual-path; Fig. 15 shows the *Brain-Mamba* architecture proposed in ref. [82].

1. BTMamba (Brain Temporal Mamba) models within-region temporal dependencies using a selective state-space module, effectively replacing recurrent units (e.g., GRUs) with a linear-time sequence processor.
2. BNMamba (Brain Network Mamba) aggregates regional states through graph-based propagation over functional connectivity matrices. These two paths are interleaved, allowing recurrent updates from Mamba to inform graph-level reasoning and *vice versa*. The model thus captures both intra-region dynamics and inter-region coupling.

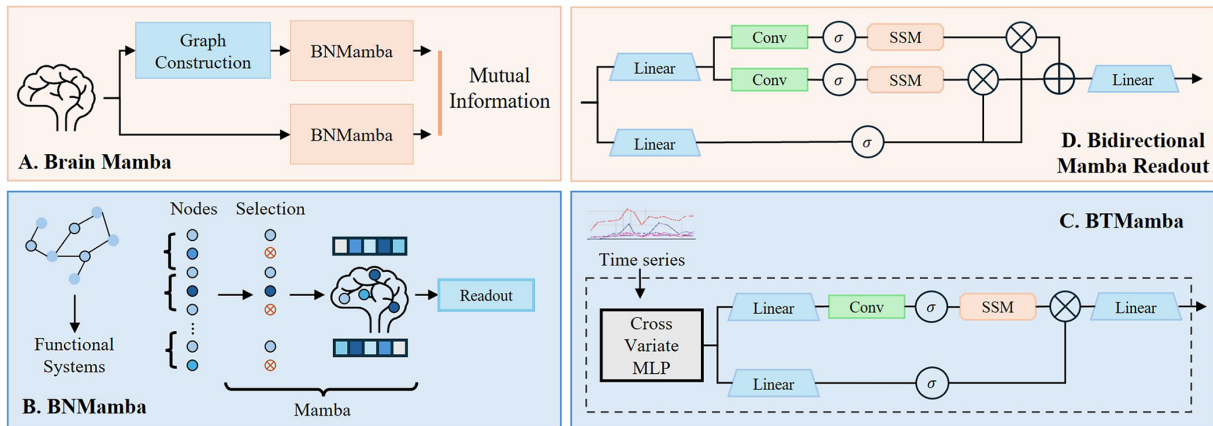


Figure 15: Schematic overview of the Brain-Mamba architecture for multimodal neuro-signal decoding (redrawn from [82]).

Multi-view Brain Network Mamba, on the other hand, improves upon the proposed architecture through the addition of the Adaptive Graph Isomorphic Information Bottleneck (AGIB) block, whose purpose is to reduce redundancy of features within each modality (fMRI, EEG). Each modality retains its respective graph structure, whereas a common Mamba Temporal Encoder is used for all of them. The architecture of AGIIBM is depicted in Fig. 16 [83].

The common encoder ensures that temporal features will be aligned across views, making it possible for the network to transfer knowledge from one modality or patient cohort to another. In contrast to fixed GNNs, the proposed design allows the model to learn temporal correspondences across modalities, which is essential for combining data from multiple sources of neuroimaging.

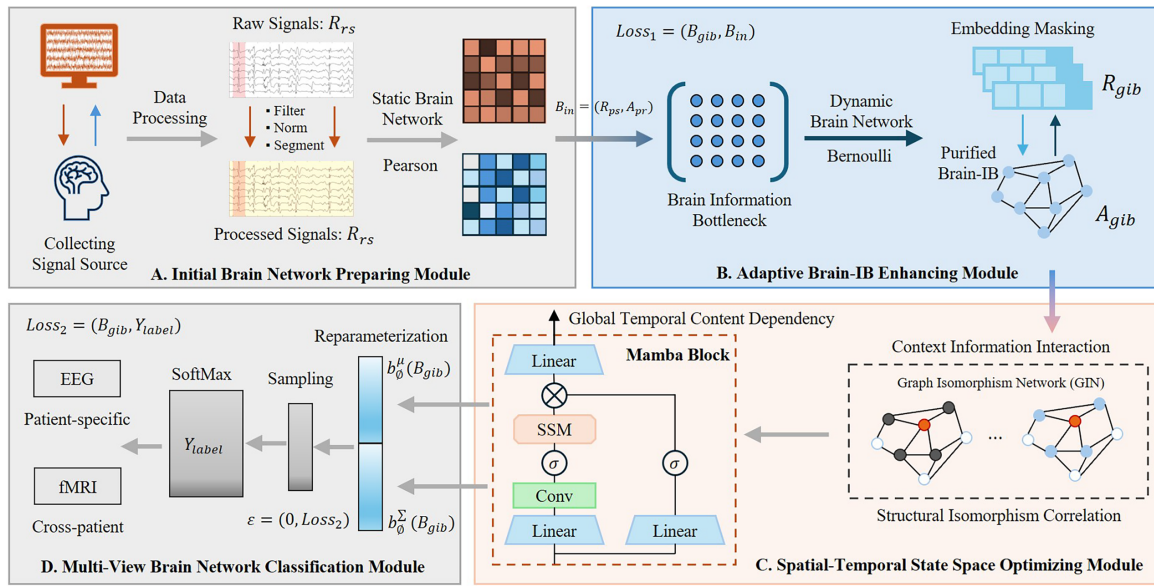


Figure 16: Schematic overview of the AGIIBM model for multi-view brain network classification (redrawn from [83]).

In both architectures, selectivity-based gating serves as a means of recursion as well as a stability control measure. They avoid exploding gradients by limiting update values while maintaining temporal continuity in noisy signals.

(3) Spatial-Sequential Coupling in Medical Imaging Graphs

However, concerning medical imaging, the relationship between Graph and Mamba is different. It goes from the time dimension of evolution towards the spatial dimension of order-awareness, since even though biomedical images are static, they have some natural spatial order dependencies that may be traversed through as if it were a sequence. The patch-level embedding features are considered a pseudo-sequence created based on scanning order or spatial proximity. In the Mamba network, features work as the memory of information within the context, while in the GNN, features are collected in the slide level. Fig. 17 describes the structure of the DG-SSM-MIL approach described in reference [84].

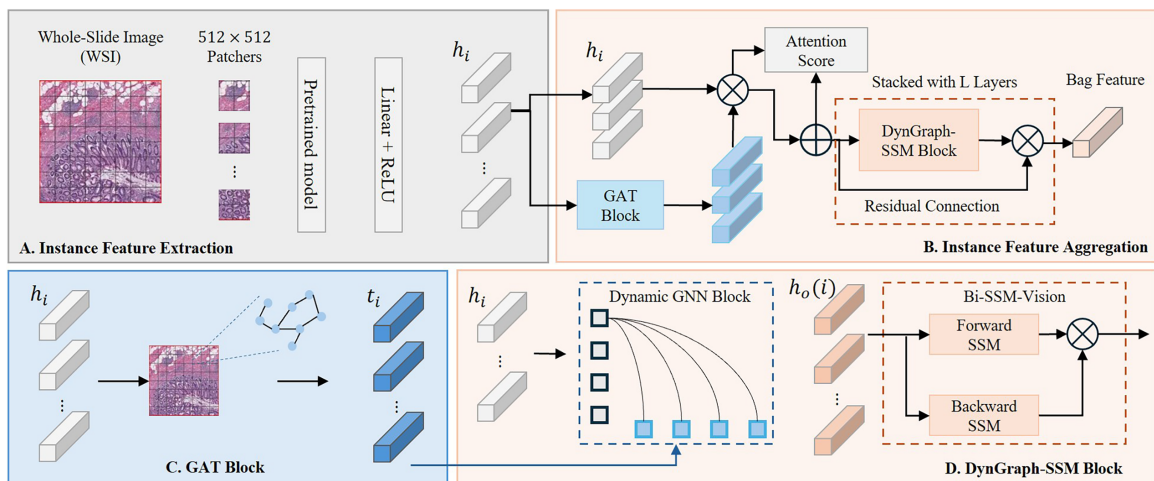


Figure 17: Schematic overview of the DG-SSM-MIL framework for whole-slide histopathology image classification (redrawn from [84]).

This architecture helps in ensuring the continuity of the local region and global consistencies, which is important for large scale analysis in histopathology. Since patches represent contiguous regions, the gate term G_t controls the process of contextual blending, thus preventing over-smoothness in heterogeneous cellular spaces.

In the GraphMamba (Medical Image Segmentation) model, this concept has been further expanded by implementing a scanning process guided by topology. The model generates a graph adaptively, depending upon the superpixels/regions adjacency, and uses GTAS to identify the scanning process order for Mamba. SCL ensures continuity at adjacent nodes, and the features of the nodes are processed according to the order in which they appear in the graph. The architecture of the GraphMamba (Medical Image Segmentation) model described in reference [85] is illustrated in Fig. 18.

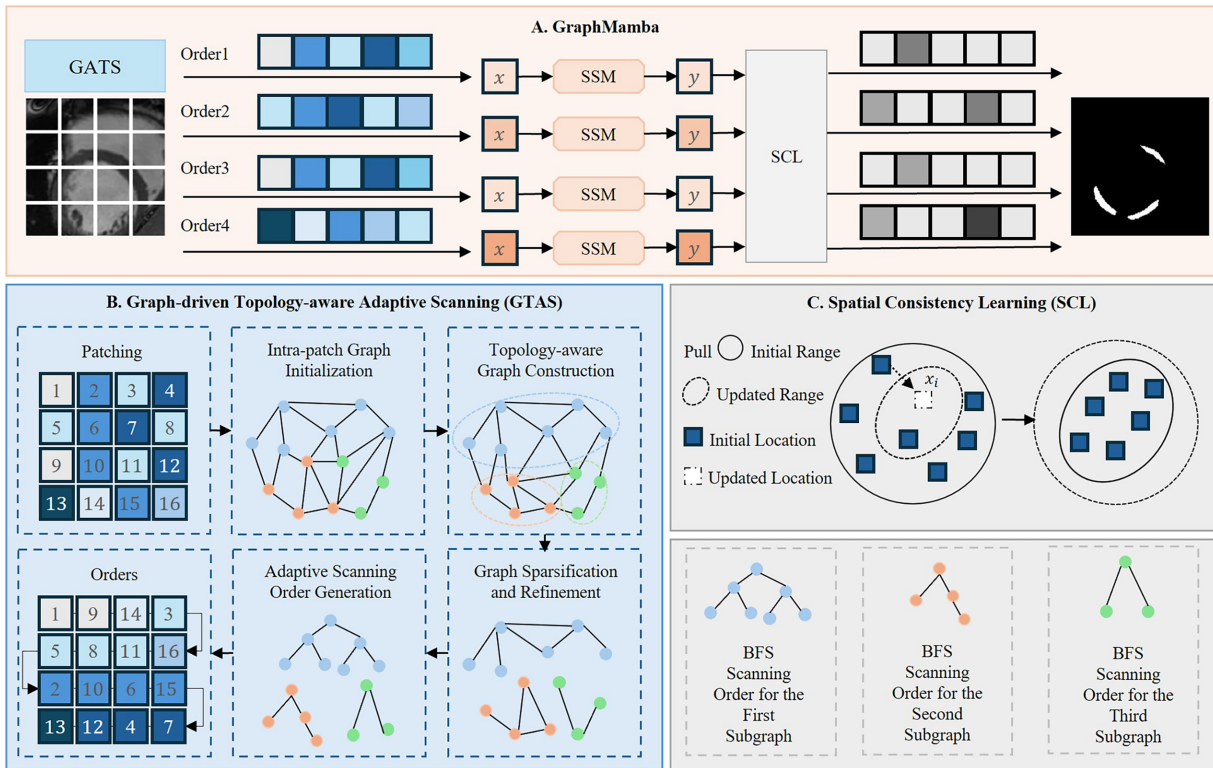


Figure 18: Schematic overview of the GraphMamba architecture for topology-aware medical image segmentation (redrawn from [85]).

The proposed architecture facilitates feature propagation without compromising the structure as a whole, which is important for anatomically-consistent segmentation applications. The experiments conducted on the LVCM, LiTS, and BTCV data reveal that topologically-constrained scanning delivers superior performance in terms of boundary detection when compared to baselines using purely CNN or transformer-based architectures.

(4) Cross-Domain Adaptability and Knowledge Transfer

Despite its vastly different application scales across molecular, neural, and imaging fields, GraphMamba demonstrates remarkable cross-domain adaptability thanks to its highly transferable underlying mechanisms. This versatility is primarily attributed to three core design features: first, the structural decoupling of the spatial graph module from the Mamba sequence module, enabling flexible reuse of sequence

processing components across tasks; second, a unified linear state update equation, providing a foundation for parameter-level transfer between different tasks; and most importantly, its state transition process is always explicitly guided by the graph adjacency matrix. In addition, the ability of the proposed approach enables the model to adapt its structure in response to different configurations of a network regardless of whether it is a molecule or a brain network by avoiding the need for changing the architecture altogether. Through the tight integration of graph topology reasoning and state-space dynamics manipulation, Graph-Mamba produces a biological computational platform that allows for the necessary adaptation to any diffusion or communication process.

4.2.3 Theoretical and Empirical Insights

The combination of graph-based learning and state-space dynamics, used in the biomedical Graph-Mamba paradigm, provides several theoretical models, all of which have been proven effective in practice. Regardless of whether it is applied to molecules, neurons, or images, this methodology always shows three main features, which include structural robustness, scale invariance, and interpretable propagation dynamics. This section highlights these key features through an analysis of the articles covered.

(1) Structural Stability and Controlled Information Flow

Recent applications, including RLMamba [108] and MambaMIL [109], employ Mamba-style selective state updates to improve feature retention and information filtering in task-specific settings. These results provide empirical support for the usefulness of recurrent selective propagation.

Being characterized by high complexity and large amounts of noise, biomedical data are known to be very hard to work with [110–112]. This property makes Mamba's stability very useful. For instance, consider Brain-Mamba [82] that can classify EEG and fMRI: it generates stable time trajectories despite the fact that the inputs are highly noisy and unstructured; moreover, it has higher convergence rates compared to attention mechanisms but does not sacrifice the signal's coherence. Such behavior can be observed in graph-based segmentation using GraphMamba [85]: by introducing GTAS to its architecture, it achieves numerical stability and prevents noise from accumulating due to its irregular graph topology—a problem faced by classical graph convolutions. In simple terms, the gate serves as a built-in regularizer that eliminates unnecessary activations while keeping the important ones. Similarly, it is also valid for molecular modeling, where the Innovative Mamba and Graph Transformer frameworks illustrate the same property for state propagation between protein residues and ligands regardless of the graph's size, achieving stable convergence without numerous gradient spikes.

(2) Cross-Scale Expressiveness and Dynamic Representation

The scalability of Graph-Mamba in terms of expressive capabilities within biological hierarchies is a common observation. Compared to conventional GNNs, which are confined by limited receptive field size and stationary adjacency matrix, Graph-Mamba employs continuous and adaptive filter mechanisms to direct signal propagation based on biological conditions.

Specifically, MambaCAttnGCN+ [81], a combination of Mamba's natural language representation and graph convolution techniques, was employed to represent sequence-based interactions of piRNA and higher-order connections among diseases. In the same way, *Multi-view Brain Network Mamba* revealed how selective state-space propagation could coordinate across different imaging data types, leading to an average of 3%–6% improvement in multi-view classification accuracy compared to fixed-topology GNNs.

Practically, Graph-Mamba models dynamically scale their state-space kernels at different levels. As illustrated in GraphMamba in medical image segmentation [85], Graph-Mamba automatically tuned the

bandwidth for spatial propagation, such that in homogeneous tissue areas, G_t would be low to maintain smooth representations; conversely, near the anatomical boundaries, G_t was higher to boost feature contrasts. As confirmed by the gate activations in the ablation study of the paper, such adaptive modulation offers empirical evidence for biological-awareness expressiveness.

(3) Interpretability and Biological Plausibility

Concerning the field of biomedical models, the importance of interpretability may be compared to that of prediction accuracy. Graph-Mamba models offer an approach allowing a clear way of analyzing the path of signals' propagation in a complicated architecture. Opposed to the opaque way used in classical attention mechanisms, Mamba uses explicit tracking of the intermediate hidden states and gating coefficients, thus allowing an analysis of how the information passes in a biological network.

An instance of this is Brain-Mamba [82]. Selective gate activations coincide with typical brain activity during cognitive functions, as confirmed by the robust activations located in the fronto-parietal and temporal areas of the brain that coincide with typical patterns of synchronization in neuroscience. It is likely that Mamba's gate activation represents real brain functions, not just statistical noise.

In medical imaging as well, Graph-Mamba [85] demonstrates interpretability. Visualization of gate maps in segmentation boundaries shows activation spikes near the borders of organs, thus confirming the existence of the model's selectivity in propagation of information in real anatomical structures. In molecular systems, the attention shifts to the state coupling of the residue subgraph with ligand subgraph.

In the evaluation of the MGDTA model [80], the authors observed that the Mamba-based representations placed greater emphasis on protein residues associated with known binding regions. This finding suggests that the learned state dynamics may capture functionally relevant patterns for drug-target affinity prediction. However, the reported correspondence provides supporting evidence rather than definitive proof of correct or causally interpretable latent trajectories.

(4) Cross-Domain Generalization and Transfer

Empirical evidence additionally supports Graph-Mamba's ability to facilitate knowledge transfer between different biological scales and modalities, which stems from the modularity of the decomposition that isolates graph-based encoders from the shared dynamics of the state space.

In AGIIBM [83], the Mamba encoder pretrained on resting-state fMRI was successfully applied for EEG classification using fine-tuning on only <10% of the parameters, yielding similar levels of accuracy.

GraphMamba [85] demonstrated that pretrained model weights on the BTCV dataset could be used for LiTS without any need for fine-tuning of the Mamba core module, suggesting that the selective state module could be reused as a biological prior in terms of spatial continuity.

The MambaCAtnGCN+ [81] approach utilized pretrained RNA encoders on disease datasets to show that learned dynamics of the Mamba model reflect biological sequence order.

Theoretical consideration shows that the use of such a selective recurrence operator is associated with a contractive mapping in the case where $\|I - G_t\| < 1$, thus ensuring a stable propagation regardless of varying topologies' cardinality. It has been found to be the key requirement for achieving domain stability both in molecule and neurograph models.

In the end, we show that Graph-Mamba manages to combine the aspects of stability, expressiveness, and interpretability within a bio-inspired state space. The use of theories of control and dynamical systems provides additional rigor for Graph-Mamba since the gating ensures bounded propagation energies and graph coupling introduces domain-specific biases into the model. Therefore, Graph-Mamba is not an artificial neural network but a mathematical basis.

4.3 Heterogeneous and Multi-Relational Graphs

Extending the Graph-Mamba framework to heterogeneous and multi-relation graphs is a logical next step. Homogeneous graphs treat all nodes as equally important; however, reality exhibits heterogeneity in terms of entity type and various contexts [113–115]. Mamba’s selective state space can regulate information exchange between different relations and time scales, and core branch analysis and multi-label classification schemes are commonly used to address structural noise associated with complex networks [116,117].

Mamba’s efficiency, combined with graph-related inductive biases, allows it to directly handle traditional graph tasks. There is a promising potential in the application of such techniques to 3D reconstruction [118], dynamic graph structure learning [119], and image feature extraction [120,121]. Such adaptability is clearly observable in the practical realm, ranging from HSI and LiDAR-based land cover classification [122] to biomedical soft tissue tracking [123], which confirms the robustness of Graph-Mamba’s fundamental concepts even within multimodal sensor networks.

Specifically for heterogeneous graph modeling [124], each node v_i and edge e_{ij} is assigned a semantic category, leading to a graph schema $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \phi, \psi)$ where $\phi(v_i)$ and $\psi(e_{ij})$ stand for the node and edge types, respectively. Heterogeneous graphs can model various types of graphs, such as social relations, knowledge ontologies, recommendation systems, and citation networks. Learning intra-type coherence and inter-type dependency becomes a core issue for such graphs. Conventional HGNNs are commonly based on either meta-path or type-specific encoders, which may ignore temporal dependency and higher-order relationships crossing relation types.

Graph-Mamba aims at addressing this issue by transforming such graphs into a continuous selective state-space. Specifically, the model employs type-aware gating to directly control the state transition of latent states. Instead of following strict message passing rules, the model performs selective routing through relation-wise subgraphs. This strategy is effective in learning non-local information with robustness.

Across recent studies, Mamba’s role in heterogeneous graph settings has expanded from temporal sequence processing to contextual reasoning over typed structures. The selective state-space has proven effective at adapting propagation strength to the semantics of each edge type.

In practice, this allows Mamba modules to act as relation-specific filters that capture high-order dependencies—for example, in social hypergraphs (friend $\rightarrow$ group $\rightarrow$ event), recommendation graphs (user $\rightarrow$ item $\rightarrow$ context), or citation networks (paper $\rightarrow$ author $\rightarrow$ venue).

Graph-Mamba’s hybrid design thus provides two complementary advantages: gating functions control which relations contribute to current state updates, thereby mitigating over-smoothing across irrelevant node types; and the state-space formulation bridges textual, symbolic, and relational modalities, enabling joint learning from structured and unstructured information sources.

The heterogeneity of data structures across domains is evident in the diversity of recent Graph-Mamba applications, four of which represent the current frontier of this research direction. Table 10 provides basic information about the papers discussed.

A distinct trend is evident across these models: Graph-Mamba architecture progress from solely spatial reasoning to relational-contextual reasoning, conceptualising the graph as a structured manifold upon which selective dynamics occur. This change shows how general Graph-Mamba is; the framework is no longer limited to time or space but can be used in any area where different types of relationships affect how a system works.

Table 10: Core literature in heterogeneous and multi-relational Graph-Mamba research.

Model	Domain	Graph Type	Core Task	Key Interaction with Mamba Mechanism
HGM-SRR [125]	Social Network Analysis	Hyper-relational graph (people ↔ group ↔ event)	High-order social relation inference	Introduces hyperedge-aware selective state propagation to capture multi-entity dependencies across social contexts.
CIM-SR [126]	Recommender Systems	User-Item heterogeneous interaction graph	Sequential recommendation and collaborative filtering	Couples Mamba sequence module with graph-based collaborative information propagation for context-adaptive recommendation.
AdaptQA [127]	Knowledge Graph & NLP	Heterogeneous knowledge graph (entities, relations, texts)	Answer generation with knowledge reasoning	Integrates Mamba state space for textual context propagation within knowledge-graph structures, bridging language and graph modalities.
EMK-KEN [128]	Scholarly Network Analysis	Attributed citation graph (paper, author, venue)	Knowledge-value assessment and citation impact prediction	Employs Mamba for temporal and semantic state tracking within heterogeneous citation networks, integrated with graph-based knowledge propagation.

Theoretically, Mamba's state-space dynamics offer a continuous-time framework for disseminating information across discrete heterogeneous structures. Each type of relation adds to a different part of the overall system dynamics, which makes it possible to understand how states change in relation to each other. Following this overview, the next subsections look at four examples of how heterogeneous Graph-Mamba systems can be used.

4.3.1 Hyper-Relational Graphs and Social Reasoning

The hyper-relational graph representation model [129] can effectively handle the multi-entity and multi-context nature of social systems. Social interaction data in [125] is modeled into a heterogeneous hypergraph, as shown in Eq. (27):

$$\mathcal{H} = (\mathcal{V}, \mathcal{E}_h, \Phi, \Psi), \quad (27)$$

Define $\mathcal{H} = (\mathcal{V}, \mathcal{E}_h, \Phi, \Psi)$ as a heterogeneous graph where $\mathcal{V}$ consists of nodes of various types, such as people, groups, and incidents. The hyperedge collection $\mathcal{E}_h$ represents high-order relationships between several elements that appear together in the same social or physical environment. The functions $\Phi(v)$ and $\Psi(e_h)$ map each node v and hyperedge e_h to their corresponding types for better feature handling based on their types.

These nodes have multimodal features, such as images, texts, and metadata that need to be preprocessed before embedding them in the latent space and building the graph. The hyperedges can be formed using either of the following two methods: (1) Hyperedges are formed when entities/individuals participate in the same event instance together; or (2) The formation of hyperedges is dependent on the co-occurrence rate or

semantic proximity of the entities (group membership, history of communication). This way, more than one individual can be connected via one hyperedge.

(1) Dataset Structure Description

Table 11 provides several benchmark datasets covering visual, textual, and social modalities, laying the foundation for this paper's evaluation of higher-order relational reasoning capabilities.

Table 11: Summary of multimodal social relation and contextual affinity benchmark datasets.

Model	Domain	Scale	Relation Categories
PISC-F [130]	Image-based social relation dataset	24,457 images; 76,573 social relations	Friend (Fri), family (Fam), couple (Cou), professional (Pro), commercial (Com), and no relation (No).
PISC-C [130]	Image-based social relation dataset	17,142 images; 64,514 social relations	<i>Intimate</i> (Int), <i>non-intimate</i> (Non), and No.
PIPA [131]	Photo album dataset for identity and relation classification	37,107 images, 2356 identities	<i>Attachment</i> (father-child (r0), mother-child (r1), grandpa-grandchild (r2), grandma-grandchild (r3)), <i>reciprocity</i> (friends (r4), siblings (r5), classmates (r6)), <i>mating</i> (lovers/spouses (r7)), <i>hierarchical power</i> (presenter-audience (r8), teacher-student (r9), trainer-trainee (r10), leader-subordinate (r11)), and <i>coalitional groups</i> (band members (r12), dance team members (r13), sport team members (r14), colleagues (r15)).

In each case, the dataset contains multi-person images that can be used to create a hypergraph where: Nodes represent people in the image; Hyperedges represent common context frames or events (all people in the image are one hyperedge). For PISC, there are annotated relations between specific pairs of people; for the hypergraph, there are implicit relations involving multiple people in the same image. On the other hand, PIPA focuses on the consistency of the same person's identity in multiple images, making relational reasoning possible. Such differences make these two datasets suitable to test the performance of Graph-Mamba in hyper-relational reasoning tasks.

The hypergraph generation algorithm for both datasets includes three main phases. Firstly, pretrained detectors are used for identifying persons in each frame, and then bounding boxes are used for extracting region-based features, which can be considered as node embeddings. Secondly, one hyperedge connects all persons detected in one image, and additional hyperedges are created when the person is visible in several images (as in PIPA). This allows forming a multi-layer hypergraph taking into account intra-scene and inter-scene relations. Lastly, visual and geometric features are embedded in the joint space resulting in combined node embeddings, suitable for hypergraph propagation.

Both datasets have inherent contextual dependencies where inter-person connections depend not only on the persons themselves but also on some common environment. Hypergraphs explicitly encode such dependencies at the level of groups, and at the same time, the Mamba model provides context propagation across hyperedges and images due to the selective state-space representation. These aspects allow for inferring

multi-level relations involving both intra-scene and inter-scene interactions. Therefore, recognizing social relations can be regarded as a proper benchmark problem for Graph-Mamba's hybrid relational state-space modeling method.

(2) HGM-SRR Model Architecture and Implementation Details

To provide additional insights into modeling hyper-relational reasoning, this research proposes the structural arrangement of the HGM-SRR model, an approach that builds upon the foundational concepts of Graph-Mamba models into higher-order relational scenarios. First introduced by Tang et al. [125], this model serves as a case study in utilizing Mamba-inspired state space modeling strategies within hypergraphs. As shown in Fig. 19, the HGM-SRR model consists of four main phases.

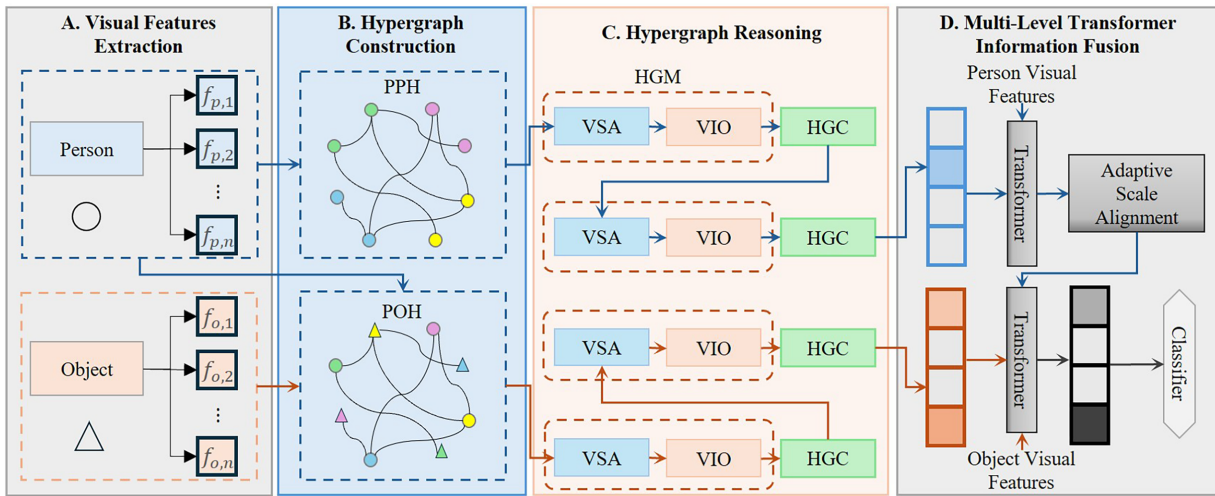


Figure 19: Overall framework of the Hypergraph Mamba (HGM-SRR) (redrawn from [125]).

The pipeline uses a two-stage encoder for learning visual representations of detected people and objects, mapping them into high-level vectors. The embedding vectors obtained through the cropping of people and objects are further organized into hypergraphs, namely the Person-Person Hypergraph (PPH) and the Person-Object Hypergraph (POH), which capture multipartite social relationships and scene semantics.

The core reasoning procedure operates on the aforementioned hypergraphs using the Hypergraph Mamba (HGM). Fig. 19 shows the architecture of HGM with its internal sequential flow. The main components of HGM include two special operators. First, the Vertex Selection Algorithm (VSA) selects necessary nodes to avoid ambiguity in the reasoning caused by the similarity between the embeddings. Next, the Vertex Interaction Operator (VIO) is applied to selectively scan vertices with the inspiration of the Mamba architecture to capture the global dependency between vertices. Following VIO in the sequence, the output is fed into a Hypergraph Convolution (HGC) layer. Finally, the structural information derived from hypergraph reasoning will be merged into the visual representation via Multi-Level Transformer (MLT) to obtain accurate relation predictions.

In summary, through combining hypergraph learning and state-space reasoning, the proposed method goes beyond the standard pairwise relationships. The key idea of this paper is extending Mamba's selective propagation to hyper-relational graphs. As shown on Fig. 20.

This allows for efficient inference at the vertex level while ensuring consistency of information in multi-hop and multi-type relationships, combining visual cues with structural reasoning when analyzing social activities.

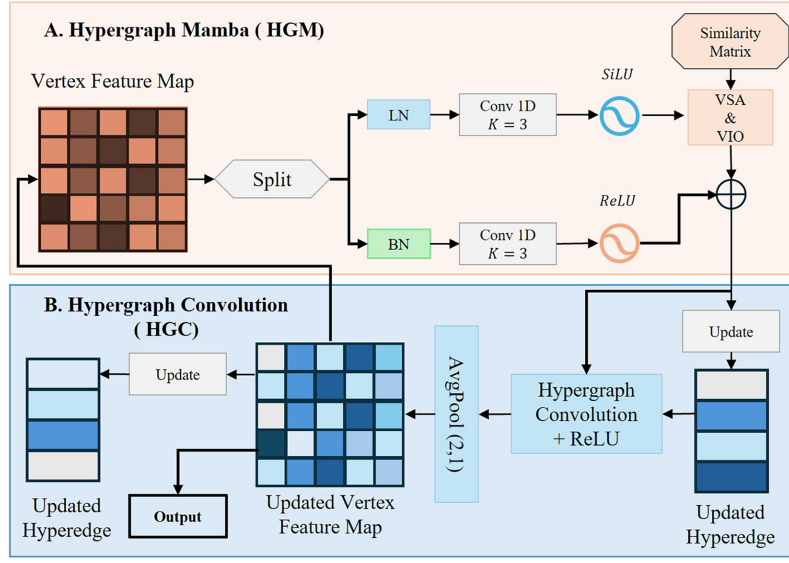


Figure 20: Internal structure of the Hypergraph Mamba (HGM) reasoning module (redrawn from [125]).

4.3.2 Sequential and Collaborative Graphs for Recommendations

The success of sequential recommendations lies in the merging of user-item interactions with the effects of changes in the behavior of users. Graph-based traditional recommendation systems do have ways of integrating collaborative information within their system; nevertheless, they often fail in dealing with temporal issues especially concerning the effects of changes in the history and strength of interactions. In contrast, sequential recommendation methods consider user histories as disconnected sequences, disregarding the complex relationship between users and items due to context sharing.

CIM-SR [126] uses the concept of heterogeneous interaction graph along with sequential state space model using Mamba. This system combines collaborative and sequential information by considering them not as separate things but together.

(1) Data Structure and Heterogeneous Graph Representation

A heterogeneous bipartite graph as shown in Eq. (28):

$$\mathcal{G} = (\mathcal{U}, \mathcal{I}, \mathcal{E}), \quad (28)$$

where: $\mathcal{U} = \{u_1, u_2, \dots, u_m\}$ represents users, $\mathcal{I} = \{i_1, i_2, \dots, i_n\}$ represents items (products, movies, or services), and $\mathcal{E} \subseteq \mathcal{U} \times \mathcal{I}$ represents user-item interaction edges derived from implicit feedback.

The edge relation e_{ui} is related to a chronological order t_{ui} and has attributes including timestamp, session ID (denoting temporal order), action type (such as click, add to cart, and purchase), and information about the content (such as category, brand, and genre). By combining the above-mentioned aspects, one can construct a time-aware heterogeneous graph where the edges not only represent the relationships between users and items but also capture the dependence between sessions [132]. As shown in Table 12.

An interesting feature of these databases is that the users are not alone; instead, they create virtual groups based on common interests in items. Eq. (29) describes the connectivity matrix between users, items, and users:

$$A_{uu'} = \frac{1}{|\mathcal{I}_u \cap \mathcal{I}_{u'}| + 1}, \quad (29)$$

where $\mathcal{I}_u$ denotes the set of items interacted with by user u . This forms a *second-order graph* that encodes indirect user similarity, complementing the direct user-item interactions in the original bipartite graph.

Table 12: Dataset in heterogeneous and multi-relational Graph-Mamba research.

Dataset	Items	Users	Actions	Avg. Seq. Length	Sparsity
Beauty	12,101	22,363	198,502	8.9	99.93%
Toys	11,924	19,412	167,597	8.6	99.93%
ML-1M	3416	6040	999,611	165.5	95.16%

The final architecture takes the form of a heterogeneous multilayer graph, which consists of both user-item and user-user subgraphs. In such a setting, the model can pass information not only through behavioral sequences but also through collaborative relationships, thus keeping track of high-order preferences in groups of similar users.

When it comes to sequential modeling, for every user, interaction data is split into context-aware segments. Every segment s_k corresponds to a subgraph G_{s_k} , representing the graph structure in terms of user-item interactions for that specific time frame. It creates a link between the graph topological structure and temporal order. Node representations are serialized for each session-graph, before being fed into the Mamba model, according to interaction order.

(2) CIM-SR Model Architecture and Implementation Details

The CIM-SR architecture, shown in Fig. 21, integrates two complementary encoding views—a *sequential view* and a *collaborative graph view*—connected via contrastive learning alignment.

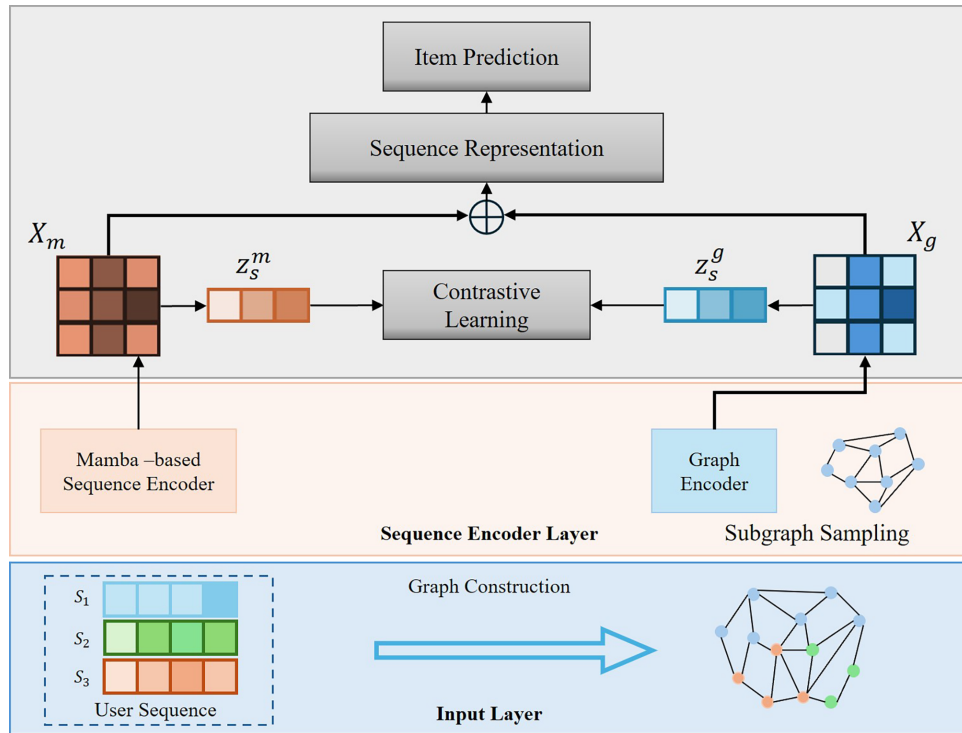


Figure 21: Structure of the CIM-SR model (redrawn from [126]).

To build a global collaborative graph $G = (V, E)$ from all the training sequences, we first make nodes that stand for items and edges that show how often items appear together in user interaction histories. The weight of the edge is inversely related to the distance between items, which measures how close they are to each other. This graph combines behavioural data from around the world, which lets items with similar interaction contexts share richer embeddings.

The Mamba-based encoder learns temporal relationships in each user sequence individually. The Mamba framework ensures linear complexity by selectively updating states and preserving only the most important historical context, an approach that diverges from the Transformer paradigm. Multiple Mamba layers result in deep representations of sequential data, successfully capturing the dynamics of evolving user preferences.

In CIM-SR, a local graph is created for each user sequence and message passing is performed using LightGCN. In this manner, collaborative information that exists between user sequences can be learned. CIM-SR uses contrastive learning to make the representations from the sequential and graph encoders work together. Positive pairs are made up of representations of the same sequence from both views, while negative pairs are made up of representations of different sequences. This alignment makes sure that embeddings learned in different relational spaces stay semantically consistent, which makes them more stable and generalisable.

Finally, we derive the representation of the sequence by merging the outputs from both the encoders and then feeding the merged representation to the softmax prediction layer that finds out the highest probability next element. The overall aim of the training process is to maximize both recommendation loss and contrastive alignment loss.

In conclusion, the introduction of CIM-SR is vital for Mamba's ability to solve sequential recommendation problems and collaborative recommendation problems. CIM-SR captures the temporal local behavior alongside the global relational context via contrastive multi-view learning on global collaborative graphs. As such, it provides more accurate predictions than Transformer-based methods with less computation time.

4.3.3 Knowledge Graphs and Textual Reasoning

Knowledge graphs (KGs) [133] and natural language form two modalities of machine reasoning: KGs encode symbolic knowledge, while text encodes contextual and unstructured knowledge. The goal of knowledge-based NLU is precisely to make use of this modality gap. But traditional approaches for GNN reasoning over KGs often overfit to the neighborhood relational information, while conventional approaches based on language models fail to ensure factual consistency in the process of inference [134–136].

In response to this problem, AdaptQA [127] proposes Graph-Mamba, which combines symbolic reasoning from graphs with selective state-space propagation in language space to allow fusion of knowledge and context when answering questions.

(1) Data Structure and Heterogeneous Graph Representation

The AdaptQA framework was evaluated on two large-scale, open-domain multi-hop question answering datasets, HotpotQA and WikiHop, both of which provide rich textual and relational information suitable for constructing knowledge graphs. Rather than being treated solely as text corpora, these datasets were transformed into structured relational graphs to enable effective reasoning through graph learning modules. Table 13 presents key statistics for the two datasets.

The preprocessing stage of AdaptQA utilizes ReadOIE and DetIE frameworks to extract entities and establish links among them in order to produce a triplet structure in the format of (subject, relation, object).

Next, the triplets extracted during the preprocessing phase serve as input to generate a question-oriented knowledge subgraph. Specifically, the subgraph involves nodes that represent entities present both in the question and support, while edges denote relations between entities established through coreferences and syntactic parsing. Edge weights, on the other hand, denote semantic similarity, calculated by the RoBERTa encoder.

Table 13: Statistics and characteristics of the HotpotQA and WikiHop datasets.

Dataset	Train	Test	Type	Characteristic
HotpotQA [137]	90,447	7405	Multi-hop QA	Requires reasoning over multiple Wikipedia documents; includes bridge and comparison questions.
WikiHop [138]	43,738	5129	Reading Comprehension	Multi-selective reading; tests entity linking and relational reasoning across multiple passages.

Thus, the graph incorporates not only explicitly mentioned Wikipedia relations but also latent semantic connections allowing the model to conduct several reasoning hops while predicting the answer. Typically, each sample involves a graph comprised of 20–40 entities and 50–100 relations, which is compact enough to allow state-space reasoning while remaining relationally complete.

(2) AdaptQA Model Description

The AdaptQA system works through a series of steps in which knowledge acquisition, graph reasoning, and language optimization are tightly integrated (as shown in Fig. 22). What makes this system different is its treatment of graphs like HotpotQA and WikiHop. The AdaptQA model does not consider graphs to be additional sources of background knowledge.

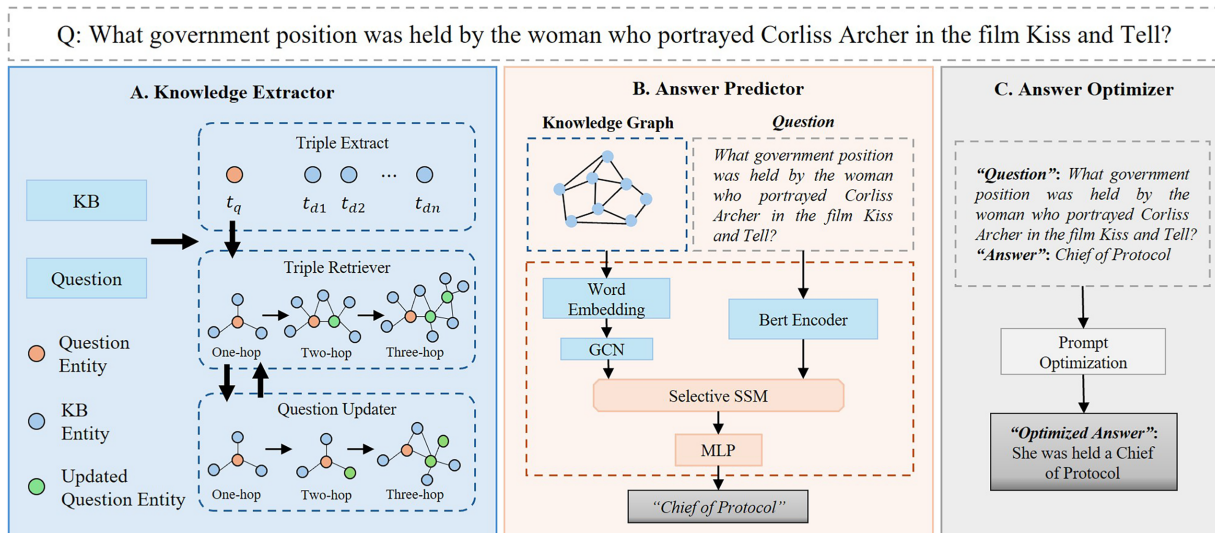


Figure 22: Structure of the AdaptQA model (redrawn from [127]).

The AdaptQA model uses the Adaptive Retrieval Module in the first stage by setting the number of hops based on the level of semantic complexity associated with each unique question. Contrary to static hop

systems that end up truncating or duplicating information, the AdaptQA model ensures contextual integrity of every entity relation. The context-aware update of the question embedding is achieved via contextual attention with each reasoning hop, making sure that the retrieved subgraph becomes increasingly relevant to the question semantics in every hop. After all hops, we obtain a question-specific knowledge subgraph consisting of relevant entities and their relations in terms of the required logical reasoning pathway to respond to the question.

In the second phase, the key component of the AdaptQA architecture is the Graph-Based Module (GBM), which implements a reasoning process over a dynamically generated knowledge graph. The GBM combines two techniques:

- (1) A light-weight graph convolution operation, which captures the semantics from adjacent nodes. Unlike other GCNs and GATs where propagation is constrained to a fixed number of steps, in the GBM, node states can be updated based on adaptive propagation intensity, where the propagation intensity is proportional to the semantic similarity score of the neighboring nodes computed during knowledge extraction.
- (2) Based on the Mamba selective-state model, the GBM uses a gate unit to filter out irrelevant information from the node states. The node has an internal state vector, which selectively aggregates incoming messages from its neighboring nodes, allowing long-term dependencies with linear computational complexity to be modeled in a single pass through the graph.

In summary, combining the above two techniques allows the GBM to capture both local and global structure in the graph while preventing over-smoothing. Moreover, the GBM is interpretable in nature, which can visualize the reasoning process across hops and demonstrate dependencies between entities. This characteristic cannot be achieved in previous Natural Answer Generation (NAG) architectures using transformers.

In the last phase, optimization of answers is done using Large Language Models (LLMs) where GPT-3 or GPT-4 models are used for post-editing of the text generated using linguistic abilities. Through the few-shot prompting method, grammatical and factual consistency and fluency of the textual output by the GBM are improved without affecting the factual content inferred by reasoning in the previous steps. In each prompt, an example containing the question, reasoning trace, and initial answer is included so that the LLM can mimic human writing without compromising on logic.

All three phases above contribute to the creation of a pipeline where graph-based reasoning and language generation techniques complement each other. Retrieval ensures semantic adequacy, inference of interpretable graph-level representations is done by GBM, and LLM helps in enhancing expressiveness. By doing so, factual and grammatical adequacy of answers have been improved.

4.3.4 Citation and Knowledge Networks

The process of transferring knowledge inside a scientific ecosystem is inherently based on the concepts of graphs and time. Nodes, which include papers, authors, and journals, are interrelated in a multitude of ways, including citations, collaboration, and similarity of topics. In other words, citation graphs are a good case study that can be analyzed using the concept of heterogeneous and temporal graph analysis, where temporal relations capture the process of creation, propagation, and evolution of knowledge over time. A prime example of this type of research is EMK-KEN [128].

(1) Data Foundations: Multi-Entity and Multi-Relational Structures

To evaluate the generalizability of EMK-KEN on different citation networks, they have tested the algorithm on ten citation networks, each representing a different type of citation pattern among domains that include physics, computer science, biomedicine, and social sciences, as shown in [Table 14](#).

Table 14: Summary of cross-domain citation and knowledge networks.

Dataset	Nodes	Edges	Feature Dim.	Density	Description/Domain
ACMHypertextECHT [139]	728	5326	771	0.3396	Hypertext conference
Oxytocin [140]	1467	86,535	770	0.3094	Biomedical
ACMIS [141]	16,148	128,027	768	0.3317	Management information systems
SNAP-HEP-TH [142]	27,770	352,807	768	0.2813	High-energy physics (theoretical)
ALTEGRAD Challenge2021 [143]	100,336	2,633,688	768	0.2671	AI text-mining challenge dataset
OGBN-ArXiv [144]	169,343	1,166,243	128	0.3094	Open Graph Benchmark
Citation Networks-V12 [145]	4,894,081	45,564,149	769	0.1743	Large-scale cross-domain
DBLP-V13 [145]	5,354,309	48,277,950	768	0.1603	Computer science bibliographic
AMN (AI + ML + NLP) [144]	60,646	264,505	772	0.1362	AI-ML-NLP interdisciplinary
Biomedicine [140]	31,546	211,883	772	0.1110	Biomedical and bioinformatics subfields

To support the EMK-KEN model architecture, each data set is converted into a graph representation. Here, each paper becomes a node in the graph, complete with its attributes, which consist not only of text embedding vectors from specialized models such as SciBERT or MatSciBERT but also metadata including the author, publishing venue, and publication year. Each citation relationship is denoted by a directed edge; when dealing with cross-disciplinary datasets such as DBLP-V13, the weight of the edges is adjusted according to the citations made to avoid any biases caused by well-cited individuals or journals.

Through this approach, the EMK-KEN model accurately maintains citation networks in both intra-domain and inter-domain formats, thus enabling EMK-KEN to track knowledge dissemination channels such as from AI to NLP or from physics to mathematics, consistent with the focus of the paper.

EMK-KEN additionally develops the Knowledge Quantification Index based on changes in graph entropy. $KQI(v_i)$ considers both the in-degree of a node and the community centrality to determine how much information a paper contributes to the graph-based information architecture. Distribution of the KQI across the dataset reflects the evolution of knowledge development across different fields, where AI/ML shows relatively high variation in KQI due to rapid innovations in these fields while biomedicine displays steady growth and high KQI peak due to consistent knowledge building.

(2) EMK-KEN Model Description

EMK-KEN model, on the other hand, is a graph-sequenced based network architecture, which was designed to calculate the worth of knowledge within a very large citation graph. The design combines both the sequencing capabilities of Mamba and the non-linear capabilities of Kolmogorov Arnold Network (KAN) [146]. This dual-branched approach is capable of calculating the worth of knowledge across different academic disciplines. The original paper by the researchers [128], from which Fig. 23 shows the design of the model, will provide further insight into the model.

The first stage involves the standardization of heterogeneous features of nodes using information extracted both from metadata as well as the text-based embeddings. Considering that each node in terms of papers can have high-dimensional and sparse features, the MetaFP module (as shown in Fig. 24) utilizes linear transformations, normalization, and activation for increasing the feature density. Furthermore, relative position embedding is also used in this module to retain the directionality of the citations and thus enable the explicit modeling of the temporal sequence of citations. Therefore, this module acts as an interface between the raw bibliometric data and the next steps.

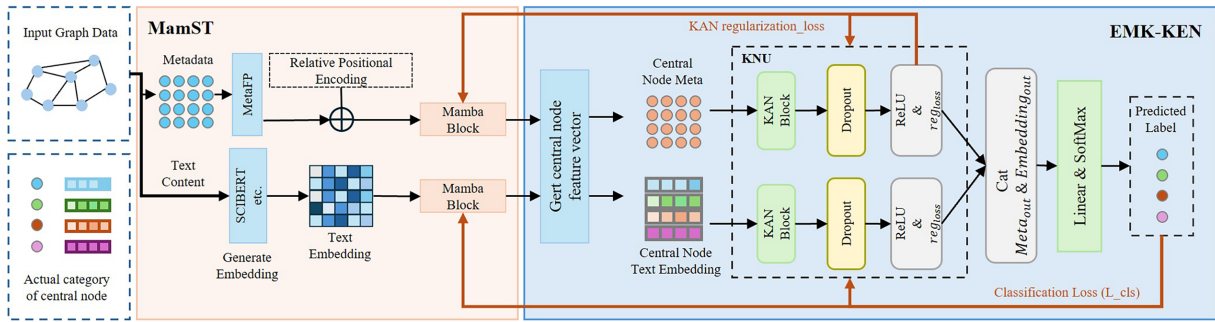


Figure 23: Structure of the EMK-KEN framework (redrawn from [128]).

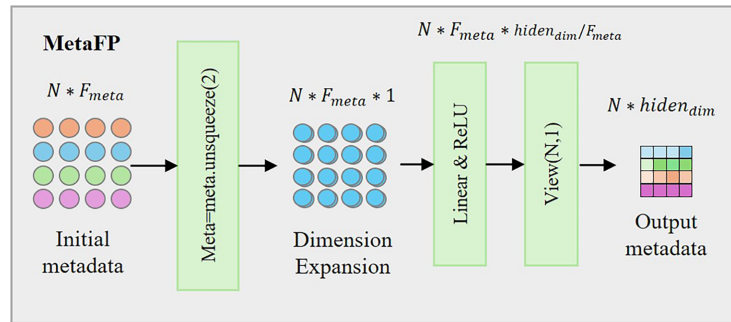


Figure 24: Structure of the MetaFP module (redrawn from [128]).

In the second phase, Mamba's selective state space model (SSM) is employed for an efficient modeling of the long-term dependencies in the citation graph. The node embeddings are viewed as elements in a temporal-structural sequence representing citation flow propagation. By employing a gating scheme, MamST can selectively update its internal state such that it can maintain important historical citations while ignoring other non-relevant edges. Unlike the traditional attention schemes that are computationally expensive, MamST can handle millions of nodes in linear time ($O(N)$). Thus, MamST serves as a dynamic memory filter that simplifies huge citation graphs into concise knowledge flows.

In the third step, KAN acts as a nonlinear projection layer following sequential encoding. Using B-spline adaptive activation functions instead of regular fully connected layers in the network, KAN provides exact control over nonlinear mapping operations. This process avoids loss of minor structural changes in the influence of knowledge, a major drawback that arises using regular linear mappings. At the same time, the KNU model utilizes dropout techniques and weight normalization for retaining training stability and improving cross-domain accuracy. The outputs from MamST and KNU are then combined to create a multiscale representation of the knowledge impact. Finally, a classification layer is employed to map the multiscale representation to a Knowledge Quantification Index (KQI), quantifying the scholarly impact in each case.

Architecturally speaking, EMK-KEN is based on the idea of extending the selective state space framework of Graph-Mamba into the academic domain. This is done using the basic property of Mamba, which is the propagation of information selectively. Since EMK-KEN's internal state only changes due to highly relevant structural/semantic input, the problem of smoothing and redundancy commonly observed in conventional deep graph neural networks is naturally prevented by design.

The selective scanning of nodes and edges was initially conceived in Graph-Mamba to combine graph topology with state-space models for dependency tracking. EMK-KEN uses the same notion in the academic

domain where the nodes refer to actual papers and edges are the citation paths between them. Instead of engaging in generic graph reasoning, EMK-KEN directly applies its selective-state mechanism to track information propagation and calculate entropy within citation networks.

In conclusion, the paradigm shift heralded by Graph-Mamba manifests across a wide spectrum of physical, biological, and informational structures. To synthesize the architectural nuances and conceptual threads developed throughout this section, [Table A2](#) provides a comprehensive cross-domain taxonomy. By encapsulating node-edge semantics, core integration mechanisms, and domain-specific empirical advantages of the 16 foundational models discussed, this matrix highlights the systemic scalability and mathematical robustness of coupling state-space dynamics with complex graph topologies.

5 Discussion and Theoretical Integration

5.1 From Domain-Specific Observations to Unified Dynamics

In all the application studies in Chapter 4, Graph-Mamba is observed to be seamlessly combining graph structures with dynamics processes. Although the studied areas cover a vast range of disciplines, the model shows three consistent tendencies regarding its operation.

Firstly, the framework serves as an excellent filter of information. Instead of aggregating the information from all neighbors regardless of the circumstances as most GNN models do, the gating of Mamba actively controls the process. It is especially important in the case of spatiotemporal and citation networks since tracking dependencies in such systems can be prone to noise if not controlled.

Secondly, the model does not require any additional mechanisms to operate at different scales. The continuous development of the latent variable makes it possible to adapt to different hierarchical levels naturally—for example, network modules or citation clusters. At its core, the model connects graph convolution to dynamical systems by blending graph structure with time dynamics.

Finally, Graph-Mamba tightly binds semantics and structure in applications when both are present, such as in case of biomedical ontologies and text or symbol-based graphs. The selective propagation of Mamba aligns the meaning of nodes with physical connections, making the two properties grow in parallel in one representation trajectory rather than separately. Graph-Mamba is an efficient mathematical framework for structured data analysis. No matter if the system includes traffic networks or molecules, the mechanism stays the same—stable, explainable, and flexible information transfer.

5.2 A Unified Framework for Structured Dynamical Learning

The cumulative insights from the preceding chapters substantiate a triadic theoretical framework that underpins Graph-Mamba, as shown in [Eq. \(30\)](#):

$$\mathcal{F}_{GM} = \langle \mathcal{G}, \mathcal{S}, \mathcal{P} \rangle, \quad (30)$$

$\mathcal{G}$ represents the graph topology, which can be described in terms of structural prior information, adjacency mappings, or relation schemes. $\mathcal{S}$ represents the state space representation, which consists of continuous latent dynamics that evolve over time or contexts. $\mathcal{P}$ represents the propagation process, which is governed by the dynamic gates $g_t \in [0, 1]$.

The triadic approach captures the two sides of the complex network structure, namely, the combinatorial aspect of the graph edges and the dynamics of the evolution described by the recurrence relations developed by Mamba. From a theoretical point of view, one of the major consequences of this approach is that the graph learning problem can be viewed as a continuum process. If $g_t = 1$ then Graph-Mamba becomes a static

message-passing algorithm that operates on a fixed graph. On the other hand, for $g_t = 0$, Graph-Mamba becomes a pure state-space dynamics problem without any restrictions on the graph topology.

This serves as an indication of the versatile nature of Graph-Mamba and its ability to be applied across several different fields due to its flexibility towards levels of coupling between the relational aspect and the dynamic aspects based on data characteristics.

In its essence, Graph-Mamba presents a coherent mathematical model that incorporates all three of these elements into the process of structured dynamic learning. This triptych approach allows for the incorporation of various models in one formal framework while providing one consistent way of looking at dynamic modeling of graphs for future events.

5.3 Interpretability and Cross-Domain Generalization

The first advantage of Graph-Mamba is the high adaptability of its mechanism of information propagation. Here the adjacency matrix of the graph dictates when the information may be propagated, while Mamba gates dictate how much and to what degree. Therefore, the association between gating patterns and influential elements becomes more apparent. The identification of the important elements in a complex network becomes can be explored by relating connection states to Mamba gating patterns.

The adaptive gating mechanism of Graph-Mamba can support generalization across structurally diverse graph-learning tasks. However, its robustness depends on factors such as graph density, dataset scale, serialization strategy, and regularization. In particular, performance on extremely sparse or small-scale graphs may be less reliable because selective state modeling can overfit limited structural information.

In terms of theory, this combination of interpretability and robustness reflects the “structural transparency principle”: the more transparent and evident the propagation mechanisms of the network become, the more stable and adaptable the network turns out to be. Explicit gating of the Graph-Mamba model acts not only as a computational component of this method but also as a main connection between structural transparency and robustness.

To summarize, the explicit treatment of spatial topological and temporal aspects of the data makes Graph-Mamba models predictable and interpretable, while at the same time enables effective transfer of such models to different domains without losing their semantic or structural consistency.

5.4 Limitations and Open Challenges

Despite the potential of Graph-Mamba, it is not a universal panacea. Researchers should carefully evaluate the necessity of integrating Mamba based on graph characteristics, as simpler architectures are often sufficient or even preferable under certain conditions:

- (1) Shallow and highly homophilous graphs (e.g., Cora, Citeseer): local aggregation in GCN/GAT is competitive, and Mamba’s added complexity does not pay off.
- (2) Lack of inherent sequential dynamics: forcing a 1D selective scan on static graphs may introduce arbitrary inductive biases; permutation-invariant message passing or Graph Transformers may be more natural.
- (3) Extremely sparse and small-scale graphs: Mamba’s selective memory can lead to overfitting; lightweight GNNs are more robust.

Even where Graph-Mamba is applicable, further optimization is required. Scalability remains a key issue: although Graph-Mamba avoids the quadratic cost of self-attention, memory access and state updating on highly dynamic graphs (e.g., social networks) must be carefully managed. Theoretical limitations also persist—stability has only been shown empirically, not proven; currently Graph-Mamba is a heuristic

algorithm, lacking convergence and energy conservation guarantees. Interpretability beyond the gating mechanism needs more research: while activation maps provide local explanations, global causality (e.g., how long-range dependencies affect the system) requires new methods, such as Lyapunov dynamical analysis or mutual information decomposition. Finally, evaluation methods must be standardised—current preprocessing varies wildly across datasets, making fair comparison impossible. A common benchmarking suite is imperative.

On the whole, these problems indicate that Graph-Mamba serves as a stepping stone, not a stopping point. Its key contribution lies in both efficiency improvements and providing a platform for further advancements.

5.5 Summary

In this chapter, all arguments revolve around one main idea: Graph-Mamba represents a generic paradigm for structured dynamic learning. From physics to biology and to social settings, Graph-Mamba proves that the heart of complex systems lies in the interaction among the graph, state spaces, and selective propagation mechanisms. While the three principles represent the pillars of modern models, they provide an integrated understanding of graph-based dynamic intelligence.

According to empirical findings, Graph-Mamba successfully reaches an optimum compromise between local connectivity and global diffusion processes. By implementing selective gate mechanisms, Graph-Mamba may reduce the propagation of task-irrelevant information through selective state updates. The ability of Graph-Mamba to achieve good generalization over heterogeneous relational data stems from its perfect balance between structure and dynamics.

From a theoretical perspective, the ternary framework provides a principled abstraction that integrates discrete and continuous learning perspectives. Graph-Mamba transcends the binary opposition between static representation learning and sequential dynamics by embedding combinatorial graph relations into a continuous state space controlled by selectivity. It provides a mathematical model of an evolutionary structure: a system that acquires relational geometry and transformational dynamics through unique propagation laws.

The fact that the learned representations may offer useful interpretive cues in some scenarios, although their consistency and causal validity remain to be established bears equal weight since it confirms Graph-Mamba's potential to function as a model of science rather than a collection of techniques used to perform some specific tasks. Namely, Graph-Mamba allows for interpretation of its inner mechanisms since adjacency propagation, gate operation, and state evolution are explicitly defined, thus becoming part of a wider context of knowledge. As a result, Graph-Mamba becomes a modeling framework, moving beyond pure prediction and towards explanations and consistency with scientific theories.

As discussed in [Section 5.4](#), the area is still undergoing rapid development. There remain many unresolved questions related to scalability issues, theoretical foundations, benchmarking, and consistency in evaluation, hindering the development of Graph-Mamba as a generic framework for dynamical graph intelligence. However, the problem of finding an acceptable trade-off between expressivity and interpretability remains at the center of research.

6 Conclusions and Future Directions

6.1 Emerging Research Directions

The concept synthesis introduced here sets the stage for Graph-Mamba as a theoretical framework for structured dynamical learning. Combining the topological information about a graph structure with the

state-space description of dynamics lays the groundwork for developing models that can account for the evolution of complex dynamical systems and their dependencies. Several potential future research directions can be identified here.

First, there is an important need to combine Graph-Mamba with foundation models on a larger scale. With the rise of general purpose models in the representation learning domain, future architectures may experiment with the application of select state-space architectures in multimodal or even language models. This will allow future developments to combine Graph-Mamba's approach with symbolic and relational learning as well as sequential data processing.

Another key point lies in the possibility of extending Graph-Mamba into a framework that includes physical constraints on the learned model. Since Graph-Mamba models are dynamic by nature, it becomes possible to learn models that account for any constraints arising from conservation laws or causal regularities. This extension of Graph-Mamba will enable it to incorporate physics in its models.

The third approach would involve extending Graph-Mamba towards cross-domain and multimodal learning. A possible direction in this respect would involve using visual, textual and numeric modalities within the same relational state space, thereby enabling organization of diverse data types by using similar principles of dynamics. This type of generalization seems especially applicable in fields like biomedical imaging, bioinformatics, and social information systems, which experience cross-modal interactions.

Continuous and life-long graph learning seems to be a promising area for future research. The mechanism of selecting gates used in Mamba could allow embedding concepts of adaptive memory into Graph-Mamba models that will consistently perform regardless of transformations occurring in relations. In other words, it may be possible to develop a model of Graph-Mamba as a dynamic system, whose structure will change simultaneously with relation changes.

In general, it seems like the path lies towards a scalable, multi-modal, and theoretical approach to Graph-Mamba. Future developments are expected not to lead to any improvements to accuracy or efficiency without integrating all the knowledge gathered on Graph-Mamba in the previous studies. It is more sensible to build upon already obtained results.

6.2 Methodological Challenges and Prospective Outlook

Even though the significance of Graph-Mamba is well understood, the implementation of Graph-Mamba is still at its preliminary phase, exhibiting several methodological problems. For instance, problems associated with scalability and efficiency continue to exist. With more graph-based data becoming available, the balance between expressiveness and resource limitations will determine the practicality of large-scale implementations.

Additionally, further theoretical research is necessary. Even though Graph-Mamba is capable of combining discrete graph representations with continuous state representations, stability, convergence, and control have yet to be proven mathematically. Further studies must be conducted in order to determine when selective propagation results in stability or interpretability.

Another limitation of Graph-Mamba concerns interpretability. Existing studies have provided preliminary evidence that activation gates and state-propagation patterns may align with task-relevant structural or semantic features. However, such observations remain largely correlational and do not establish causal relationships or guarantee consistent semantic interpretability across applications. Therefore, Graph-Mamba should be examined using broader interpretability frameworks that combine gate visualization with attribution analysis, perturbation tests, counterfactual evaluation, and domain-specific validation.

Standardized benchmarking is still a practical yet important limitation. Studies have been performed in different fields, but they present differences in relation to data set properties, performance measures, and preprocessing steps. Diversity makes comparison impossible and hinders the process of synthesis, while the implementation of community standards for testing dynamic graphs could provide better empirical grounds for studying Graph-Mamba.

On the whole, one can conclude that both architectural and conceptual development should be taken into account as essential factors in Graph-Mamba's evolution. In particular, how the above issues will be handled will determine if Graph-Mamba will become a theory of structured dynamical intelligence or just a series of models for solving certain tasks.

6.3 Conclusions

Graph-Mamba represents an emerging direction in graph learning that connects graph-structured representation with selective state-space modeling. Instead of treating graphs only as static collections of nodes and edges, Graph-Mamba views relational data as structured systems whose hidden states evolve under both topological constraints and input-dependent dynamics. In this sense, it provides a useful bridge between graph-based reasoning and dynamical modeling.

Across spatio-temporal, biomedical, heterogeneous, geometric, and visual domains, existing studies show that Graph-Mamba can adapt selective propagation to different graph semantics, data structures, and task requirements. Its main value lies in combining local structural priors with long-range state evolution, offering a potential balance among expressiveness, efficiency, robustness, and interpretability. However, this advantage remains context-dependent and should be understood in relation to specific graph types, serialization strategies, and evaluation settings.

This survey also has several limitations. First, the formal literature corpus is based mainly on WoS-indexed SCI-EXPANDED and CPCI-S records, which provides a controlled and reproducible basis for analysis but may underrepresent very recent preprints, conference-first studies, or works using alternative terminology. Second, the review is organized primarily by representative application domains rather than by a fully architecture-centered taxonomy. This choice reflects the current application-driven development of Graph-Mamba, but it may not capture every fine-grained methodological distinction among graph types, scanning strategies, serialization methods, and integration points. Third, this survey does not provide a unified benchmark or leaderboard-style comparison of performance, model size, parameter count, memory consumption, and runtime. The reviewed studies differ substantially in tasks, datasets, graph construction methods, evaluation metrics, prediction horizons, and hardware settings, and many original papers do not consistently report parameter size or computational cost. Therefore, cross-model comparison is discussed qualitatively where available rather than presented as a direct ranking.

More broadly, Graph-Mamba suggests that graph learning can be interpreted as a process of controlled structural evolution. This perspective may extend beyond the current Mamba architecture to other state-space or dynamical operators that preserve the principle of selective propagation on structured data. Future research may therefore move from model-specific designs toward more general frameworks for topology-aware state evolution.

In summary, Graph-Mamba is important not because it replaces existing GNNs, Transformers, or recurrent models, but because it highlights a new way of thinking about structured learning. It emphasizes that complex relational intelligence may arise from the interaction between topology, dynamics, and selective control. This makes Graph-Mamba a representative step toward future models that integrate graph structure with dynamic state-space reasoning.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Availability of Data and Materials: No new primary data were generated or analyzed in this review. All data and information discussed in the article were obtained from previously published studies, which are cited in the manuscript.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1: Relevant data sources.

Dataset	Links
Common Traffic and Mobility Datasets	https://drive.google.com/file/d/16xbiRvh5jJx2A4Sww9dEMeZ_Y0AEoQTM/view
Milan Cellular Traffic Dataset	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/UTLAHU
NTU RGB+D 60/120 [72]	https://rose1.ntu.edu.sg/dataset/actionRecognition/
UAV-Human [73]	https://github.com/sutdcv/UAV-Human
NW-UCLA [74]	https://www.kaggle.com/datasets/akshayjain22/n-ucla
Davis [95]	https://github.com/hkmztrk/DeepDTA/tree/master/data/davis
KIBA [96]	https://github.com/hkmztrk/DeepDTA/tree/master/data/kiba
BindingDB [97]	https://www.bindingdb.org/rwd/bind/index.jsp
MNDR 4.0 [93]/piRDisease v1.0 [94]	http://www.rnadisease.org/
BVFC and BVFC-MEG [98]	https://things-initiative.org
ADHD-200 [99]	https://fcon_1000.projects.nitrc.org/indi/adhd200/
TUH-EEG [100]	https://isip.piconepress.com/projects/nedc/html/tuh_eeg/
MPI-EEG [101]	https://fcon_1000.projects.nitrc.org/indi/retro/MPI_LEMON/downloads/download_EEG.html
HCP (Mental States) [102]	https://www.humanconnectome.org/study/
CHB-MIT [103]	https://physionet.org/content/chbmit/1.0.0/
Siena EEG Dataset [104]	https://physionet.org/content/siena-scalp-eeg/1.0.0/
ABIDE-I [105]	https://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html
BRACS	https://bracs.icar.cnr.it/
TCGA-NSCLC	https://portal.gdc.cancer.gov/projects/TCGA-LUAD
TCGA-RCC	https://portal.gdc.cancer.gov/projects/TCGA-KIRC
CAMELYON16	https://camelyon16.grand-challenge.org/
LVCN (Left Ventricle Cine-MRI)	https://competitions.codalab.org/competitions/17094
BTCV	https://www.synapse.org/!Synapse:syn3193805
Amazon Datasets	http://jmcauley.ucsd.edu/data/amazon/
MovieLens-1M	https://grouplens.org/datasets/movielens/1m/
People in Photo Albums (PIPA)	https://exposing.ai/pipa/
People in Social Context (PISC)	https://zenodo.org/records/1059155

Table A2: Mamba-specific taxonomy of representative Graph-Mamba models: graph construction, scanning/serialization, and gating/selection mechanisms.

Model Name	Primary Task	Graph Construction/Node Semantics	Core Graph-Mamba Integration Mechanism	Scanning/Serialization	Gating/Selection
STMAGRN [63] (2025)	Spatio-temporal Traffic Forecasting	Dynamic traffic sensor graph; nodes = road sensors	Mamba encoder + STM dynamic graph generation + GCRN decoder.	Uni-directional	Mamba selective SSM; GCRN gates
STGNN-M [64] (2024)	Cellular Traffic Prediction	Base-station correlation graph; nodes = cellular base stations	Bi-Mamba for temporal dynamics; GCN/GAT for nearby and distant base-station relations.	Bi-directional	Standard Mamba selection
MGCN [65] (2025)	Long-term Traffic Forecasting	Road sensor graph; nodes = road network sensors/detectors	Chebyshev GCN for spatial features + Bi-Mamba for long-term temporal modeling.	Bi-directional	SILU-gated Mamba output
Transfer-Mamba [66] (2025)	Few-shot Cross-city Traffic Prediction	Cross-city traffic graphs; nodes = road sensors/detectors	Mamba + adaptive GCN pretraining; knowledge clustering/querying for transfer.	Uni-directional	Standard selective SSM
SkeMamba [67] (2025)	Skeleton-based Action Recognition	Anatomical skeleton graph; nodes = human joints	ATT converts skeleton graphs to sequences; DSTMba models spatial/temporal paths.	Bi-directional tree + dual spatial/temporal paths	ST-GateMba
ActionMamba [68] (2025)	Skeleton-based Action Recognition	Skeleton graph; nodes = joints	ACE encoding + SM-GCN local aggregation + ST-Mamba global modeling.	Multi-axis spatial/temporal scans	MLP-SILU gating
MGDTA [80] (2024)	Drug-Target Affinity Prediction	Drug molecular graph; nodes = atoms,	Graph Transformer for drugs; Mamba for protein sequences; feature fusion.	Uni-directional	Standard selective scan
MambaCAtnGCN+ [81] (2025)	piRNA-Disease Association Prediction	Heterogeneous piRNA-disease graph; nodes = piRNAs and diseases	MambaTextCNN + GCN + cross-attention + PUL.	Uni-directional	Selective SSM + gated MLP
Brain-Mamba [82] (2024)	Brain Activity Modeling	Brain graph/time series; nodes = brain units, e.g. ROIs, channels, or voxels.	BTMamba for temporal signals; BNMamba for graph-based brain-region aggregation.	Functional ordering + bidirectional readout	Selective SSM + learnable fusion gate
AGIIBM [83] (2025)	Multi-view Brain Network Classification	Dynamic brain network; nodes = brain regions or EEG channels	Graph information bottleneck + GIN + Mamba temporal modeling.	Not specified	GIB sigmoid-Bernoulli graph masking + Mamba feature filtering
DG-SSM-MIL [84] (2025)	Whole-Slide Image Classification	WSI patch graph; nodes = image patches/instances	GAT local graph modeling + DynGraph-SSM/Bi-SSM global sequence modeling.	Bi-directional Bi-SSM	Attention fusion + selective SSM

(Continued)

Table A2 (continued)

Model Name	Primary Task	Graph Construction/Node Semantics	Core Graph-Mamba Integration Mechanism	Scanning/Serialization	Gating/Selection
GraphMamba [85] (2026)	Medical Image Segmentation	Medical image patch/region graph; nodes = image patches or anatomical regions	GTAS generates graph-guided scan orders; SCL enforces spatial consistency.	Adaptive topology-aware BFS scan	Selective scan + SCL
HGM-SRR [125] (2025)	Hybrid-type Social Relation Recognition	Person-person/object hypergraphs; vertices = persons or objects	PPH/POH + VSA filtering + VIO Mamba vertex interaction + MLT fusion.	Multi-route/multi-directional	VSA selection + selective SSM
CIM-SR [126] (2025)	Sequential Recommendation	Global item graph; nodes = items	Mamba sequence encoder + LightGCN graph encoder + contrastive alignment.	Uni-directional	Mamba adaptive selection gating
AdaptQA [127] (2025)	Natural Answer Generation	Multi-hop knowledge graph; nodes = entities	Adaptive KG retrieval + graph-based Mamba + LLM answer refinement.	Uni-directional	Mamba selective filtering
EMK-KEN [128] (2025)	Citation Network Classification	Citation ego-network; nodes = papers, central node = target paper	MamST encodes metadata/text with Mamba; KAN classifies central node.	Uni-directional	Selective propagation/forgetting

References

1. Wang Z, Zhu H, Liu P, Sun L. Social engineering in cybersecurity: a domain ontology and knowledge graph application examples. *Cybersecurity*. 2021;4(1):31. doi:10.1186/s42400-021-00094-6.
2. Sun AY, Jiang P, Mudunuru MK, Chen X. Explore spatio-temporal learning of large sample hydrology using graph neural networks. *Water Resour Res*. 2021;57(12):e2021WR030394. doi:10.1029/2021wr030394.
3. Lyu J, Xu G, Bao F, Zhou Y, Liu Y, Lu S, et al. Multi-label classification model using graph convolutional neural network for social network nodes. *Comput Model Eng Sci*. 2026;146(2):1–10. doi:10.32604/cmesci.2025.075239.
4. Wei Y, Wang X, He X, Nie L, Rui Y, Chua TS. Hierarchical user intent graph network for multimedia recommendation. *IEEE Trans Multimed*. 2022;24:2701–12. doi:10.1109/tmm.2021.3088307.
5. Zhang XM, Liang L, Liu L, Tang MJ. Graph neural networks and their current applications in bioinformatics. *Front Genet*. 2021;12:690049. doi:10.3389/fgene.2021.690049.
6. Corso G, Stark H, Jegelka S, Jaakkola T, Barzilay R. Graph neural networks. *Nat Rev Meth Primers*. 2024;4(1):17. doi:10.1038/s43586-024-00294-7.
7. Veličković P. Everything is connected: graph neural networks. *Curr Opin Struct Biol*. 2023;79:102538. doi:10.1016/j.sbi.2023.102538.
8. Zheng W, Xu G, Lu S, Lyu J, Bao F, Yin L. GNN: core branches, integration strategies and applications. *Comput Model Eng Sci*. 2026;146(1):1–10. doi:10.32604/cmesci.2025.075741.
9. Gao C, Zheng Y, Li N, Li Y, Qin Y, Piao J, et al. A survey of graph neural networks for recommender systems: challenges, methods, and directions. *ACM Trans Recomm Syst*. 2023;1(1):1–51. doi:10.1145/3568022.
10. Khemani B, Patil S, Kotecha K, Tanwar S. A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. *J Big Data*. 2024;11(1):18. doi:10.1186/s40537-023-00876-4.
11. Liu M, Wang Z, Ji S. Non-local graph neural networks. *IEEE Trans Pattern Anal Mach Intell*. 2022;44(12):10270–6. doi:10.1109/TPAMI.2021.3134200.
12. Luan S, Hua C, Lu Q, Zhu J, Zhao M, Zhang S, et al. Revisiting heterophily for graph neural networks. *Adv Neural Inf Process Syst*. 2022;35:1362–75. doi:10.5555/3600270.3600370.
13. Suresh S, Budde V, Neville J, Li P, Ma J. Breaking the limit of graph neural networks by improving the assortativity of graphs with local mixing patterns. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*; 2021 Aug 14–18; Online. p. 1541–51. doi:10.1145/3447548.3467373.
14. Liu S, Ying R, Dong H, Li L, Xu T, Rong Y, et al. Local augmentation for graph neural networks. *Int Conf Mach Learn*. 2022;162(6):14364–78. doi:10.1007/s13042-026-03101-9.
15. Kim H, Lee BS, Shin WY, Lim S. Graph anomaly detection with graph neural networks: current status and challenges. *IEEE Access*. 2022;10(9):111820–9. doi:10.1109/ACCESS.2022.3211306.
16. He X, Wang Y, Fan W, Shen X, Juan X, Miao R, et al. Mamba-based graph convolutional networks: tackling over-smoothing with selective state space. In: *Proceedings of the 34th International Joint Conference on Artificial Intelligence*; 2025 Aug 16–22; Montreal, QC, Canada. p. 5345–53. doi:10.24963/ijcai.2025/595.
17. Qian F, Bai L, Cui L, Li M, Du H, Wang Y, et al. Exploring the over-smoothing problem of graph neural networks for graph classification: an entropy-based viewpoint. In: *Proceedings of the 34th International Joint Conference on Artificial Intelligence*; 2025 Aug 16–22; Montreal, QC, Canada. p. 3235–44. doi:10.24963/ijcai.2025/360.
18. Akansha S. Over-squashing in graph neural networks: a comprehensive survey. *Neurocomputing*. 2025;642(3):130389. doi:10.1016/j.neucom.2025.130389.
19. Sun C, Li C, Lin X, Zheng T, Meng F, Rui X, et al. Attention-based graph neural networks: a survey. *Artif Intell Rev*. 2023;56(2):2263–310. doi:10.1007/s10462-023-10577-2.
20. Ying C, Cai T, Luo S, Zheng S, Ke G, He D, et al. Do transformers really perform badly for graph representation? *Adv Neural Inf Process Syst*. 2021;34:28877–88.
21. Özçelik R, de Ruiter S, Criscuolo E, Grisoni F. Chemical language modeling with structured state space sequence models. *Nat Commun*. 2024;15(1):6176. doi:10.1038/s41467-024-50469-9.
22. Alonso CA, Sieber J, Zeilinger MN. State space models as foundation models: a control theoretic overview. In: *Proceedings of the 2025 American Control Conference (ACC)*; 2025 Jul 8–10; Denver, CO, USA. p. 146–53.

23. Muca Cirone N, Orvieto A, Walker B, Salvi C, Lyons T. Theoretical foundations of deep selective state-space models. *Adv Neural Inf Process Syst*. 2024;37:127226–72. doi:10.52202/079017-4041.
24. Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces. In: *Proceedings of the First Conference on Language Modeling (CoLM)*; 2024 Oct 7–9; Philadelphia, PA, USA.
25. Zhang H, Zhu Y, Wang D, Zhang L, Chen T, Wang Z, et al. A survey on visual Mamba. *Appl Sci*. 2024;14(13):5683. doi:10.3390/app14135683.
26. Liu X, Zhang C, Huang F, Xia S, Wang G, Zhang L. Vision Mamba: a comprehensive survey and taxonomy. *IEEE Trans Neural Netw Learn Syst*. 2026;37(2):505–25. doi:10.1109/TNNLS.2025.3610435.
27. Salam A, Mahmud R, Islam T, Mukta S, Shatabda S. A comprehensive survey on Mamba: architectures, challenges, and opportunities. *Computer*. 2025;58(8):64–76. doi:10.1109/MC.2025.3571322.
28. Wang Z, Kong F, Feng S, Wang M, Yang X, Zhao H, et al. Is Mamba effective for time series forecasting? *Neurocomputing*. 2025;619(3):129178. doi:10.1016/j.neucom.2024.129178.
29. Jian B, Pan J, Ghahremani M, Rueckert D, Wachinger C, Wiestler B. Mamba? Catch the hype or rethink what really helps for Image registration. In: *Biomedical image registration*. Cham, Switzerland: Springer; 2024. p. 86–97. doi:10.1007/978-3-031-73480-9_7.
30. Wang C, Tsepa O, Ma J, Wang B. Graph-Mamba: towards long-range graph sequence modeling with selective state spaces. *arXiv:2402.00789*. 2024. doi:10.48550/arxiv.2402.00789.
31. Behrouz A, Hashemi F. Graph Mamba: towards learning on graphs with state space models. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*; 2024 Aug 25–29; Barcelona, Spain. p. 119–30. doi:10.1145/3637528.3672044.
32. Han D, Wang Z, Xia Z, Han Y, Pu Y, Ge C, et al. Demystify Mamba in vision: a linear attention perspective. *Adv Neural Inf Process Syst*. 2024;37:127181–203. doi:10.52202/079017-4039.
33. Sun J, Chen K, Wang S, Zhang Y, Xu Z, Wu X, et al. DGFE-Mamba: mamba-based 2D image segmentation network. *J Bionic Eng*. 2025;22(4):2135–50. doi:10.1007/s42235-025-00711-x.
34. Liu J, Yang H, Zhou HY, Xi Y, Yu L, Li C, et al. Swin-UMamba: Mamba-based UNet with ImageNet-based pretraining. In: *Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2024*; 2024 Oct 6–10; Marrakesh, Morocco. p. 615–25. doi:10.1007/978-3-031-72114-4_59.
35. Yin L, Wang L, Lu S, Xu G, Bao F, Zheng W. GMUG: a U-Net processed GRU optimized graph-Mamba network for wetland classification. *JAWRA J Am Water Resour Assoc*. 2026;62(3):e70132. doi:10.1111/1752-1688.70132.
36. Chu Y, Liu Y, Zhang Y, Hou X, Yu L, Peng Z. TESSM: tree-based selective state space models for efficient join order selection learning. In: *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*; 2024 Oct 21–25; Boise, ID, USA. p. 374–83. doi:10.1145/3627673.3679742.
37. Wang J, Lang J, Yang LZ, Li H. Learning dynamic brain network representation based on graph Mamba architecture. In: *Proceedings of the 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*; 2024 Dec 3–6; Lisbon, Portugal. p. 5143–50. doi:10.1109/BIBM62325.2024.10822717.
38. Hu J, Guo D, Si Z, Liu D, Diao Y, Zhang J, et al. MOL-Mamba: enhancing molecular representation with structural & electronic insights. *AAAI*. 2025;39(1):317–25. doi:10.1609/aaai.v39i1.32009.
39. Atitallah SB, Rabah CB, Driss M, Boulila W, Koubaa A. Exploring graph Mamba: a comprehensive survey on state-space models for graph learning. *arXiv:2412.18322*. 2024. doi:10.48550/arxiv.2412.18322.
40. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71.
41. Sarkis-Onofre R, Catalá-López F, Aromataris E, Lockwood C. How to properly use the PRISMA statement. *Syst Rev*. 2021;10(1):117. doi:10.1186/s13643-021-01671-z.
42. Durbin J, Koopman SJ. *Time series analysis by state space methods*. Oxford, UK: Oxford University Press; 2012.
43. Mews S, Koslik JO, Langrock R. How to build your latent Markov model: the role of time and space. *Stat Model*. 2025;25(6):481–507. doi:10.1177/1471082x251355681.
44. Gu A, Goel K, Ré C. Efficiently modeling long sequences with structured state spaces. *arXiv:2111.00396*. 2021.
45. Yao Z, Zheng Q, Zuo J, Zhang Y. GPS-Mamba: graph permutation scanning state space model for multivariate time series forecasting. *Expert Syst Appl*. 2026;311(12):131373. doi:10.1016/j.eswa.2026.131373.

46. Zhang L, Huang S, Ma D, Zhang D, Zhang L. Brain network Mamba: a bi-directional state-space model for brain network analysis on rs-fMRI. In: Proceedings of the International Workshop on Machine Learning in Medical Imaging; 2025 Sep 23; Daejeon, Republic of Korea. p. 224–33. doi:10.1007/978-3-032-09513-8_22.
47. Liu Z, Wang X, Wang B, Huang Z, Yang C, Jin W. Graph odes and beyond: a comprehensive survey on integrating differential equations with graph neural networks. In: Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2; 2025 Aug 3–7; Toronto, ON, Canada. p. 6118–28. doi:10.1145/3711896.3736559.
48. Liu Y, Tian Y, Zhao Y, Yu H, Xie L, Wang Y, et al. VMamba: visual state space model. *Adv Neural Inf Process Syst*. 2024;37:103031–63.
49. Pandey AP, Varghese AJ, Patil S, Xu M. A comparative study on dynamic graph embedding based on Mamba and transformers. *arXiv:2412.11293*. 2024. doi:10.48550/arxiv.2412.11293.
50. Adoni HWY, Nahhal T, Krichen M, Aghezzaf B, Elbyed A. A survey of current challenges in partitioning and processing of graph-structured data in parallel and distributed systems. *Distrib Parallel Databases*. 2020;38(2):495–530. doi:10.1007/s10619-019-07276-9.
51. Zhang S, Tong H, Xu J, Maciejewski R. Graph convolutional networks: a comprehensive review. *Comput Soc Netw*. 2019;6(1):11. doi:10.1186/s40649-019-0069-y.
52. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. *Int Conf Mach Learn*. 2017;7(6):1263–72. doi:10.1007/978-3-030-40245-7_10.
53. Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, et al. T-GCN: a temporal graph convolutional network for traffic prediction. *IEEE Trans Intell Transport Syst*. 2020;21(9):3848–58. doi:10.1109/tits.2019.2935152.
54. Liu J, Ong GP, Chen X. GraphSAGE-based traffic speed forecasting for segment network with sparse data. *IEEE Trans Intell Transp Syst*. 2022;23(3):1755–66. doi:10.1109/tits.2020.3026025.
55. Vrahatis AG, Lazaros K, Kotsiantis S. Graph attention networks: a comprehensive review of methods and applications. *Future Internet*. 2024;16(9):318. doi:10.3390/fi16090318.
56. Huang K, Zhai J, Zheng Z, Yi Y, Shen X. Understanding and bridging the gaps in current GNN performance optimizations. In: Proceedings of the 26th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming; 2021 Feb 27; New York, NY, USA. p. 119–32. doi:10.1145/3437801.3441585.
57. Shi D, Lin L, Han A, Wang Z, Guo Y, Gao J. When graph neural networks meet dynamic mode decomposition. *Int Conf Learn Represent*. 2025;2025:19691–719.
58. Chen RT, Rubanova Y, Bettencourt J, Duvenaud D. Neural ordinary differential equations. In: Proceedings of the Proceedings of the 32nd International Conference on Neural Information Processing Systems; 2018 Dec 3–8; Red Hook, NY, USA. p. 6572–83.
59. Asikis T, Böttcher L, Antulov-Fantulin N. Neural ordinary differential equation control of dynamics on graphs. *Phys Rev Research*. 2022;4(1):013221. doi:10.1103/physrevresearch.4.013221.
60. Dong H, Lee GH. PS-Mamba: spatial-temporal graph Mamba for pose sequence refinement. In: Proceedings of the 2025 IEEE/CVF International Conference on Computer Vision (ICCV); 2025 Oct 19–25; Honolulu, HI, USA. p. 8568–78. doi:10.1109/ICCV51701.2025.00802.
61. Cheng L, Zhu P, Guo Y, Gao C, Wang Z, Tang K. SourceDetMamba: a graph-aware state space model for source detection in sequential hypergraphs. In: Proceedings of the 34th International Joint Conference on Artificial Intelligence; 2025 Aug 16–22; Montreal, QC, Canada. p. 2749–57. doi:10.24963/ijcai.2025/306.
62. Ren Y, Meng Z, Chen K, Sun W, Chen H. Spatio-temporal hypergraph-driven evolutionary Graph-Mamba method for remaining useful life prediction. *Adv Eng Inform*. 2026;69(8):103925. doi:10.1016/j.aei.2025.103925.
63. Zhang X, Zhang Y, Meng X. Spatio-temporal Mamba dynamic graph convolutional recurrent network for traffic prediction. *IEEE Trans Artif Intell*. 2026;7(1):225–38. doi:10.1109/TAI.2025.3571378.
64. Li J, Pu X, Xia P. STGNNM: spatial-temporal graph neural network with Mamba for cellular traffic prediction. In: Proceedings of the 2024 16th International Conference on Wireless Communications and Signal Processing (WCSP); 2024 Oct 24–26; Hefei, China. p. 1187–92. doi:10.1109/WCSP62071.2024.10827036.
65. Lin W, Zhang Z, Ren G, Zhao Y, Ma J, Cao Q. MGCN: Mamba-integrated spatiotemporal graph convolutional network for long-term traffic forecasting. *Knowl Based Syst*. 2025;309(4):112875. doi:10.1016/j.knosys.2024.112875.

66. Cheng S, Qu S, Zhang J. Transfer-Mamba: selective state space models with spatio-temporal knowledge transfer for few-shot traffic prediction across cities. *Simul Model Pract Theory*. 2025;140(6):103066. doi:10.1016/j.simpat.2025.103066.
67. Zhao J, Dai J, Zhou F, Pan J, Xu H. Dual-path spatio-temporal Mamba for skeleton-based action recognition. *Visual Comput*. 2025;41(9):6507–19. doi:10.1007/s00371-025-03950-5.
68. Wen J, Liu D, Zheng B. ActionMamba: action spatial-temporal aggregation network based on Mamba and GCN for skeleton-based action recognition. *Electronics*. 2025;14(18):3610. doi:10.3390/electronics14183610.
69. Li Y, Yu R, Shahabi C, Liu Y. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. *arXiv:1707.01926*. 2017. doi:10.48550/arxiv.1707.01926.
70. Song C, Lin Y, Guo S, Wan H. Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting. *AAAI*. 2020;34(1):914–21. doi:10.1609/aaai.v34i01.5438.
71. Barlacchi G, De Nadai M, Larcher R, Casella A, Chitic C, Torrissi G, et al. A multi-source dataset of urban life in the city of Milan and the Province of Trentino. *Sci Data*. 2015;2(1):150055. doi:10.1038/sdata.2015.55.
72. Sedmidubsky J, Elias P, Zezula P. Benchmarking search and annotation in continuous human skeleton sequences. In: *Proceedings of the 2019 on International Conference on Multimedia Retrieval*; 2019 Jun 10–13; Ottawa, ON, Canada. p. 38–42. doi:10.1145/3323873.3325013.
73. Li T, Liu J, Zhang W, Ni Y, Wang W, Li Z. UAV-human: a large benchmark for human behavior understanding with unmanned aerial vehicles. In: *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2021 Jun 20–25; Nashville, TN, USA. p. 16266–75.
74. Wang J, Nie X, Xia Y, Wu Y, Zhu SC. Cross-view action modeling, learning, and recognition. In: *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*; 2014 Jun 23–28; Columbus, OH, USA. p. 2649–56. doi:10.1109/cvpr.2014.339.
75. Zheng T, Yao H, Zhao S, Jiang K, Xiao Y. GraphMamba: whole slide image classification meets graph-driven selective state space model. *Pattern Recognit*. 2025;167(1):111768. doi:10.1016/j.patcog.2025.111768.
76. Khan S, Dambandkhameneh F, Shaikh N, Nie Y, Venugopal R, Li X. SlideMamba: entropy-based adaptive fusion of GNN and Mamba for enhanced representation learning in digital pathology. *Sci Rep*. 2026;16(1):4253. doi:10.1038/s41598-025-34367-8.
77. Wang C, Xie Y, Chen Q, Zhou Y, Wu Q. A comprehensive analysis of Mamba for 3D volumetric medical image segmentation. *Pattern Recognit*. 2026;173(3):112701. doi:10.1016/j.patcog.2025.112701.
78. Li H, Fan J, Cai W. VGM-UNet: a hybrid visual graph deformable Mamba with Fourier neural operator U-Net for medical image segmentation. *Neural Netw*. 2026;201:108890. doi:10.1016/j.neunet.2026.108890.
79. Zhou Y, Sun L, Xiong X, Ti G, Yang S. GCNet-Mamba: leveraging state space models and CNN for medical image classification. *Expert Syst Appl*. 2026;303(8):130733. doi:10.1016/j.eswa.2025.130733.
80. Han K, Shi C, Wang Z, Liu W, Li Z, Wang Z, et al. Innovative Mamba and graph transformer framework for superior protein-ligand affinity prediction. *Microchem J*. 2024;206:111444. doi:10.1016/j.microc.2024.111444.
81. Yao D, Li X, Zhan X, Zhang B, Zhang J. MambaCAtnGCN+: a comprehensive framework integrating MambaTextCNN, cross-attention and graph convolution network for piRNA-disease association prediction. *Sci Rep*. 2025;15(1):25058. doi:10.1038/s41598-025-07641-y.
82. Behrouz A, Hashemi F. Brain-Mamba: encoding brain activity via selective state space models. In: *Proceedings of the 15th Conference on Health, Inference, and Learning*; 2024 Jun 27–28; New York, NY, USA. p. 233–50.
83. Dong C, Sun D, Yu Z, Luo B. Multi-view brain network classification based on adaptive graph isomorphic information bottleneck Mamba. *Expert Syst Appl*. 2025;267(1):126170. doi:10.1016/j.eswa.2024.126170.
84. Ding F, Cai C, Li J, Liu M, Jiao Y, Wu Z, et al. Classification of whole-slide pathology images based on state space models and graph neural networks. *Electronics*. 2025;14(10):2056. doi:10.3390/electronics14102056.
85. Yu C, Zhang H, Pu C, Lv S, Yu J, Wu X, et al. GraphMamba: graph-driven spatial order-aware Mamba for medical image segmentation. *Pattern Recognit*. 2026;171(3):112231. doi:10.1016/j.patcog.2025.112231.
86. Shui Z, Karypis G. Heterogeneous molecular graph neural networks for predicting molecule properties. In: *Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM)*; 2020 Nov 17–20; Sorrento, Italy. p. 492–500. doi:10.1109/icdm50108.2020.00058.

87. Zang X, Zhao X, Tang B. Hierarchical molecular graph self-supervised learning for property prediction. *Commun Chem*. 2023;6(1):34. doi:10.1038/s42004-023-00825-5.
88. Li MM, Huang K, Zitnik M. Graph representation learning in biomedicine and healthcare. *Nat Biomed Eng*. 2022;6(12):1353–69. doi:10.1038/s41551-022-00942-x.
89. Yi HC, You ZH, Huang DS, Kwok CK. Graph representation learning in bioinformatics: trends, methods and applications. *Brief Bioinform*. 2022;23(1):bbab340. doi:10.1093/bib/bbab340.
90. Mswahili ME, Jeong YS. Transformer-based models for chemical SMILES representation: a comprehensive literature review. *Heliyon*. 2024;10(20):e39038. doi:10.1016/j.heliyon.2024.e39038.
91. Wu CK, Zhang XC, Yang ZJ, Lu AP, Hou TJ, Cao DS. Learning to SMILES: BAN-based strategies to improve latent representation learning from molecules. *Brief Bioinform*. 2021;22(6):bbab327. doi:10.1093/bib/bbab327.
92. Wu T, Tang Y, Sun Q, Xiong L. Molecular joint representation learning via multi-modal information of SMILES and graphs. *IEEE/ACM Trans Comput Biol Bioinform*. 2023;20(5):3044–55. doi:10.1109/TCBB.2023.3253862.
93. Chen J, Lin J, Hu Y, Ye M, Yao L, Wu L, et al. RNADisease v4.0: an updated resource of RNA-associated diseases, providing RNA-disease analysis, enrichment and prediction. *Nucleic Acids Res*. 2023;51(D1):D1397–404. doi:10.1093/nar/gkac814.
94. Muhammad A, Waheed R, Ali Khan N, Jiang H, Song X. piRDisease v1.0: a manually curated database for PiRNA associated diseases. *Database*. 2019;2019:baz052. doi:10.1093/database/baz052.
95. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol*. 2011;29(11):1046–51. doi:10.1038/nbt.1990.
96. Tang J, Szwarzda A, Shakyawar S, Xu T, Hintsanen P, Wennerberg K, et al. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *J Chem Inf Model*. 2014;54(3):735–43. doi:10.1021/ci400709d.
97. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res*. 2007;35(suppl 1):D198–201. doi:10.1093/nar/gkl999.
98. Behrouz A, Delavari P, Hashemi F. Unsupervised representation learning of brain activity via bridging voxel activity and functional connectivity. In: *Proceedings of the 41st International Conference on Machine Learning*; 2024 Jul 21–27; Vienna, Austria. p. 3347–81.
99. Bellec P, Chu C, Chouinard-Decorte F, Benhajali Y, Margulies DS, Craddock RC. The neuro bureau ADHD-200 preprocessed repository. *NeuroImage*. 2017;144:275–86. doi:10.1016/j.neuroimage.2016.06.034.
100. Shah V, von Weltin E, Lopez S, McHugh JR, Veloso L, Golmohammadi M, et al. The temple university hospital seizure detection corpus. *Front Neuroinform*. 2018;12:83. doi:10.3389/fninf.2018.00083.
101. Babayan A, Erbey M, Kumral D, Reinelt JD, Reiter AMF, Röbbig J, et al. A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults. *Sci Data*. 2019;6(1):180308. doi:10.1038/sdata.2018.308.
102. Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K. The WU-minn human connectome project: an overview. *NeuroImage*. 2013;80(156869):62–79. doi:10.1016/j.neuroimage.2013.05.041.
103. Shoeb AH. Application of machine learning to epileptic seizure onset detection and treatment [dissertation]. Cambridge, MA, USA: Massachusetts Institute of Technology; 2009.
104. Detti P, Vatti G, Zabalo Manrique de Lara G. EEG synchronization analysis for seizure prediction: a study on data of noninvasive recordings. *Processes*. 2020;8(7):846. doi:10.3390/pr8070846.
105. di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol Psychiatry*. 2014;19(6):659–67. doi:10.1038/mp.2013.78.
106. Cameron C, Yassine B, Carlton C, Francois C, Alan E, Andrés J, et al. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Front Neuroinform*. 2013;7:41. doi:10.3389/conf.fninf.2013.09.00041.
107. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*. 2002;15(1):273–89. doi:10.1006/nimg.2001.0978.

108. Wang M, Tong G. RLMamba: integrating residual learning with Mamba for long-term time series forecasting. *Expert Syst Appl.* 2025;278(2):127362. doi:10.1016/j.eswa.2025.127362.
109. Yang S, Wang Y, Chen H. Mambamil: enhancing long sequence modeling with sequence reordering in computational pathology. In: *Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2024*; 2024 Oct 6–10; Marrakesh, Morocco. p. 296–306. doi:10.1007/978-3-031-72083-3_28.
110. Yu S, Li A, Chen Y, Wang D, Tang X. Heterogeneous network-based algorithms in the biomedical data mining: a review from technical perspective. *Inform Health.* 2024;1(2):111–22. doi:10.1016/j.infoh.2024.07.002.
111. Yue L, Tian D, Chen W, Han X, Yin M. Deep learning for heterogeneous medical data analysis. *World Wide Web.* 2020;23(5):2715–37. doi:10.1007/s11280-019-00764-z.
112. Rajendran S, Pan W, Sabuncu MR, Chen Y, Zhou J, Wang F. Learning across diverse biomedical data modalities and cohorts: challenges and opportunities for innovation. *Patterns.* 2024;5(2):100913. doi:10.1016/j.patter.2023.100913.
113. Chen C, Ma W, Zhang M, Wang Z, He X, Wang C, et al. Graph heterogeneous multi-relational recommendation. *AAAI.* 2021;35(5):3958–66. doi:10.1609/aaai.v35i5.16515.
114. Li Z, Zhao Y, Zhang Y, Zhang Z. Multi-relational graph attention networks for knowledge graph completion. *Knowl Based Syst.* 2022;251(11):109262. doi:10.1016/j.knosys.2022.109262.
115. Zhao Y, Dong J, Wang W, Duan H. A multi-typed multi-relational heterogeneous graph neural network model for complex networks. *Knowl Based Syst.* 2025;329(20):114291. doi:10.1016/j.knosys.2025.114291.
116. Meng F, Zhao S, Guo Z, Lan Y, Pang B. Mamba-GTC: cross-view contrastive learning with state space modeling for heterogeneous graph representation. *Knowl Based Syst.* 2026;335(12):115161. doi:10.1016/j.knosys.2025.115161.
117. Yang Y, Liu X. GMDNet: graph-Mamba dual-branch network for hyperspectral image classification. *Int Conf Artif Intell Pattern Recognit.* 2025;13993:526–35. doi:10.1117/12.3092534.
118. Dong H, Chharia A, Gou W, Carrasco FV, De la Torre F. Hamba: single-view 3D hand reconstruction with graph-guided bi-scanning Mamba. *arXiv:2407.09646.* 2024. doi:10.48550/arxiv.2407.09646.
119. Yuan H, Sun Q, Wang Z, Fu X, Ji C, Wang Y, et al. DG-Mamba: robust and efficient dynamic graph structure learning with selective state space models. *AAAI.* 2025;39(21):22272–80. doi:10.1609/aaai.v39i21.34382.
120. Yang A, Li M, Ding Y, Fang L, Cai Y, He Y. GraphMamba: an efficient graph structure learning vision Mamba for hyperspectral image classification. *IEEE Trans Geosci Remote Sensing.* 2024;62(86):1–14. doi:10.1109/tgrs.2024.3493101.
121. Ding R, Luong KD, Rodriguez E, da Silva ACAL, Hsu W. Combining graph neural network and Mamba to capture local and global tissue spatial relationships in whole slide images. *Sci Rep.* 2025;15(1):18261. doi:10.1038/s41598-025-99042-4.
122. Yin L, Wang L, Lu S, Xu G, Lyu J, Zheng W. GMVG: a ViT embedded graph-Mamba network for HSI/LiDAR land cover classification. *Glob NEST J.* 2026;28(6):1–11. doi:10.30955/gnj.08129.
123. Zhang L, Zhou Y, Tian J, Guo F, Zhang X, Yang B. HTM: a hybrid triangular modeling framework for soft tissue feature tracking. *Comput Model Eng Sci.* 2025;145(3):3949–68. doi:10.32604/cmes.2025.071869.
124. Bing R, Yuan G, Zhu M, Meng F, Ma H, Qiao S. Heterogeneous graph neural networks analysis: a survey of techniques, evaluations and applications. *Artif Intell Rev.* 2023;56(8):8003–42. doi:10.1007/s10462-022-10375-2.
125. Tang W, Qing L, Wang P, Li L, Zhu C. Hypergraph Mamba reasoning-based social relation recognition. *IEEE Trans Image Process.* 2025;34:4814–29. doi:10.1109/tip.2025.3592551.
126. Zhang Z, Yang B, Yin S. A collaborative information enhanced Mamba-based sequential recommendation model. In: *Proceedings of the 2025 10th International Conference on Computer and Communication System (ICCCS)*; 2025 Apr 18–21; Chengdu, China. p. 167–72. doi:10.1109/icccs65393.2025.11069788.
127. Liu Z, Li J, Huang Y, Cui N, Pei L. Knowledge-based natural answer generation via effective graph learning. *Knowl Based Syst.* 2025;316(5–6):113288. doi:10.1016/j.knosys.2025.113288.
128. Liu C, Qu Z. EMK-KEN: a high-performance approach for assessing knowledge value in citation network. *Knowl Based Syst.* 2025;326(46):113949. doi:10.1016/j.knosys.2025.113949.

129. Rosso P, Yang D, Cudré-Mauroux P. Beyond triplets: hyper-relational knowledge graph embedding for link prediction. In: Proceedings of the Web Conference 2020; 2020 Apr 20–24; Taipei, Taiwan. p. 1885–96. doi:10.1145/3366423.3380257.
130. Li J, Wong Y, Zhao Q, Kankanhalli MS. Dual-glance model for deciphering social relationships. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22–29; Venice, Italy. p. 2669–78. doi:10.1109/iccv.2017.289.
131. Zhang N, Paluri M, Taigman Y, Fergus R, Bourdev L. Beyond frontal faces: improving person recognition using multiple cues. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015 Jun 7–12; Boston, MA, USA. p. 4804–13. doi:10.1109/CVPR.2015.7299113.
132. Qin X, Yuan H, Zhao P, Liu G, Zhuang F, Sheng VS. Intent contrastive learning with cross subsequences for sequential recommendation. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining; 2024 Mar 4–8; Merida, Mexico. p. 548–56. doi:10.1145/3616855.3635773.
133. Hogan A, Blomqvist E, Cochez M, D'amato C, De Melo G, Gutierrez C, et al. Knowledge graphs. *ACM Comput Surv.* 2021;54(4):1–37. doi:10.1145/3447772.
134. Jagvaral B, Lee WK, Roh JS, Kim MS, Park YT. Path-based reasoning approach for knowledge graph completion using CNN-BiLSTM with attention mechanism. *Expert Syst Appl.* 2020;142(1):112960. doi:10.1016/j.eswa.2019.112960.
135. Peng C, Xia F, Naseriparsa M, Osborne F. Knowledge graphs: opportunities and challenges. *Artif Intell Rev.* 2023;56(11):13071–102. doi:10.1007/s10462-023-10465-9.
136. Ji S, Pan S, Cambria E, Marttinen P, Yu PS. A survey on knowledge graphs: representation, acquisition, and applications. *IEEE Trans Neural Netw Learn Syst.* 2022;33(2):494–514. doi:10.1109/TNNLS.2021.3070843.
137. Yang Z, Qi P, Zhang S, Bengio Y, Cohen W, Salakhutdinov R, et al. HotpotQA: a dataset for diverse, explainable multi-hop question answering. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing; 2018 Oct 31–Nov 4; Brussels, Belgium. p. 2369–80. doi:10.18653/v1/d18-1259.
138. Welbl J, Stenetorp P, Riedel S. Constructing datasets for multi-hop reading comprehension across documents. *Trans Assoc Comput Linguist.* 2018;6(1):287–302. doi:10.1162/tacl_a_00021.
139. Anderson MWR, Millard D. Hypertext's meta-history: documenting in-conference citations, authors and keyword data, 1987–2021. In: Proceedings of the 33rd ACM Conference on Hypertext and Social Media; 2022 Jun 28–Jul 1; Barcelona, Spain. p. 96–106. doi:10.1145/3511095.3531271.
140. Leng G, Leng RI, Ludwig M. Oxytocin—a social peptide? Deconstructing the evidence. *Philos Trans R Soc B.* 2022;377(1858):20210055. doi:10.1098/rstb.2021.0055.
141. Luo P. ACM IS abstract and citation network1. *Harv Dataverse.* 2014. doi:10.7910/DVN/27695.
142. Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: densification laws, shrinking diameters and possible explanations. In: Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining; 2005 Aug 21–24; Chicago, IL, USA. p. 177–87. doi:10.1145/1081870.1081893.
143. Cohan A, Feldman S, Beltagy I, Downey D, Weld D. SPECTER: document-level representation learning using citation-informed transformers. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; 2020 Jul 5–10; Online. p. 2270–82. doi:10.18653/v1/2020.acl-main.207.
144. Hu W, Fey M, Zitnik M, Dong Y, Ren H, Liu B, et al. Open graph benchmark: datasets for machine learning on graphs. *Adv Neural Inf Process Syst.* 2020;33:22118–33.
145. Tang J, Zhang J, Yao L, Li J, Zhang L, Su Z. ArnetMiner: extraction and mining of academic social networks. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2008 Aug 24–27; Las Vegas, NV, USA. p. 990–8. doi:10.1145/1401890.1402008.
146. Somvanshi S, Javed SA, Islam MM, Pandit D, Das S. A survey on Kolmogorov-Arnold network. *ACM Comput Surv.* 2026;58(2):1–35. doi:10.1145/3743128.