



Review

Graph Attention Networks: A Comprehensive Review of Methods and Applications

Aristidis G. Vrahatis ¹, Konstantinos Lazaros ¹ and Sotiris Kotsiantis ^{2,*}

¹ Department of Informatics, Ionian University, 49100 Corfu, Greece; aris.vrahatis@ionio.gr (A.G.V.); konlazaros@gmail.com (K.L.)

² Department of Mathematics, University of Patras, 49100 Patras, Greece

* Correspondence: kotsiantis@upatras.gr

Abstract: Real-world problems often exhibit complex relationships and dependencies, which can be effectively captured by graph learning systems. Graph attention networks (GATs) have emerged as a powerful and versatile framework in this direction, inspiring numerous extensions and applications in several areas. In this review, we present a thorough examination of GATs, covering both diverse approaches and a wide range of applications. We examine the principal GAT-based categories, including Global Attention Networks, Multi-Layer Architectures, graph-embedding techniques, Spatial Approaches, and Variational Models. Furthermore, we delve into the diverse applications of GATs in various systems such as recommendation systems, image analysis, medical domain, sentiment analysis, and anomaly detection. This review seeks to act as a navigational reference for researchers and practitioners aiming to emphasize the capabilities and prospects of GATs.

Keywords: graph attention networks; graph neural networks; graph convolution networks



Citation: Vrahatis, A.G.; Lazaros, K.; Kotsiantis, S. Graph Attention Networks: A Comprehensive Review of Methods and Applications. *Future Internet* **2024**, *16*, 318. <https://doi.org/10.3390/fi16090318>

Academic Editors: Panagiotis Papageorgas, Dimitrios Piromalis and Dionisis Kandris

Received: 25 July 2024

Revised: 14 August 2024

Accepted: 30 August 2024

Published: 3 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the last decade, the graphical representation of data has gained widespread prominence in multiple fields, ranging from social networks to molecular biology. Complex systems are everywhere, from societal structures demanding collaboration between billions of individuals to the harmonious functioning of billions of neurons in our brains. As a result, the graphical representation of data has emerged as a pivotal technique for capturing the complexity of these systems, allowing us to represent them as networks of entities and their interactions [1]. By depicting a complex system as a network comprising entities and their interactions, we can scrutinize their relationships, enabling us to acquire a more profound insight into the foundational structures and patterns that govern them [2].

Graph neural networks (GNNs), designed specifically to handle graph-structured data, play a crucial role in unlocking the full potential of such representations. In the last decade, a plethora of graph neural network (GNN) subcategories have been proposed to address the unique challenges of learning on graph-structured data. These subcategories include the graph convolutional networks (GCNs), which leverage convolutional operations on graph data to capture local neighborhood information; GraphSAGE, an inductive learning framework that generates embeddings for nodes by sampling and aggregating information from their local neighborhoods; ChebNets, which utilize Chebyshev polynomial spectral filters for efficient graph convolutions in the spectral domain; and Graph Isomorphism Networks (GINs), designed to capture the structural information of a graph by considering both node features and graph topology through a learnable aggregation function. These diverse architectures reflect the ongoing research efforts to enhance the performance and versatility of GNNs across various applications.

Graph attention networks (GATs) are a promising subcategory of graph neural networks (GNNs). They introduce the attention mechanism to GNNs, allowing them to dynamically weigh the importance of neighboring nodes during the aggregation process,

which helps capture complex relationships and dependencies in graph-structured data. This adaptability enables GATs to learn local patterns while preserving global graph structure information, enhancing their performance in various tasks across multiple domains.

In this review, we present a concise yet thorough exploration of graph attention networks (GATs), a key area in graph-based deep learning. Our paper is organized to guide readers through the core concepts and recent advancements in this field. We start with the basics of graph convolution networks (GCNs) in Section 2, where we also introduce graph attention networks and their advanced form, GATv2. This section aims to clarify the fundamental principles and functionalities of these networks.

Following this, Section 3 delves into the main categories of GATs, highlighting their flexibility and effectiveness in various settings. In categorizing the various types of graph attention networks (GATs) for this review, a deliberate and strategic approach was taken to ensure that the categories reflect both the distinct mechanisms employed by these networks and their specific applications in diverse domains. The primary objective was to offer a comprehensive yet organized framework that facilitates a deeper understanding of how different GAT architectures are optimized for particular challenges within graph-structured data. The categories were carefully chosen to highlight the unique contributions of each GAT variant, emphasizing the specific mechanisms they utilize, such as global attention, multi-layer stacking, or spatial considerations, which directly influence their effectiveness in capturing complex patterns and relationships within graphs. This categorization is essential for providing clarity in a rapidly evolving field, where the proliferation of GAT models can otherwise lead to confusion or overlap in understanding their capabilities and limitations.

The selection of these specific categories—Global Attention GATs, Multi-Layer GATs, graph-embedding GATs, spatial GATs, Variational GATs, and Hybrid GATs—was driven by the need to encapsulate the broad spectrum of techniques and methodologies that have emerged within the GAT framework. Each category represents a distinct approach to enhancing the core graph attention mechanism, tailored to address specific challenges such as the need for global context, multi-layer feature abstraction, or spatial awareness. For instance, Global Attention GATs were categorized separately to underscore their ability to capture overarching patterns across an entire graph, which is crucial for tasks where understanding distant node relationships is vital. Similarly, Multi-Layer GATs were distinguished for their capacity to aggregate and refine information through multiple layers, facilitating the learning of higher-order features. By organizing the review into these targeted categories, we aim to provide readers with a clear, structured, and detailed roadmap of the various GAT techniques, enabling them to more effectively select and apply the appropriate GAT model for their specific research needs or application domains. Then, in Section 4, we showcase the practical applications of GATs in different domains, illustrating their real-world impact. The paper concludes with a “Discussion” Section 5 that critically examines the current challenges and future directions for GAT research, offering a forward-looking perspective.

This review stands out for its clear and accessible language, making complex concepts understandable. Figure 1 below visually summarizes the key areas and applications of GAT-based tools covered in our paper. Our aim is to provide a valuable resource for both newcomers and seasoned researchers in the field, contributing a fresh and comprehensive view of the dynamic and evolving world of graph attention networks.

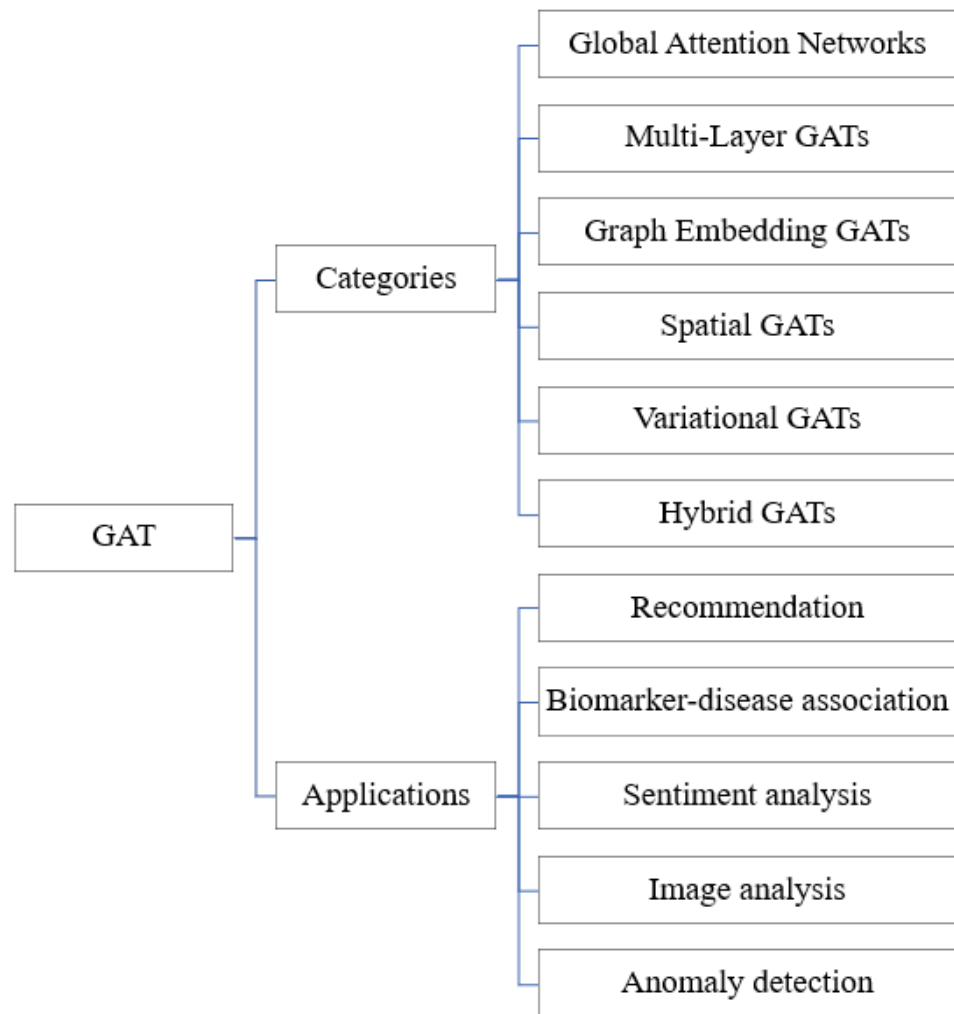


Figure 1. The applications and categories of GAT-based tools analyzed in this work.

2. Graph Neural Networks

Graph neural networks (GNNs) are a class of deep learning models that have been developed specifically to process data represented in the form of graphs or networks. Graphs are powerful data structures that encode relationships between objects, and GNNs enable one to leverage these relationships to extract meaningful features to perform tasks such as node classification, edge prediction, and graph-level inference. GNNs aim to learn embeddings that capture both the structural information of the graph and the features of the nodes as can be seen in Figure 2 below. They achieve this by passing messages between nodes in a graph, aggregating information from neighboring nodes and updating each node’s representation. These message-passing and updating steps are repeated over multiple layers, allowing the network to capture increasingly complex relationships between nodes. The final node embeddings can then be used for downstream tasks [2]. In a simple GNN, the vector representation h_A for a node A with neighbors N_A is computed as follows:

$$h_A = \sum_{i \in N_A} x_i W^T$$

where W is the weight matrix of the neural network and x_i is the input vector of the neighbors of A . Since we are talking about neural networks, such operations are converted to matrix multiplications for convenience and higher efficiency. Therefore, the above relation in matrix multiplication form is as follows:

$$H = \tilde{A}^T XW^T$$

where H is the vector representation matrix for all nodes, $\tilde{A} = A + I$ (where I is the unit matrix such that self-loops are included) is the adjacency matrix of the input graph, and X is the matrix of nodes.

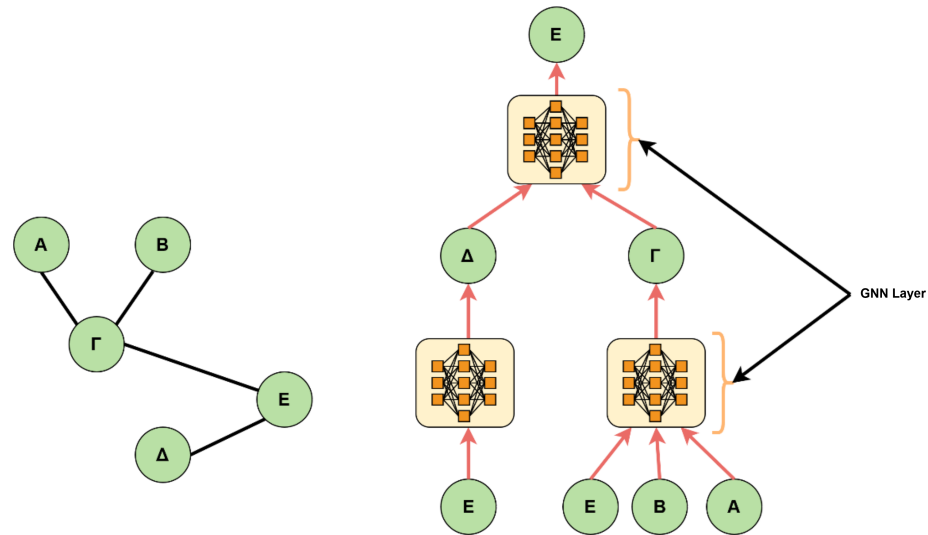


Figure 2. Illustration of the input graph (left) and the corresponding computation graph (right) depicting the process by which a graph neural network (GNN) computes the vector representation of node E by aggregating information from its neighboring nodes.

2.1. Graph Convolution Networks

The graph convolutional network (GCN) architecture serves as a fundamental paradigm for graph neural networks (GNNs), which was introduced in [3]. The primary objective of this architecture is to develop a computationally efficient version of convolutional neural networks (CNNs) for graph-based data. Specifically, it aims to approximate the graph convolution operation in graph signal processing. The GCN has emerged as a widely used and flexible architecture in various scientific domains, and it is frequently employed as a benchmark for graph data analysis. In comparison to tabular or image data, graph data exhibit varying numbers of neighbors for each node, rendering traditional graph neural networks inadequate. This discrepancy in neighborhood size poses a significant challenge that must be addressed. One solution involves dividing node embeddings by their respective degrees, i.e., the number of edges incident to each node. This process, known as degree normalization, ensures a fair comparison of nodes despite their differing numbers of neighbors. Degree normalization is accomplished via matrix multiplication, whereby each node embedding in the graph is multiplied by the degree matrix raised to the power of $-1/2$ as can be seen in Figure 3. The resulting normalized graph representation accounts for the variability in neighborhood size and enables effective analysis. The mathematical relation for computing vector representations for the nodes of a graph is now as follows:

$$H = \tilde{D}^{-1/2} \tilde{A}^T \tilde{D}^{-1/2} XW^T$$

where $\tilde{D} = D + I$ with D being the degree matrix for each node.

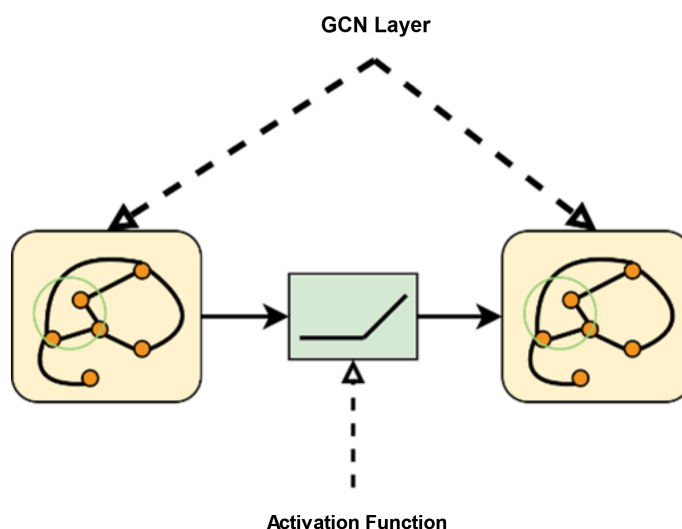


Figure 3. Visualization of graph convolutional network (GCN) layers. Unlike classical graph neural networks (GNNs), GCNs incorporate the degree of each node to enable effective normalization during the aggregation process.

2.2. Graph Attention Networks

Graph attention networks (GATs) [2,4] mark a significant theoretical progression from graph convolutional networks (GCNs). At the heart of GATs is the principle that certain nodes are more crucial than others. This idea, while not entirely novel and somewhat reflected in GCNs, is advanced in GATs. In GCNs, nodes with fewer neighbors gain more importance due to a normalization coefficient that depends primarily on node degrees. However, the limitation of this GCN approach is its sole dependence on node degrees for determining node importance.

In contrast, GATs aim to create weighting factors that consider not just the node degrees but also the significance of node features. This is where GATs diverge from GCNs: their method of assigning scale factors during the aggregation of neighborhood information. GCNs use a non-parametric scaling factor derived from a normalization function, whereas GATs employ an attention mechanism to allocate scaling factors. This attention-based approach allows GATs to assign greater weights to more important nodes during neighborhood aggregation. This key difference grants GATs a finer degree of control over how information flows within intricate graph structures, making them more adaptable for handling complex data. On the other hand, GCNs are generally more effective in scenarios where the graph's structure is clearly defined, and the significance of each node is more or less uniform.

To realize this functionality, the graph attention layer in GATs performs several operations on graph-structured data. Initially, each node undergoes a joint linear transformation through a weight parameter matrix W , setting the stage for further processing and analysis.

Following the initial transformation in graph attention networks (GATs), the process continues with the computation of attention coefficients. These coefficients represent the non-normalized attention weights calculated pairwise between neighboring nodes. At this stage, the z embeddings of two adjacent nodes are concatenated, forming a combined vector. This concatenated vector is then subjected to a dot product operation with a learnable weight vector, effectively integrating the node features into the attention mechanism.

Subsequently, to introduce nonlinearity into the model, the LeakyReLU activation function is applied to the result of this dot product operation. The next step involves normalizing the attention coefficients to maintain consistency across all nodes. This is achieved through the application of the softmax function, ensuring that the coefficients are comparable and appropriately scaled.

In the aggregation phase, the model combines embeddings from various neighbors, guided by the calculated attention weights. However, an important consideration in GATs is the potential instability of self-attention. To address this issue, the concept of multi-head attention is employed. This approach involves creating multiple attention mechanisms or ‘heads,’ each with its distinct parameters. These multiple heads operate in parallel, enhancing the model’s capacity and stability. Each head computes its own output, which is subsequently integrated to form the final output. Typically, the outputs of these heads are concatenated in intermediate layers of the network, while averaging is used in the final layer to consolidate the information gleaned from different perspectives. In GATs, attentional weights are determined implicitly by comparing inputs to each other (a process known as self-attention) as can be seen in Figure 4. The above mathematical relation concerning the computation of vector representations of nodes is transformed as follows:

$$h_i = \sum_{j \in N_i} a_{ij} W x_j$$

where a_{ij} are the attention weights calculated dynamically by the network.

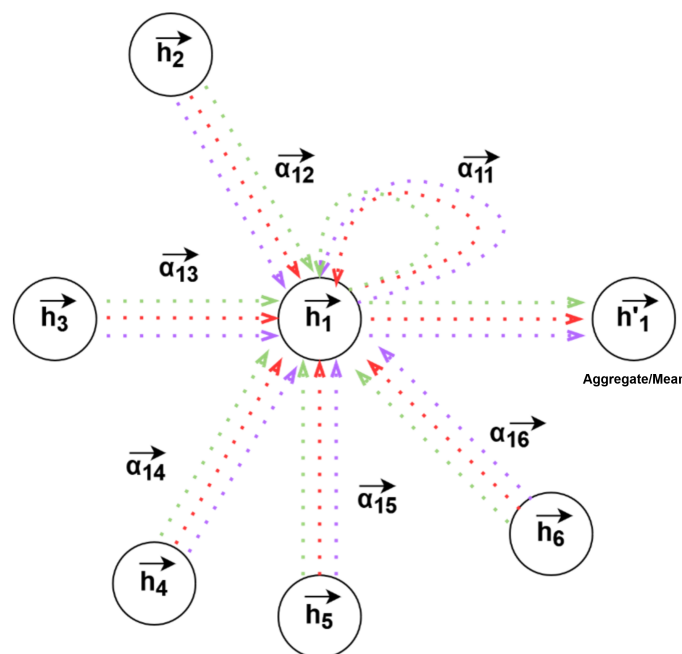


Figure 4. Illustration of the multi-head attention mechanism with three attention heads. Each arrow is color-coded to represent independent calculations of attention weights. The aggregated features from each head are subsequently merged or averaged to produce the final vector representation of a node.

2.3. Graph Attention Network Version 2 (GATv2)

GATv2 [2,4] is an extension of graph attention networks that addresses the issue of static attention observed in the original GATs. Static attention refers to a condition where the ranking of attention for the primary nodes is identical for all query nodes. GATv2 resolves this issue by introducing dynamic attention, which allows for the flexibility of attention weights to vary according to the query node. This is achieved by altering the attention coefficient calculation process. Specifically, the embeddings of the two nodes are concatenated and then subjected to a nonlinear Leaky ReLU activation function. The resulting output is then multiplied by a learnable weight vector through a dot product operation. This approach enables GATv2 to better model the graph structure and capture important node interactions by allowing for dynamic adjustments in attention weights.

3. Graph Attention Network Categories

In this section, we present six distinct categories of graph attention networks (GATs), which are differentiated by their core methodological approaches, the application of attention mechanisms, the nature of features captured, and the techniques utilized to address particular challenges in graph learning. In our assessment, these six categories represent an appropriate classification for the diverse range of GAT methodologies. We believe that this taxonomy effectively captures the different approaches in graph learning. In parallel, we have compiled a table (which can be found in Section 3.6) featuring the most cited studies in this field, providing a clear and organized overview of the influential and impactful research on GAT methodologies (see Table 1).

3.1. Global Attention Networks

Global Attention Graph Neural Networks represent a sophisticated approach in the field of graph-based deep learning, designed to capture and utilize both local and global contextual information within graph-structured data. These networks enhance traditional graph neural networks by introducing attention mechanisms that allow the model to focus on the most relevant parts of the graph, thereby improving its ability to capture intricate patterns and relationships. By leveraging global attention, these networks are capable of understanding the broader context in which nodes and edges exist, which is crucial for tasks where the relationships between distant nodes or the overall structure of the graph play a significant role. This ability to aggregate information from across the entire graph, rather than just from immediate neighbors, enables Global Attention GATs to produce more accurate and robust node embeddings, leading to better performance in a wide range of applications, from recommendation systems to sentiment classification and beyond.

The CGAT model (Contextualized Graph Attention Network) [5] successfully captures both local and non-local contexts in knowledge graphs for enhanced recommendation methods. By combining a user-specific graph attention system, a biased random walk process, and an item-specific attention system, CGAT demonstrates superior performance compared to existing approaches. Addressing the challenges of global sequence contexts and structural syntax in aspect–category sentiment classification, the BiGAT model [6] employs graph attention networks, Biaffine modules, and aspect-specific mask operations. This method improves the capture of relations between words, resulting in better classification and outperforming existing methods.

The HFGAT framework (hybrid framework based on GAT) [7] offers a novel approach to predicting metabolic pathways by combining global and local characteristics of compounds. This method outperforms traditional machine learning and graph convolutional network-based methods, providing valuable insights for drug discovery. GAT_SCNet [8] showcases its effectiveness in recognizing various categories of road markings using point clouds from Mobile Laser Scanning systems. With impressive results exceeding 91% across three criteria, this method sets a new state-of-the-art standard, particularly for linear road markings. The RA-AGAT model [9] addresses stock prediction and recommendation tasks by exploiting intercorrelation and temporal features. When tested on the Chinese A-share market, RA-AGAT outperformed existing approaches. A learnable feature map filtration module and an influence-based graph attention network [10] were introduced for visual position recognition in environments with extreme appearance changes. This approach yielded better results than existing methods, demonstrating its adaptability and effectiveness. Furthermore, the hierarchical graph attention network (HGAT) [11] addresses the challenge of obtaining full global information in semi-supervised node classification. Tests on four datasets revealed state-of-the-art results, with a sensitivity analysis further highlighting HGAT's ability to collect global structure information and transfer node features effectively. The model makes predictions according to the following formula:

$$H^{\text{out}} = \text{softmax} \left(\frac{1}{K} \sum_{k=1}^K \alpha^k W^k H_l^* \right)$$

where $H^{\text{out}} \in \mathbb{R}^{|V| \times |Y|}$ is the prediction of nodes belonging to the class $y_i \in |Y|$, and H_1^* is the concatenated node representation of H_1 and H_{2l+1} .

The category of Global Attention Networks also includes the Holistic Graph Neural Network (HGNN), a two-fold architecture that introduces a global-based attention mechanism for learning and generating node embeddings [12]. By incorporating global features that summarize the overall behavior of the graph, in addition to local semantic and structural information, the HGNN ensures that each individual node is aware of the global behavior of the graph outside its local neighborhood. A more sophisticated hierarchical global feature extraction mechanism is also proposed as a variant of HGNN, further exploring diverse global pooling strategies to derive highly expressive global features. This approach exemplifies the continuous innovation in the field, as researchers strive to develop more effective methods for leveraging both local and global information within graph-structured data.

In summary, the category of Global Attention Networks showcases the versatility and effectiveness of these models across a wide range of applications. The studies presented demonstrate how these networks have been successfully employed to tackle diverse challenges, such as recommendation systems, sentiment classification, drug discovery, road marking recognition, stock prediction, and visual position recognition. The consistent improvements and state-of-the-art results achieved in each application highlight the potential of Global Attention Networks in capturing and utilizing complex relationships and global features within graph-structured data. As research in this area continues to advance, we can expect further enhancements in model performance and an expansion of applications that benefit from these innovative approaches.

3.2. Multi-Layer Graph Attention Networks

Multi-Layer Graph Attention Networks (Multi-Layer GATs) are a subcategory of graph attention networks (GATs), which involve stacking multiple attention layers to capture complex higher-level features in graph-structured data. Multi-Layer GATs stack multiple layers of attention to further enhance their ability to learn from graph-structured data. In this architecture, each layer learns a new set of node features based on the features of neighboring nodes, with the output from one layer being fed as input to the next layer. This enables Multi-Layer GATs to learn more abstract, higher-level features by aggregating information from a larger neighborhood of nodes in the graph.

Towards this direction, FinGAT [13] is a deep learning-based model that leverages Multi-Layer GATs for stock recommendations, capturing short- and long-term temporal patterns from stock price timelines using fully connected graphs among stocks and sectors. In a related work, sparse graph attention networks (SGATs) [14] were designed to identify and sparse out irrelevant or noisy edges in graph-structured data by learning sparse attention coefficients. Another study introduced HGHAN [15], which identified hacker groups using a heterogeneous graph attention network (HAN), outperforming other heterogeneous graph node embedding algorithms. A graph-based circRNA–disease association prediction method [16] was also presented, showing improved performance compared to existing methods. The DuGa-DIT model [17] is a dual gated graph attention model with dynamic iterative training, addressing problems in traditional entity matching techniques and validating its effectiveness on benchmark datasets and a cross-lingual personalized search case. A three-channel approach, including a Heterogeneous Edge-enhanced graph ATtention network (HEAT) [18], was proposed to improve the decision-making and planning modules of autonomous vehicles by accurately predicting multiple agent trajectories. Robust Representation Learning (RRL-GAT) [19] was developed for more accurate multi-label image characterization, employing a Class Attention Graph Convolution Module (C-GAT) and an Adaptive Graph Attention Convolution Module (A-GAT) to detect the communication structure of categories and assess the dynamic connection between objects. A deep learning model based on the hierarchical graph attention network for miRNA–disease associations (HGANMDA) [20] was developed to predict miRNA–disease associations, out-

performing existing methods. A virtual sensor-based imputed graph attention network [21] was proposed to improve anomaly detection in complex equipment with incomplete data. The MV-GAN model [22] was developed for travel recommendation, incorporating user and product representations from multiple sources and applying a view-level attention mechanism for efficient combining of node representations. The CellVGAE model [23] is a graph autoencoder model for studying scRNA-seq data, outperforming other deep learning architectures in terms of training times. GAT-LI [24] is a graph learning and interpretation system that utilizes GAT2 and GNNExplainer to better understand autism spectrum disorders and the underlying biological mechanisms.

The MDGAT architecture and matcher [25] applied an attention method to 3D point clouds, improving data association between 3D points for LiDAR-based SLAM systems and mapping. An innovative topology-adaptive, high-speed transient stability assessment (HSTSA) scheme using a novel multi-graph attention network with a residual structure (ResGAT) and a new piece-wise transient stability index (PSI) [26] was proposed, showing superior accuracy and resistance to different scenarios when tested on an IEEE 39-bus system and IEEE 300-bus system. HGATMDA [27] is a new method for predicting miRNA–disease associations, using a heterogeneous graph to extract features and applying a neural network for predictions. The method outperformed current approaches in terms of prediction performance, and three case studies validated its efficacy with 50 validated miRNA–disease pairs. An innovative graph neural network-based system for short text categorization [28] utilized both limited labeled data and large unlabeled data, outperforming existing state-of-the-art methods in both transductive and inductive learning. In another study [29], graph attention was proposed for expression comprehension to identify objects in images based on their descriptions in natural language. Node attention and edge attention were used to capture information related to the objects and their relationships with their environment, revealing a superior method compared to other solutions.

The above-mentioned papers demonstrate the versatility and robustness of Multi-Layer Graph Attention Networks (GATs) in addressing a wide range of complex problems across various domains. The innovative approaches and models presented in these studies highlight the ability of GATs to efficiently handle graph-structured data, extract meaningful features, and improve performance compared to traditional techniques. These advancements in graph-based deep learning have the potential to drive further research and development in diverse fields, including finance, cybersecurity, healthcare, transportation, and natural language processing.

3.3. Graph-Embedding GATs

Graph-embedding GATs, a subcategory of graph attention networks, focus on learning latent coordinates or embeddings for each graph node, effectively capturing the underlying graph structure. By combining the strengths of graph-embedding techniques and attention mechanisms, these networks can efficiently represent and process graph-structured data, enabling them to address a wide range of problems across various domains. Graph-embedding techniques transform the graph structure into a low-dimensional continuous space that retains the essential properties of the original graph. GATs, on the other hand, use attention mechanisms to weigh the importance of neighboring nodes when aggregating features.

In graph-embedding GATs, the attention mechanism is combined with graph-embedding techniques to learn more expressive node representations, capturing both the local and global structure of the graph. Graph-embedding GATs have been successfully applied in various domains, showing promise in improving the performance of graph-based machine learning models and driving further advancements in graph deep learning research.

Towards this direction, MS-GAT (Multi-relational Synchronous Graph Attention Network) leverages graph-embedding techniques to capture subtle interactions in traffic systems, such as data coupling between spatial and temporal dimensions [30]. By learning node embeddings that encapsulate these intricate relationships, MS-GAT is able to provide

enhanced traffic analysis and prediction. Tested on five real-world datasets, this model demonstrates the power of graph-embedding GATs in handling complex graph data, achieving better results than existing solutions. Similarly, a novel heterogeneous graph neural network framework with a hierarchical attention mechanism is introduced to manage multi-relational data with large amounts of entities and relations [31]. This framework utilizes graph-embedding GATs to learn expressive node representations that effectively capture the underlying structure of heterogeneous graphs. The approach, tested on a variety of heterogeneous graph tasks, further highlights the potential of graph-embedding GATs in various graph-based applications, as it outperforms existing state-of-the-art models.

HFGAT [7], a novel hybrid framework, integrates global and local characteristics of compounds for predicting their metabolic pathways. This approach has proven valuable for drug discovery, outperforming traditional machine learning methods and graph convolutional network-based techniques in multi-class classification accuracy and F1 scores. Similarly, MKGAT [32] is a computational framework designed to predict miRNA–disease correlations, with successful validation for three human cancers. Graph4Web [33] is a relation-aware graph attention network for web service classification. By parsing web service descriptions into a dependency graph and leveraging pre-trained BERT embeddings, Graph4Web demonstrates improved classification performance compared to seven baseline methods. Another unified framework for graph attention networks [34] combines graph context information and node representations to achieve better performance in semi-supervised node classification. A new framework for explainable recommendation using a knowledge graph attention network model [35] achieves high recommendation accuracy and provides interpretable visual explanations. DTIHNC [36] integrates heterogeneous networks and cross-modal similarities to better understand drug–target interactions, outperforming existing methods. A GAT model in a study [37] accurately predicts Parkinson’s disease (PD) by combining morphological, structural, and functional features, identifying regions with the greatest PD impact. A novel source code model [38] combines abstract syntax trees (ASTs) and control flow graphs (CFGs) with a graph attention mechanism-based neural network, improving program classification and defect prediction. A system using heterogeneous graphs, self-enhanced graph attention networks, and tri-aggregator neural networks [39] identifies drug–virus associations, outperforming existing models and identifying potential SARS-CoV-2 treatments.

GANLDA (graph attention network for lncRNA–disease associations) [40] predicts lncRNA–disease associations by combining lncRNA and disease heterogeneous data, showing significant improvement over current methods. CAMT (Context-Aware method to learn invocations patterns and descriptions for Mashup Tagging) [41], a context-aware mashup tagging algorithm, leverages neural networks to consider high-level linkages in two graphs and a multi-head attention mechanism to differentiate adjoining mashups’ significance. SGANM (self-adaptation graph attention network via meta-learning) [42], a self-adaptive graph attention network, employs meta-learning to quickly recognize new fault types with few samples, outperforming other few-shot learning algorithms. GATrust [43], a trust assessment framework for online social networks, integrates information from different aspects and uses graph attention and convolutional networks to generate accurate trust predictions, showing improvements of 4.3% and 5.5%. GSCS [44] employs graph attention networks, an RNN-based sequence model, and a transformed embedding layer to generate high-quality Java method summaries. In predicting drug–drug interactions (DDIs) for poly-drug treatments, DGAT-DDI [45] encapsulates information from drug sources, targets, and self-roles, outperforming state-of-the-art models. The Attention-Gated Conditional Random Fields (AG-CRF) model [46] learns and fuses representations for pixel-level prediction, achieving state-of-the-art performance in monocular depth estimation, object contour prediction, and semantic segmentation. GAT_SCNet [8] is a graph attention network for recognizing road markings from point clouds generated by Mobile Laser Scanning systems. With results exceeding 91% under three criteria, GAT_SCNet sets a new state-of-the-art in linear road marking recognition. MV-GAN [22], a travel recommendation model, incorpo-

rates user and product representations from multiple sources and efficiently combines node representations to capture user and product patterns. To capture the latent embeddings of both users and products, the model incorporates an embedded propagation layer linking the user and product entities. The embedding update mechanisms for users and products are governed by propagation-based and pooling-based rules, which can be formulated as follows:

$$\mathbf{u}_i^{0,l+1} = \sigma\left(\mathbf{W}^{l+1} \times \left(\mathbf{u}_i^{0,l} + \text{AGG}\left(\mathbf{v}_j^{0,l} \mid j \in \mathcal{N}_{ui}\right) + \mathbf{b}^{l+1}\right)\right), \quad (1)$$

$$\mathbf{v}_j^{0,l+1} = \sigma\left(\mathbf{W}^{l+1} \times \left(\mathbf{v}_j^{0,l} + \text{AGG}\left(\mathbf{u}_i^{0,l} \mid i \in \mathcal{N}_{vj}\right) + \mathbf{b}^{l+1}\right)\right), \quad (2)$$

where $\mathbf{u}_i^{0,l} \in \mathbb{R}^{D \times 1}$ is the free embedding of user u_i on the l -th layer, and $\mathbf{v}_j^{0,l} \in \mathbb{R}^{D \times 1}$ is the free embedding of product v_j on the l -th layer. D is the embedding size. $\mathbf{W}^{l+1} \in \mathbb{R}^{D \times D}$ and $\mathbf{b}^{l+1} \in \mathbb{R}^{D \times D}$ are the learned weight and learned bias at step $l + 1$. σ is a nonlinear activation function, specifically LeakyReLU. \mathcal{N}_{ui} represents the travel products clicked by user u_i , and \mathcal{N}_{vj} denotes the users that click item v_j . $\text{AGG}(\cdot)$ is the aggregation function, such as averaging or max-pooling operation.

AR-KGAN [47] jointly embeds fact triplets and logical rules to complete knowledge bases with enhanced accuracy, outperforming existing methods on three benchmark datasets. GAT-GO (Graph Attention Networks for Gene Ontology) is a graph attention network (GAT) method designed to substantially improve protein function prediction by leveraging predicted structure information and protein sequence embedding. Traditional computational methods for protein function prediction can be fast, but they often lack satisfactory accuracy. GAT-GO aims to address this limitation by taking advantage of recent breakthroughs in protein structure prediction and protein language models. The GAT-GO method [48] integrates both structural information and sequence embedding to better predict protein functions. By combining these two sources of information, GAT-GO is able to capture both local and global features of proteins, which, in turn, enhances its predictive power. The graph attention mechanism in GAT-GO helps the model to focus on the most relevant parts of the protein structure and sequence, effectively improving the accuracy of the predictions. The KGAN-CDA (Knowledge Graph Attention Network for CircRNA–Disease Association) model [49] presents a novel computational approach to predicting circRNA–disease connections, capturing both low-level and high-level neighbor information from diverse associations. Its performance surpasses existing methods, as demonstrated by cross-validation results and a case study. In the HGATLDA (Heterogeneous Graph Attention Network for lncRNA–Disease Association) framework [50], the authors leverage node features, relationships in the network, heterogeneous topological structures, and semantic information from metapaths to accurately predict lncRNA–disease associations, showcasing the framework’s effectiveness. The study [51] proposes a novel Heterogeneous Relational Graph (HRG) and a Multiplex Relational Graph Attention Network (MRGAT), along with a connecting embedding (ConnectE) model for the Knowledge Graph Entity Typing (KGET) task. This approach leads to improved entity type predictions and better integration of entity typing tuple with entity relation triples for enhanced entity classification. Finally, MGA-Net [52] introduces a novel few-shot learning (FSL) model that combines data augmentation, embedding network, and graph attention network to address the issue of insufficient data for Synthetic Aperture Radar (SAR) target classification.

3.4. Spatial GATs

Spatial graph attention networks (spatial GATs) are specialized graph neural networks designed to focus specifically on the spatial relationships between nodes in a graph. Unlike traditional GATs that may consider general node connections, spatial GATs emphasize the importance of spatial proximity and spatial dependencies in the graph structure. They employ an attention mechanism to dynamically prioritize the influence of neighboring nodes based on their spatial relevance, allowing the model to effectively capture spatial patterns

and local interactions. This makes spatial GATs particularly useful in applications like urban planning, traffic flow prediction, and remote sensing, where understanding the spatial context and relationships between entities is crucial for accurate analysis and prediction.

In video anomaly detection, deep learning-based models such as the Spatial–Temporal Graph Attention Network (STGA) [53] have been developed to capture the spatial and temporal relationships in video data. STGA leverages spatial and temporal attention to capture local information from the graph, achieving state-of-the-art results on popular benchmarks. Another recent work in graph neural networks is the SA-GAT model [54], which introduces a Substructure Interaction Attention module to improve graph classification performance. The SA-GAT outperformed existing graph kernel and graph neural network approaches, demonstrating its effectiveness for complex graph analysis.

In the domain of multi-domain dialogue systems, a multi-task learning framework that incorporates graph attention networks (GATs) was proposed to tackle cross-domain slot sharing and dialogue act temporal planning. The study demonstrated superior performance compared to existing methods on the MultiWOZ 2.0 and 2.1 datasets [55]. A graph-based VCA model employing graph attention networks was proposed to simulate land use change by quantifying spatial interaction between urban entities. The proposed model was tested with data from Queensland and showed superior performance compared to existing CA models. The study also highlighted the importance of tuning discrete topological orders to improve calibration efficiencies [56].

The model employs a high-order neighborhood extension to iteratively explore the neighbors of each node by utilizing the original adjacency matrix A . This process collects all identified neighbors to form a new high-order adjacency matrix. The k th-order ($k = 1, 2, 3, \dots$) adjacency matrix A^k is defined to represent the nodes connected through k hops. Mathematically, A^k is derived by multiplying matrix A by itself $k - 1$ times, such that $A^1 = A$, $A^2 = A \times A$, and, in general, $A_k = A \times A \times \dots \times A^k$ (k times).

To capture the cumulative neighborhood information up to K hops, the K -order adjacency matrix \tilde{A}^K is introduced, which includes all neighbors from one hop to K hops. This matrix is computed by summing all adjacency matrices from each order: $\tilde{A}^K = \sum_{k=1}^K A^k$, where K is a hyperparameter representing the maximum order of the neighborhood considered in the model.

In the field of Smart Enterprise Management System (EMS) knowledge graphs, a Tensor-based Graph Attention Network called MR-GAT was proposed to enhance the accuracy and depth of the fusion of dense and multi-relational knowledge graphs. A Relation Attention Mechanism and joint Entity and Relation Alignment Framework were used to enhance Knowledge Fusion accuracy. The study of three datasets demonstrated the superiority of the proposed MR-GAT in representation learning for Knowledge Fusion on Smart EMS [57].

A method combining a graph-level attention network and a graph neural network (GNN) has been presented to predict lncRNA–disease associations. The proposed method, called gGATLDA, outperformed other methods and demonstrated its efficacy in identifying lncRNAs associated with different types of cancer such as breast cancer, gastric cancer, prostate cancer, and renal cancer [58]. A hybrid framework called HFGAT was introduced, which combined both global and local characteristics of a compound to predict its metabolic pathways. The framework leveraged Spatial graph attention networks and provided a valuable method for drug discovery by determining all the metabolic reactions involved in the decomposition and synthesis of pharmaceutical compounds. The proposed method outperformed traditional machine learning methods and graph convolutional network-based methods in terms of multi-class classification accuracy and F1 scores [7].

TrajGAT is a deep learning-based system that integrates spatial and temporal data from GPS trajectories to recover missing observations [59]. It splits the problem into two parts: trajectory prediction based on existing data, and substitution of the missing observations. A rule-based graph attention network and a vectorized lane-level map were used to analyze dynamic spatial patterns, while an encoder–decoder structure extracted

and fused temporal features. TrajGAT showed superior performance compared to other models with a strong resilience and wide range of missing trajectory rates. In EEG analysis, a deep learning-based approach was proposed for automated epileptic seizure detection, utilizing spatial and temporal data from EEG channels [60].

A deep spatial–temporal convolutional graph attention network was proposed to capture traffic dependencies among different regions, leveraging multi-resolution transformer networks combined with attentive graph neural networks and convolutional networks [61]. Additionally, a channel-aware recalibration residual network was designed to inject spatial contextual signals into the framework, resulting in a 5% performance improvement.

MetaSTGAT [62] was introduced as a meta-learning system for efficient management of traffic signals in urban transportation systems and public transportation. The dynamic weight generation system could detect changing features of graph nodes, enabling the system to adjust to dynamic intersection traffics, leading to a reduction in travel time by up to 19.30%. An attention-based spatiotemporal graph attention network [63] was proposed for traffic flow forecasting, which could accurately capture past, current, and future temporal relationships. The model was evaluated against traditional and prevalent methods, demonstrating superior performance when predicting medium to long-term traffic flow forecasting.

An attention-based spatiotemporal graph attention network [63] was proposed and evaluated for traffic flow forecasting, achieving accurate computation of past, present, and future temporal relationships. Comparing the results to traditional and prevalent methods, this model outperformed the baselines when predicting medium- to long-term outcomes. EvoSTGAT [64] is a Changing Spatiotemporal Graph Attention Network that was introduced for predicting approaching pedestrian paths, using a fluctuating and dynamic attention strategy to account for the social impact of current procedural pedestrians. The model was tested on two intricate datasets, verifying its proficiency.

FTPG uses a graph attention network (GAT) to accurately and sturdily predict traffic information at intersections [65]. In addition, an approach for estimating the position of the starting point of the queue was proposed, and a spatiotemporal residual graph attention network (ST-RGAN) was employed to further improve prediction precision. GAT_SCNet, ref. [8] was designed for recognizing various categories of road markings from point clouds produced by Mobile Laser Scanning systems (MLSs). Tests were conducted on a total of 100km captured by different MLS systems, with results exceeding 91% under three criteria (average precision, recall, and F1).

ASTGAT [66] is an adaptive spatial GAT that simultaneously discovers dynamic graph structures and spatial–temporal associations for traffic flow prediction. It outperformed existing models on the METR-LA and PEMS-BAY datasets, as demonstrated by both numerical and visual evaluations of forecasted traffic flow. For predicting pedestrian trajectories in unpredictable environments, a graph attention-based model called PTPGC was introduced in [67]. It leverages TCN and ConvLSTM in a dynamic graph to capture pedestrian features and spatial connections. Results on two datasets demonstrate PTPGC's improved trajectory prediction performance over existing baselines. A novel self-adaptive graph attention network, SGANM, is introduced for quick recognition of new fault types with limited data samples. SGANM utilizes meta-learning strategies for enhanced meta-knowledge learning ability, resulting in improved few-shot learning performance on benchmark datasets and a practical platform [42]. A mutually supervised few-shot segmentation network was proposed in [68] for image segmentation. The model utilizes feature maps from the middle convolution layers and merges the support image and query image into a two-sided graph, using graph attention network for graph reasoning for segmentation. Experimental results demonstrate the effectiveness of the proposed model. KGANCDA is introduced for predicting circRNA–disease connections. The model utilizes a knowledge graph attention network to capture low-level and high-level neighbor information from diverse associations. The KGANCDA model outperforms existing methods in cross-validation and a case study demonstrates its effectiveness for predicting circRNA–disease connections [49]. A

scheduling algorithm based on graph attention networks was proposed for controlling multiple robots in a warehouse, which outperformed existing methods when trained with imitation learning. The approach required no expert knowledge and was applicable to large-scale problems [69].

The Spatiotemporal Gated Graph Attention Network (STGGAT) [70] model was introduced for predicting traffic flow by utilizing License Plate Recognition (LPR) records to calculate average travel times and volume transition relationships. The model combined a gated recurrent unit layer, a graph attention network layer with edge features, a bidirectional long short-term memory, and a residual structure. Validation with the LPR system of Changsha, China showed that STGGAT was more accurate and reliable than existing baseline models and was capable of inductive learning and fault tolerance.

EGAT [71] was created for object detection in high spatial resolution remote sensing imagery (HSRI). The network was able to access important information of the rationale graphs to distinguish between different spatial–semantic correlations by introducing a long short-term memory (LSTM) device into EGAT. The experiments showed that the strategy provided state-of-the-art results and was an effective method for object detection in HSRI.

SSGAT [72], a novel model for HSI classification, was designed to combine information of labeled and unlabeled samples in an unsupervised manner in order to address the issue of the insufficiency of labeled samples. Evaluation results showed its effectiveness and superiority over existing approaches. A self-adaptive graph convolution network (GCN), an encoder–decoder network, and a conditional random field (CRF) algorithm were utilized to enhance the semantic segmentation of laser scanning (LS) data. Promising results were achieved for multiple objects using the ParisLille-3D, Semantic3D, and vKITTI datasets [73]. TAGAT [74] is designed to consider type-related information during the embedding process and used a hierarchical attention mechanism to increase interpretability of the embedding space and improve reasoning performance.

A unique air temperature forecasting system was created using graph attention networks and gated recurrent units, resulting in better predictions than existing benchmark models [75]. HLGAT was introduced as a novel deep graph model for person re-identification tasks. By using the attention mechanism to aggregate local features, it captured both intra- and inter-local relations and outperformed current state-of-the-art methods [76]. A graph-based deep learning framework has been proposed to model dependencies between objects and triplets in Visual Relationship Detection. By leveraging graph attention network and encoding prior knowledge in graph generation, the framework improved the performance of VRD over existing state-of-the-art models [77]. Graph attention has also been proposed for expression comprehension by identifying objects in images based on their descriptions in natural language. By utilizing node and edge attention to capture and utilize information related to objects and their relationships with their environment, the approach showed superior performance compared to other methods [29].

3.5. Variational GATs

Variational Graph Attention Networks (VGAT) represent an advanced architecture within the domain of graph neural networks, designed to handle the complexities inherent in various types of data. They integrate the principles of graph attention mechanisms with variational inference, enabling it to effectively capture and leverage intricate relationships between nodes in a graph. This approach is particularly useful in scenarios where data are heterogeneous or multimodal, as it can seamlessly integrate and analyze diverse data types, such as in the fields of power systems, drug–target interactions, and medical image processing.

A VGAT-based [78] approach was proposed to distinguish between transient rotor angle instability and short-term voltage instability in power systems. A label-smoothing technique was incorporated to address the issue of label inaccuracy. The approach was tested on an eight-machine 36-bus system and Northeast China Power System, showing improved performance over GCNs and other machine learning methods.

A VGAT has been proposed for integration of heterogeneous networks and cross-modal similarities of drugs, proteins, diseases, and side effects to better understand drug–target interactions [36]. A denoising autoencoder and VGAT were used to reduce the feature dimensionality, while a multiple-layer convolutional neural network was used to further enhance the results. Experiments had shown that the DTIHNC method yielded better results than other state-of-the-art methods. In the field of medical image processing a VGAT has been trained on a dataset of 302 CT scans in order to distinguish complicated Pneumonia-Related PPE from uncomplicated PPE and normal cases achieving a high accuracy of 86.7% [79]. QPGAT [80] combined quantum probability and graph attention mechanisms to accurately map the relationship between pieces of text and model each text node as a particle in a superposition state. It was applied to two complex NLP tasks: emotion–cause pair extraction and joint dialog act recognition and sentiment classification, and the network was found to be competitive with other methods on both tasks.

PSCR [81], was developed, to create FBN structures in order to diagnose autism spectrum disorder (ASD). The PSCR and VGAT framework achieved an accuracy of 72.40% in diagnosing ASD, exceeding the results from other FBN construction methods and classification frameworks.

A seizure detection method that utilizes graph attention networks to detect epilepsy through EEG data has been proposed [82]. The method employs a graph structure to leverage the positional correlations between different EEG signals and addressed data imbalance through the focal loss in its loss function. It achieved an accuracy, sensitivity, and specificity of 98.89%, 97.10%, and 99.63%, respectively. MGAT was introduced [83] as a graph-embedding approach that utilized an attention mechanism to capture various types of relationships within multi-view networks. The model showed its effectiveness in outperforming existing baselines on several real-world datasets. An MGAT has also been proposed for personalizing recommendations based on user interaction data from different sources [84]. The method utilized a gated attention mechanism to refine personal interests according to the modality, allowing it to recognize complex patterns and generate better-targeted recommendations than current methods. Experimental results on the Tiktok and MovieLens datasets showed that MGAT outperformed existing methods.

In this model, both users and items are associated with unique IDs. A common approach to representing this ID information is through embedding, where the ID is transformed into a vectorized representation. Specifically, a user u and an item i are projected into vectors e_u and e_i , respectively, which encapsulate their general characteristics. Additionally, within individual interaction graphs, each item i is associated with a pre-existing feature, denoted as $e_{m,i}$, that emphasizes its characteristics in the m -th modality. Furthermore, an additional embedding $e_{m,u}$ is assigned to each user u to capture the user's preference within the m -th modality. The complete set of embeddings is summarized as follows:

$$\mathbf{E} = \{e_u, e_i, e_{m,u}, e_{m,i} \mid u \in U, i \in I, m \in M\}$$

where $e_u, e_{m,u} \in \mathbb{R}^{|U| \times d}$ and $e_i, e_{m,i} \in \mathbb{R}^{|I| \times d}$, N and M denote the numbers of users and items, respectively, and d is the embedding size. It is worth noting that e_i, e_u , and $e_{m,u}$ are randomly initialized and trained during optimization, while $e_{m,i}$ is derived from fixed features via a trainable neural network.

3.6. Hybrid GATs

Hybrid Graph Attention Networks (GATs) are a type of computational framework that combine multiple technologies to perform optimally in diverse tasks. These networks have been used in various applications, including miRNA–disease connection discovery, multi-agent dynamic traffic control, knowledge graph completion, character representation, and session-based recommendation.

GATMDA [85], is a computational framework that applied graph attention network and multi-source information to uncover miRNA–disease connections. GATMDA achieved

an average AUC of 0.9566 and identified a high percentage of verified candidates. DQ-GAT [86] is a graph attention-based deep Q learning model that was proposed for multi-agent dynamic traffic scenarios. DQ-GAT yielded higher success rates and better balancing between safety and efficiency than previous deep learning and traditional rule-based approaches in both seen and unseen scenarios. A multi-relational graph attention network (MRGAT) was proposed for knowledge graph completion [87]. Experimental results demonstrated the superiority of MRGAT compared to other models. In the field of character representation, a polymorphic grapheme attention network has been presented [88] to dynamically capture the relevance between characters and their corresponding words. This improved the performance of character representation in all four dimensions and resulted in significant performance improvements in experiments.

CKSR is a knowledge-aware session-based recommendation model, proposed in [89], that combined a cross-session graph and a knowledge graph to create a cross-session knowledge graph. The CKSR model used this graph to capture transition patterns among interacted items and outperformed other SBR methods in experiments on two benchmark datasets. A multi-view framework for anomaly detection of Internet routing traffic has been presented [90]. It utilizes seasonal and trend decomposition combined with a graph attention network to recognize relationships among multiple features and correlations in time. The framework achieved improved anomaly detection performance, with an F1 score of 96.3% and 93.2% on balanced and imbalanced datasets, respectively. Moreover, the proposed framework could be extended to detect unseen anomalous events. GTGenie [91] was proposed as a computational model for discovering biomarker–disease associations. The model made use of graph attention networks and pre-trained BERT-based models and achieved competitive performance on a variety of benchmark datasets. OmicsGAT [92] is a graph attention network, designed for analyzing RNA-seq data to identify deeper relationships between gene expression and network structure for cancer subtype analysis, patient stratification, and cell clustering. The algorithm's multi-head attention mechanism gave attention coefficients to a sample's neighbors, allowing the capture of information related to a sample more effectively and giving visibility into the importance of particular neighbors in any specific sample's cancer subtype analyses.

An RVTR [93] graph attention network-based RNA virus transmission network representation model made use of natural language processing and a graph context loss function to more accurately train the model in detecting asymptomatic propagators of COVID-19. GTGAT [94] is a gated tree-based graph attention network that improved upon the success of graph attention networks (GAT) for transductive and inductive reasoning in generalized knowledge graphs. The approach was successful in transductive tests and outperformed existing methods when applied to medical knowledge graphs in inductive tasks. GANet [95], is a deep learning-based model, which was proposed to obtain point correspondences from two-view images, resulting in improved mean Average Precision (mAP) by up to 1.5% and 0.6% on the YFCC and SUN3D datasets, respectively. GCHGAT [96] is a group-constrained hierarchical graph attention network that has been presented for predicting pedestrian trajectories, incorporating pair-wise and group-wise interactions between individuals. GCHGAT achieved superior performance with the smallest prediction error when compared to other methods on the ETH and UCY datasets. In the field of fake news detection, a graph attention network-based approach was created to identify potential fake news on social networks by analyzing user–user connections and content graphs. The method demonstrated improved accuracies and F1 scores compared to existing methods, as demonstrated through experiments conducted on two datasets [97]. TGAT [98] is a multi-relational graph attention network framework that was introduced to bridge the gap between IoT data and intelligent applications and services. TGAT can capture rich interactions between mixed triples, entities, and relationships, while the Tucker model was employed to reduce storage and calculation consumption. TGAT achieved up to 7.6% improvement in the hits@1 accuracy compared to other models on real-world heterogeneous graphs. A model [99] that combined convolutional neural networks and

a graph neural network was presented for predicting crowd flow in cities partitioned into irregular regions based on road networks and functionality. The model utilized a location-aware and time-aware graph attention mechanism called Semantic Graph Attention Network (Semantic-GAT) that relied on dynamic node attribute embedding and multi-view graph reconstruction. Experimental results showed that the model effectively reduced the prediction error.

A heterogeneous graph attention network was proposed to detect and forecast food safety risks based on various variables. The proposed risk profile was tested on a dataset from China and was found to be accurate [100]. GraphReg [101] is a deep learning technique that models the effect of non-coding genetic variations on target gene expression. It uses 3D interactions from chromosome conformation capture tests to predict gene expression and is better than current state-of-the-art deep learning approaches. It can also predict direct transcription factor targets by *in silico* deletion of other transcription factor binding motifs. In the field of autonomous driving, DRL-GAT-SA [102] has been proposed. It is an autonomous driving safety system that combines graph attention reinforcement learning and dynamic safety assurance for efficient driving in uncertain conditions. A deep semantic information propagation method [103] was proposed for aligning a single labeled source domain and multiple unlabeled target domains. Experiments on four public datasets showed that this method outperformed other prominent domain adaptation methods. A virtual sensor-based imputed graph attention network [21] that could improve anomaly detection in complex equipment with incomplete data by combining real signals and simulated signals with a graph attention network has also been presented. A Learning from Demonstration (LfD) approach using graph attention networks (GATs) was presented in [104]. The approach demonstrated how a robot could learn non-interactive, interactive, and uni- or bi-manipulative operations from a human by receiving human hand paths and goal object poses. It was tested on simulated data and through real-life experiments, and the results showed that the robot could successfully learn the tasks. EGRET [105] is a transfer learning-based approach, which had yielded increased evidence of protein–protein interaction sites over alternate methods. EGRET was available on open-source GitHub, and the network behavior had been studied to explain how the decisions were made. A single-pixel compressive direction of arrival (DoA) estimation technique has been proposed [106]. It uses a graph attention network (GAT)-based deep-learning framework that leveraged metasurface antenna-based coded aperture to reduce the physical layer dimension. This approach excited the far-field sources incident on the aperture using a set of spatio-temporally incoherent modes to encode and compress the spectrum of the sources into a single channel, eliminating the need for a reconstruction step. In a low signal-to-noise ratio (SNR) environment, the proposed GAT integrated a single-pixel DoA framework that could accurately retrieve the DoA information.

RelMN [107], is a deep sparse graph attention network for object recognition and relationship categorization in visual scenes. It classified object pairs from denser graphs into foreground and background groups, utilized sparse graphs with message passing, and surpassed state-of-the-art results on multiple benchmark datasets. PGAT [108] is a path-enhanced bidirectional GAT that accurately predicts quality indices of a manufacturing process. It utilized graph attention networks to learn the relationships between the different machines in the multistage process and incorporated dependency path information into the machine features. A masked loss function was used to address the label noise problem and batch training could be used for improved efficiency. Experiments on a real-world production line dataset validated the effectiveness of the proposed approach compared to existing methods. PD-RGAT [109] is a model that has been proposed for aspect-based sentiment analysis, which considered phrase information and direction of dependency when constructing the graph. It was able to effectively capture long-range dependencies and investigate aspect–sentiment polarities, with similar effectiveness to state-of-the-art models.

In addition, a graph attention network model for predicting potential miRNA–disease associations from existing data outperformed the state-of-the-art models [110]. It potentially provides a better understanding of numerous biological processes relevant to human disease. Value Decomposition with Graph Attention Network (VGN) [111], is an approach that factored the joint reward function into individual reward functions. Two graph neural network-based algorithms, VGN-Linear and VGN-Nonlinear, were developed and tested on the StarCraft Multiagent Challenge (SMAC) benchmark. VGN methods surpassed existing value-based multi-agent reinforcement simulations in difficult tasks. GATs and ADFP-AC [112] were used for PBT chemical screening and eight new classes of PBT chemicals were identified from the Inventory of Existing Chemical Substances in China. AL-NEGAT [113] is an adversarial learning-based node edge GAT that has been designed for the identification of autism spectrum disorder (ASD) based on multimodal MRI data. This model achieved higher accuracy and better generalizability than existing state-of-the-art methods, making it a promising tool for the identification of brain disorders. The combination of a convolutional neural network (CNN) and graph neural network (GNN) in WFCG for HSI (hyperspectral image) classification has been proposed [114]. It utilized a GAT and CNN to create a weight fusion of their respective features, achieving comparable results to state-of-the-art methods. ASEGAT, is a brain cortical surface parcellation method [115] that employed a graph attention module and a squeeze-and-excitation module to recognize node features and incorporated anatomical prior information to improve the accuracy of region labeling. Experiments on a public dataset showed that ASEGAT outperformed other methods, achieving an accuracy of 90.65% and a dice score of 89.00%. CGAT was presented for ground-based remote sensing cloud classification [116]. It made use of Context Attention Coefficients and two transformation matrices to enhance the discrimination of aggregated features (AFs). Additionally, a new GCD was published, and experiments were conducted to prove the effectiveness of the model. GNNImpute, is a single-cell dropout imputation method, that was introduced and tested on various real datasets [117]. It was found that GNNImpute provided accurate and effective imputation of dropouts and reduced dropout noise. It also produced good metric results on clustering, with ARI and NMI reaching 0.8199 and 0.8368, respectively. HGAT-AMR, a deep graph learning method, was proposed to predict anti-TB drug resistance [118]. This method enabled the consideration of incomplete phenotypic profiles and provided attention scores to identify genes and SNPs associated with drug resistance.

The Hierarchical Graph Attention Network (HGAT) [119] was proposed to accurately analyze sentiment information with its complicated semantic relations for e-commerce platforms. Experiments were conducted, and it was found to outperform other baselines. A MARL problem was proposed to manage resources in a cellular network using GAT and DRL, leading to an efficient inter-slice resource management strategy [120]. GAT was applied over both DQN and A2C, and its superiority was verified through simulations. GATNNCDA [121] was developed as a novel method that could efficiently model and integrate circRNA–disease relationships and predict new potential associations. A molecule-editing graph attention network (MEGAN) was created as a neural model for automated synthesis planning, based on models that described a chemical reaction as a model of graph edits, similar to arrow pushing [122]. The authors adapted it for retro-synthesis and enlarged the dataset size, providing state-of-the-art accuracy in standard benchmarks.

Table 1. Top-cited graph attention network (GAT)-based papers the last years with a minimum of 10 citations.

Model Name	Year	Citation Count	Cites Per Year
N/A [29]	2019	115	28.75
Mgat [84]	2020	44	14.67
N/A [68]	2022	40	40.00

Table 1. Cont.

Model Name	Year	Citation Count	Cites Per Year
WFCG [114]	2022	39	39.00
HGAT [77]	2020	37	12.33
HGAT [28]	2021	34	17.00
Mgat [83]	2020	32	10.67
N/A [123]	2020	31	10.33
MAGAT [124]	2021	28	14.00
N/A [125]	2021	25	12.50
GATrust [43]	2022	25	25.00
N/A [103]	2022	23	23.00
GATMDA [126]	2021	22	11.00
MGA-Net [52]	2022	20	20.00
Hawk [127]	2021	19	9.50
ResGAT [26]	2021	19	9.50
MEGAN [122]	2021	19	9.50
GANLDA [40]	2022	19	19.00
SRGAT [128]	2021	18	9.00
PD-RGAT[109]	2022	18	18.00
HLGAT [76]	2021	18	9.00
HGAT [11]	2020	16	5.33
RRL-GAT [19]	2022	16	16.00
N/A [82]	2021	16	8.00
HEAT [18]	2022	15	15.00
ASTGAT [66]	2022	14	14.00
RA-AGAT [9]	2022	14	14.00
SSGAT [72]	2022	12	12.00
MDGAT [25]	2021	12	6.00
FTPG [65]	2022	12	12.00
Gchgat [49]	2022	11	11.00
HGATLDA [50]	2022	10	10.00
STGGAT [82]	2022	10	10.00
PSCR [81]	2021	10	5.00
EGAT [71]	2022	10	10.00
KGAT [35]	2022	10	10.00

A novel message-dependent attention mechanism was introduced to improve graph neural networks in multi-agent path planning [124], and the model was more capable on large-scale path planning tasks in a decentralized setting, consistently achieving better results than other benchmark models. Hawk [127] is a malware detection framework for Android applications that modeled relationships between Android entities to reveal hidden, higher-order relationships to better detect malicious behaviour. It offered the highest detection accuracy, with a detection time of 3.5 ms for out-of-sample applications and a much quicker training time at 50× faster than traditional methods. CRF-GAT [129], combined the advantages of conditional random fields and graph attention networks to

enable the semi-supervised fault diagnosis of motors. An optimized algorithm with the technique of Clustering with Adaptive Neighbor was also added for graph construction with high accuracy in the diagnosis of motor conditions. A graph attention network scheduler has been presented [123]. It imitated experts’ advice to quickly create near-perfect timetables for robot teams of varying sizes, and could scale up to huge, previously unseen tasks. The research had indicated that this scheduling network had been able to find highly efficient solutions for around 90% of the tests suggested for scheduling two to five robots and up to 100 tasks much more effectively than prior techniques.

4. Applications of Graph Attention Networks

Graph attention networks (GAT) have been widely adopted for graph representation learning, with several Python libraries offering optimized implementations. For the convenience of the reader, we have compiled a list of popular GAT libraries in Table 2. Most of these libraries support standard graph representation learning (GRL) tasks and offer seamless integration with popular machine learning frameworks such as PyTorch, TensorFlow, and JAX (see Table 3).

Table 2. Overview of key domains, case studies, challenges, and applications of graph attention networks.

Domain	Case Study	Problem	Applications of GATs
Healthcare and Bioinformatics	Drug–Drug Interaction Prediction	Predicting potential interactions between drugs is crucial for drug safety and efficacy. Traditional methods may not fully capture the complex relationships between different drugs and their effects on the human body.	GATs can model drug–drug interaction networks by treating drugs as nodes and interactions as edges. The attention mechanism helps to focus on the most relevant interactions, improving the accuracy of predictions.
	Protein–Protein Interaction Networks	Understanding protein interactions is essential for drug discovery and understanding biological processes. Protein-protein interaction (PPI) networks are complex and require sophisticated models to accurately predict interactions.	GATs are applied to PPI networks by treating proteins as nodes and their interactions as edges. The attention mechanism enables the model to focus on the most biologically relevant interactions, improving predictive performance.
Social Network Analysis	Community Detection	Identifying communities within social networks is important for understanding the structure and dynamics of social groups. Traditional methods often struggle with the overlapping and hierarchical nature of communities in large social networks.	GATs can be used to detect communities by focusing on the most influential connections within a network. The attention mechanism allows the model to distinguish between strong and weak ties, which is crucial for accurately identifying communities.
	Fake News Detection	The spread of fake news on social media is a significant problem, and identifying fake news early is critical. Traditional methods may not effectively capture the complex relationships between users and the content they share.	GATs can be applied to social networks where nodes represent users or news articles, and edges represent interactions (e.g., shares or likes). The attention mechanism allows the model to focus on the most suspicious interactions, improving the detection of fake news.

Table 2. Cont.

Domain	Case Study	Problem	Applications of GATs
Finance and Economics	Fraud Detection in Financial Transactions	Detecting fraudulent transactions in financial networks is challenging due to the complex and evolving nature of financial interactions. Traditional methods may fail to capture subtle patterns indicative of fraud.	GATs can be used to model financial transaction networks, where nodes represent entities (e.g., accounts) and edges represent transactions. The attention mechanism helps in focusing on unusual patterns of transactions that are likely to be fraudulent.
	Stock Market Prediction	Predicting stock market movements involves analyzing complex relationships between different stocks, sectors, and external factors. Traditional models may not effectively capture these relationships.	GATs can be applied to stock market graphs, where nodes represent stocks and edges represent relationships (e.g., co-movement or industry links). The attention mechanism helps in identifying the most influential factors affecting stock prices.
Natural Language Processing (NLP)	Document Classification	Classifying documents based on their content can be challenging when the documents have complex structures or when the relationships between different parts of the text are important.	GATs can be applied to document graphs, where nodes represent words or sentences, and edges represent syntactic or semantic relationships. The attention mechanism helps in focusing on the most relevant parts of the document for classification.
	Machine Translation	Machine translation requires understanding the relationships between words and phrases in sentences. Traditional methods may struggle to capture these relationships effectively, especially in complex sentences.	GATs can be used in translation models by treating words as nodes and their relationships as edges in a sentence graph. The attention mechanism allows the model to focus on the most important word relationships, improving translation quality.
Autonomous Vehicles and Robotics	Traffic Flow Prediction	Predicting traffic flow in urban environments is complex due to the dynamic nature of traffic and the numerous factors that influence it, such as road networks, weather, and accidents.	GATs can be applied to traffic networks, where nodes represent intersections or road segments, and edges represent traffic flow between them. The attention mechanism allows the model to focus on the most critical road segments, improving the accuracy of traffic predictions.
	Path Planning for Autonomous Robots	Autonomous robots need to navigate complex environments, which requires efficient path planning. Traditional methods may not effectively capture the complex relationships between different parts of the environment.	GATs can be used to model the environment as a graph, where nodes represent locations and edges represent possible paths. The attention mechanism helps the robot focus on the most relevant paths for efficient navigation.
Chemistry and Material Science	Molecular Property Prediction	Predicting the properties of molecules, such as their toxicity, reactivity, or solubility, is a key task in drug discovery and material science. Traditional models may not fully capture the complex interactions between atoms in a molecule.	GATs can be applied to molecular graphs, where nodes represent atoms and edges represent chemical bonds. The attention mechanism helps in focusing on the most important atomic interactions, improving the accuracy of property predictions.
Telecommunications	Network Anomaly Detection	Detecting anomalies in telecommunication networks is crucial for maintaining network security and performance. Traditional methods may not effectively capture complex, evolving patterns of network traffic.	GATs can be used to model telecommunication networks, where nodes represent devices or servers, and edges represent communication links. The attention mechanism helps in focusing on abnormal patterns, improving the detection of anomalies.

In areas such as recommendation systems, GATs have revolutionized the capability to deliver personalized and relevant suggestions, catering to diverse sectors like entertainment, travel, finance, health, and e-commerce. Their ability to intricately map user–item relationships has led to significant improvements in recommendation precision and personalization. Similarly, in biomedical research, GATs have played a pivotal role in uncovering complex biomarker–disease associations, aiding in the advancement of disease diagnosis and treatment strategies. Their impact extends to natural language processing, particularly in sentiment analysis, where they excel in extracting emotions and opinions from textual

data, ranging from customer feedback to aspect-based sentiment analysis. Moreover, in the field of image analysis, GATs are utilized for sophisticated tasks such as hyperspectral image classification, image super-resolution, and denoising, thanks to their refined attention mechanisms. Finally, in the critical domain of anomaly detection, crucial in sectors like cybersecurity and fraud detection, GATs have proven to be invaluable in identifying deviations from normal patterns in data. The widespread adoption and success of GAT-based algorithms across these diverse fields highlight their flexibility, robustness, and transformative potential in managing and interpreting complex data structures. Table 2, provides a comprehensive overview, encapsulating some of the most significant applications, challenges, and practical use cases associated with graph attention networks (GATs). This summary aims to highlight the critical advancements and areas of interest where GATs have demonstrated remarkable potential, as well as the inherent obstacles that continue to shape ongoing research and development in this domain. Through this synthesis, the reader is offered a clearer understanding of the multifaceted impact and future directions of GATs in various fields.

It is easy to understand that graph attention networks (GATs) have emerged as a powerful deep learning technique for modeling graph-structured data. With their versatility and effectiveness, they have attracted significant attention in the research community and have shown remarkable performance in many different knowledge domains.

Table 3. Popular Python frameworks for graph neural networks (GNNs) and graph attention networks (GATs).

Name	Language	Repository	Framework-Related Paper
GAT	Python	https://github.com/PetarV-/GAT	[130]
pyGAT	Python	https://github.com/Diego999/pyGAT	[130]
keras-gat	Python	https://github.com/danielegrattarola/keras-gat	[130]
pytorch_geometric	Python	https://github.com/pyg-team/pytorch_geometric	[131]
GATv2	Python	https://nn.labml.ai/graphs/gatv2/index.html	[132]
anomaly-detection-resources	Python	https://github.com/yzhao062/anomaly-detection-resources	N/A
dgl	Python	https://github.com/dmlc/dgl	[133]
dgl-lifesci	Python	https://github.com/aws-labs/dgl-lifesci	[134]
dgl-ke	Python	https://github.com/aws-labs/dgl-ke	[135]
benchmarking-gnns	Python	https://github.com/graphdeeparth/benchmarking-gnns	[136]
graph4nlp	Python	https://github.com/graph4ai/graph4nlp	[137]
GNN-RecSys	Python	https://github.com/je-dbl/GNN-RecSys	N/A
GNNLens2	Python	https://github.com/dmlc/GNNLens2	[138]
RNAglib	Python	https://github.com/Jonbroad15/RNAGlib	[139]
OpenHGNN	Python	https://github.com/BUPT-GAMMA/OpenHGNN	[140]
tgl	Python	https://github.com/amazon-science/tgl	[141]
gtrick	Python	https://github.com/sangyx/gtrick	N/A

4.1. Recommendation

The recommender system has emerged as a critical information service in the contemporary landscape of the internet. At the heart of the recommender system lies the fundamental objective of providing relevant and valuable recommendations to a group of users regarding a set of items, products, or services that are likely to align with their preferences and interests. This objective entails a sophisticated mechanism that leverages user feedback and various sources of data to make informed predictions and generate personalized recommendations that cater to the unique needs and preferences of each user [142]. Notably, graph neural networks (GNNs) have emerged as a recent state-of-the-art approach in this domain as they have shown great potential for improving the effectiveness and efficiency of the recommender system [143]. In this vein, graph attention networks (GATs) have been increasingly employed in the development of diverse recommender system applications, including but not limited to movie and media recommendation [144], travel and tourism recommendation [22,145], finance and stock recommendation [9,146], health and medicine recommendation [147], web services and API recommendation [148,149], as

well as e-commerce and package recommendation [150,151]. This highlights the versatility and potential of GATs in advancing the state-of-the-art of the recommender system domain.

4.2. Biomarker–Disease Association

The identification of the relationship between molecular biomarkers and different types of ailments is of paramount importance in elucidating the underlying molecular mechanisms of the disease. Circular RNA (circRNA) is a distinctive class of non-coding RNA molecules that play a pivotal role in a diverse range of pathological conditions. They represent a critical component of the non-coding RNA repertoire, contributing significantly to the modulation of gene expression by sequestering microRNAs (miRNAs) in a regulatory manner [152]. Long non-coding RNA (lncRNA) is a class of RNA molecules that surpass a threshold of 200 nucleotides in length. They modulate the expression of genes at various stages of gene regulation, including transcriptional, RNA processing, translational, and post-translational levels, through intermolecular interactions with nucleic acids and proteins. These molecules serve as molecular guides that facilitate the recruitment of transcription factors to specific DNA binding sites. Additionally, lncRNAs can serve as decoys, which impede the binding of certain proteins to other proteins or nucleic acids [153]. MicroRNAs (miRNAs) are a class of small, single-stranded, non-coding RNA molecules that typically range from 21 to 23 nucleotides in length. These molecules are ubiquitously present in organisms such as plants, animals, and certain viruses. miRNAs play crucial roles in modulating RNA silencing and post-transcriptional regulation of gene expression. Specifically, they participate in regulating gene expression by interfering with the translation of messenger RNA molecules into functional proteins [154]. PIWI-interacting RNAs (piRNAs) are a category of small silencing RNAs that are exclusive to animals and are differentiated from other classes of small RNAs such as microRNAs (miRNAs) and small interfering RNAs (siRNAs). piRNAs function to suppress transposable elements, govern gene expression, and defend against viral infection. They serve as guides for PIWI proteins to recognize and cleave target RNA, and also play a critical role in the promotion of heterochromatin assembly and DNA methylation [155].

Graph attention networks (GATs) have emerged as a promising tool for predicting associations between biomarkers and diseases [91]. Specifically, GATs have been utilized to predict disease associations with a diverse set of biomarkers, including circular RNAs (circRNAs) [49,156], long non-coding RNAs (lncRNAs) [40,50,58,157], microRNAs (miRNAs) [32,85,110,158], PIWI-interacting RNAs (piRNAs) [159], and microbes [126,160]. By leveraging the unique features and relationships of these biomarkers, GATs can effectively capture the complex patterns underlying biomarker–disease associations. These efforts have shown great promise in advancing our understanding of the complex interplay between biomarkers and disease states, and hold great potential for improving disease diagnosis, treatment, and prevention in the future.

4.3. Sentiment Analysis

Sentiment analysis is a natural language processing (NLP) technique that entails the examination and evaluation of digital textual data to identify and categorize the underlying emotional tone, specifically as positive, negative, or neutral. This approach is widely utilized by organizations to ascertain and classify prevailing opinions concerning a particular product, service, or idea. In pursuit of this objective, the utilization of graph attention networks has been suggested for sentiment analysis and has been applied in diverse categories of sentiment analysis, such as aspect-based sentiment analysis [109,161–167], Sentiment Analysis for Multiple Entities and Aspects [168], and Aspect Category Sentiment Classification [6]. Collectively, these endeavors underscore the capacity of graph attention networks as a valuable instrument for sentiment analysis.

4.4. Image Analysis

Image analysis encompasses a range of techniques and methods employed to process an image by breaking it down into its fundamental constituents with the aim of extracting meaningful information. These methods can involve tasks such as shape recognition, edge detection, noise reduction, object quantification, and texture analysis or image quality evaluation. To this end, GATs have been put forward as a formidable tool for diverse forms of image processing. Specifically, GATs have been suggested for applications such as hyperspectral image classification [72,114,125,169], image super-resolution [128,170], image denoising [171], change detection in remote sensing images [172], image multi-label classification [19,173], and text–image summarization [19].

4.5. Anomaly Detection

Anomaly detection, which can also be referred to as outlier detection or novelty detection, is a method of identifying exceptional items, events, or observations that significantly deviate from the majority of the data and do not adhere to a well-defined notion of normal behavior. Such instances may raise suspicion of being generated by a distinct mechanism or seem incongruous with the remaining set of data. Anomaly detection is utilized in various domains such as cyber security, medicine, machine vision, statistics, neuroscience, law enforcement, and financial fraud detection [174]. Due to their robustness and versatility, graph attention networks (GATs) have been proposed for anomaly detection tasks. GAT architectures have been proposed and investigated for anomaly localization in microwave imaging [175], video anomaly detection [53], time series anomaly detection [176,177], and BGP anomaly detection [90], among others. Through these endeavors it is evident that GATs have the potential to be effective in detecting and localizing anomalies in diverse datasets.

5. Discussion

Graph attention networks (GATs) represent an enhancement to graph convolutional networks (GCNs), a class of neural networks specifically designed for processing data represented as graphs. By incorporating attention mechanisms, GATs offer greater flexibility and parameterization for graph convolutions, allowing them to more effectively capture and leverage the structural intricacies of graph data. These networks can be categorized based on the type of attention mechanism employed, the techniques used to extract information from the graph structure, and the specific domain applications for which they have been optimized.

Over time, GATs have evolved into several specialized categories, each tailored to address distinct challenges in graph-structured data processing. Global GATs focus on capturing global features within a graph by using attention scores that are independent of the graph's edges. This approach enables them to identify and leverage broader patterns that span the entire graph, making them particularly effective in scenarios where a comprehensive global context is crucial. In contrast, Multi-Layer GATs enhance the learning of intricate, higher-level features by stacking multiple attention layers. Each layer aggregates information from neighboring nodes, progressively refining node features and enabling the network to capture more abstract and complex patterns from a larger neighborhood within the graph.

Graph-embedding GATs combine graph-embedding techniques with attention mechanisms to learn latent coordinates for each graph node, effectively representing the underlying graph structure. This category excels in capturing both local and global graph structures, making it versatile across various domains. Spatial GATs focus on capturing local graph information by integrating both spatial and temporal attention mechanisms. This capability makes them particularly effective in applications like video anomaly detection, where the spatial and temporal relationships between nodes are critical. Variational GATs introduce a probability-based regularization technique that constrains graph attention scores, thereby mitigating overfitting. This makes them highly effective in complex, multimodal data environments where overfitting is a significant concern. Finally, Hybrid GATs combine

multiple technologies from these different categories to perform optimally across diverse tasks, such as miRNA–disease connection discovery and session-based recommendations, offering a flexible and powerful framework for a wide range of applications.

Attention mechanisms in GATs can vary, encompassing dot product attention, additive attention, and even more advanced or customized attention models. The methods employed by GATs to obtain information from graph structures are diverse, involving edge-labeling approaches, aggregation functions, general integrators, softmax normalization, and edge masking.

In this paper, we emphasized five key categorizations: Global Attention Networks, which use attention scores independent of graph edges to capture global features; Multi-Layer GATs, which harness multiple attention layers to better capture intricate higher-level features; graph-embedding GATs, which utilize graph-embedding techniques to learn latent coordinates for each graph node, representing the underlying graph structure; spatial GATs, which select both spatial and temporal attention to capture local graph information; and Variational GATs, which introduce a probability-based regularization technique to constrain graph attention scores, thereby mitigating overfitting. We also include a hybrid category comprising methods that integrate more than one of the aforementioned categories (Table 4).

Graph attention networks (GATs), while based on a core architecture, exhibit substantial technical differences across various applications, adapting uniquely to the specific needs of each domain. In recommendation systems and biomarker–disease association, GATs are tailored to analyze interaction graphs, whether they be user–item in the former or biomolecular networks in the latter. The attention mechanism in these contexts is pivotal for identifying intricate patterns, such as user preferences or biomarker linkages to diseases. This is in stark contrast to their use in sentiment analysis and image analysis, where the focus shifts to natural language processing and spatial relationships, respectively. In sentiment analysis, GATs are designed to understand the semantic relationships in text, while, in image analysis, they concentrate on visual features and pixel connections.

Table 4. Overview of different GAT-based architectures, core concepts, and their advantages.

Model	Core Idea	Attention Mechanism	Advantages
Original GAT	Introduces attention mechanisms to graph neural networks (GNNs), allowing the model to learn the importance (attention weights) of neighboring nodes when aggregating information. The attention mechanism is applied to each pair of nodes and their edges.	Uses a single-layer feedforward neural network to compute attention scores, followed by a softmax function to normalize these scores.	Suitable for small to moderately sized graphs but can become computationally expensive for very large graphs due to the pairwise attention calculation.
Multi-Head Attention GAT	Extends the original GAT by using multiple attention mechanisms (heads) in parallel. This allows the model to capture more complex relationships by combining different attention heads.	Each head computes its own attention scores, and the outputs are either concatenated or averaged.	Improves the expressive power and stabilizes the learning process, making the model more robust to noise.
GATv2	An improvement over the original GAT that redefines the attention mechanism to make it more expressive and less sensitive to the order of node pairs.	Instead of computing attention scores as a single linear combination of features, GATv2 computes them using a more flexible approach that allows for asymmetric attention scores, which better captures complex node relationships.	Provides better performance on certain tasks, particularly where the direction of the edge plays a significant role.
Sparse GAT	A variation designed to handle large-scale graphs with many nodes and edges. Sparse GATs reduce the computational burden by focusing only on a subset of neighbors when computing attention, instead of all possible neighbors.	Often uses techniques like sampling or clustering to limit the number of neighbors considered during attention calculation.	Scalable to much larger graphs while maintaining reasonable performance, making them more practical for real-world applications like social networks or biological networks.
Hierarchical GAT (H-GAT)	Introduces a hierarchical structure to GATs, where attention is computed at multiple levels of graph granularity. This approach captures both local and global graph structures.	Combines attention scores at different hierarchical levels, allowing the model to learn from different scales of the graph.	Particularly useful for large and complex graphs, where both micro (local node connections) and macro (overall graph structure) views are important.
Temporal GAT	Adapts GATs for dynamic graphs where the structure evolves over time. It incorporates temporal information into the attention mechanism.	Combines traditional attention with time-aware mechanisms, such as temporal encoding or recurrent neural networks (RNNs), to handle the evolving nature of the graph.	Essential for applications like transaction networks, where the sequence and timing of interactions are crucial.
Edge-Weighted GAT	Incorporates edge weights directly into the attention mechanism, making the model more sensitive to the strength or significance of connections between nodes.	Modifies the attention computation to include edge weights, which influence the importance of neighboring nodes during information aggregation.	Useful for graphs where edges have varying levels of importance, such as in recommendation systems or weighted social networks.

Similarly, in anomaly detection, GATs are engineered to identify deviations from standard patterns, necessitating a heightened sensitivity to outliers, which differs significantly from their application in more pattern-consistent domains like recommendation systems or biomarker associations. Across all these domains, GATs are fine-tuned through variations in attention mechanisms, node feature representations, and integration with domain-specific data processing techniques, ensuring that their application is highly specialized and effective for the particular challenges and data characteristics of each field.

Recent advancements in the field of graph attention networks (GATs) have paved the way for their application in novel and diverse areas of research, showcasing their potential

to address complex challenges across various scientific domains. Notably, several groundbreaking studies published this year highlight the expanding utility of GATs, particularly in fields such as polymer science, drug discovery, and bioinformatics, where traditional methods have often fallen short.

One such study introduced GATBoost [178], a comprehensive framework designed to improve property prediction in polymers. This research demonstrated the power of GATs in mining important substructures within polymer graphs that are highly correlated with key properties such as the glass transition temperature (T_g). By integrating GATs with XGBoost-based supervised learning, the study achieved high accuracy in predictions, significantly enhancing the efficiency of property prediction processes and reducing experimental time. This approach not only delivers precise predictions but also offers direct visualization of the crucial polymer substructures, underscoring the interpretability of GAT models in material science.

In the realm of drug discovery, the AttentionMGT-DTA [179] model presents a multimodal attention-based approach for predicting drug–target affinity (DTA), a critical step in the drug design process. By utilizing GATs to represent molecular graphs and binding pocket graphs, and integrating two attention mechanisms to explore interactions between different protein modalities and drug–target pairs, this model has set new benchmarks in DTA prediction accuracy. The high interpretability of AttentionMGT-DTA, particularly in modeling interaction strengths between drug atoms and protein residues, further highlights the capability of GATs to enhance understanding and decision-making in drug development.

Another significant contribution this year comes from the bioinformatics domain with the ML-FGAT model [180], which addresses the prediction of multi-label protein subcellular localization (SCL). This model combines GATs with feature-generative adversarial networks and linear discriminant analysis to predict protein SCL with remarkable accuracy. The study emphasizes the robustness and interpretability of GATs, particularly through the analysis of attention weight parameters, and demonstrates their effectiveness across diverse datasets, including newly constructed ones such as human, virus, plant, and SARS-CoV-2.

These studies collectively underscore the rapidly growing interest in and application of GATs in cutting-edge research, reflecting their versatility and potential to revolutionize complex problem-solving across a wide range of scientific disciplines.

Despite the significant advancements and diverse applications of graph attention networks (GATs), several limitations and challenges remain that hinder their broader applicability and effectiveness in certain scenarios. One of the primary concerns lies in the scalability of GATs when applied to large-scale graphs. As the size of the graph increases, the computational and memory demands of attention mechanisms, especially those that require full pairwise attention computations, can become prohibitive. This issue is exacerbated when dealing with dense graphs, where the number of edges grows quadratically with the number of nodes, leading to an explosion in the number of attention coefficients that need to be computed and stored. Moreover, the attention mechanism's reliance on softmax normalization can introduce challenges related to numerical stability and gradient vanishing, particularly in deeper networks with multiple attention layers. These factors can limit the practical usability of GATs in large-scale, real-world applications, where computational efficiency and scalability are crucial.

Another significant limitation of GATs is their susceptibility to overfitting, especially when dealing with small or noisy datasets. The flexibility and expressiveness of attention mechanisms, while beneficial in capturing intricate patterns and relationships within the data, can also lead to the model learning spurious correlations that do not generalize well to unseen data. This issue is particularly pronounced in Variational GATs, where the introduction of probabilistic regularization techniques, although aimed at mitigating overfitting, may not always be sufficient in complex, multimodal environments. Additionally, the interpretability of the learned attention weights in GATs can be problematic. While the attention mechanism theoretically offers insights into the importance of different nodes or edges, in practice, the learned attention scores can be difficult to interpret, especially when

the network is deep or the graph structure is complex. This lack of interpretability poses challenges in applications where understanding the decision-making process is critical, such as in biomedical or financial domains.

This review provides a comprehensive and systematic exploration of the diverse graph attention network (GAT) techniques and their wide-ranging applications, offering a detailed understanding of this rapidly advancing field. By categorizing GATs based on their unique mechanisms and domain-specific optimizations, we aim to equip readers with the critical insights necessary to effectively navigate and apply these networks in their research. The analysis presented herein not only elucidates the current state of GAT methodologies but also underscores the significant potential of these networks to drive innovation in solving complex, graph-structured data problems. As GATs continue to evolve, understanding their nuanced technical variations and their implications across different domains will be essential for researchers and practitioners aiming to harness the full power of these sophisticated models.

Looking forward, graph attention networks hold immense potential for unlocking new avenues of research and application in emerging scientific fields. One promising area is quantum computing, where GATs could be employed to model and analyze quantum states and interactions, offering new insights into quantum information processing. Another exciting opportunity lies in the field of personalized medicine, where GATs could be used to integrate and analyze multi-omics data, including genomics, proteomics, and metabolomics, to predict patient-specific treatment responses and disease trajectories [181]. Additionally, the integration of GATs with natural language processing could lead to breakthroughs in understanding complex linguistic structures and enhancing machine translation systems. The ongoing development of GATs for real-time applications, such as autonomous systems and smart cities, also presents fertile ground for research, where the ability to process and interpret dynamic, large-scale graph data in real time could lead to significant advancements. As these networks continue to mature, the exploration of their applications in such novel and interdisciplinary fields will undoubtedly open up new research frontiers and drive further innovation.

Author Contributions: Conceptualization, S.K. and A.G.V.; methodology, S.K., A.G.V. and K.L.; investigation, S.K., A.G.V. and K.L.; writing—original draft preparation, A.G.V. and K.L.; writing—review and editing, S.K.; visualization, K.L.; supervision, S.K. and A.G.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

References

1. Barabási, A.L. *Network Science*; Cambridge University Press: Cambridge, UK, 2016.
2. Labonne, M. *Hands-On Graph Neural Networks Using Python*; Packt: Birmingham, UK, 2023.
3. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2016**, arXiv:1609.02907.
4. Abdel-Basset, M.; Moustafa, N.; Hawash, H.; Tari, Z. *Responsible Graph Neural Networks*; CRC Press: Boca Raton, FL, USA, 2023.
5. Liu, Y.; Yang, S.; Xu, Y.; Miao, C.; Wu, M.; Zhang, J. Contextualized Graph Attention Network for Recommendation With Item Knowledge Graph. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 181–195. [[CrossRef](#)]
6. Shan, Y.; Che, C.; Wei, X.; Wang, X.; Zhu, Y.; Jin, B. Bi-graph attention network for aspect category sentiment classification. *Knowl.-Based Syst.* **2022**, *258*, 109972. [[CrossRef](#)]
7. Yang, Z.; Liu, J.; Shah, H.A.; Feng, J. A novel hybrid framework for metabolic pathways prediction based on the graph attention network. *BMC Bioinform.* **2022**, *23*, 329. [[CrossRef](#)] [[PubMed](#)]
8. Fang, L.; Sun, T.; Wang, S.; Fan, H.; Li, J. A graph attention network for road marking classification from mobile LiDAR point clouds. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102735. [[CrossRef](#)]
9. Feng, S.; Xu, C.; Zuo, Y.; Chen, G.; Lin, F.; XiaHou, J. Relation-aware dynamic attributed graph attention network for stocks recommendation. *Pattern Recognit.* **2022**, *121*, 108119. [[CrossRef](#)]

10. Qin, C.; Zhang, Y.; Liu, Y.; Coleman, S.; Du, H.; Kerr, D. A visual place recognition approach using learnable feature map filtering and graph attention networks. *Neurocomputing* **2021**, *457*, 277–292. [[CrossRef](#)]
11. Li, K.; Feng, Y.; Gao, Y.; Qiu, J. Hierarchical graph attention networks for semi-supervised node classification. *Appl. Intell.* **2020**, *50*, 3441–3451. [[CrossRef](#)]
12. Rassil, A.; Chougrad, H.; Zouaki, H. Holistic Graph Neural Networks based on a global-based attention mechanism. *Knowl.-Based Syst.* **2022**, *240*, 108105. [[CrossRef](#)]
13. Hsu, Y.L.; Tsai, Y.C.; Li, C.T. FinGAT: Financial Graph Attention Networks for Recommending Top-KK Profitable Stocks. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 469–481. [[CrossRef](#)]
14. Ye, Y.; Ji, S. Sparse Graph Attention Networks. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 905–916. [[CrossRef](#)]
15. Xu, Y.; Fang, Y.; Huang, C.; Liu, Z. HGHAN: Hacker group identification based on heterogeneous graph attention network. *Inf. Sci.* **2022**, *612*, 848–863. [[CrossRef](#)]
16. Cao, R.; He, C.; Wei, P.; Su, Y.; Xia, J.; Zheng, C. Prediction of circRNA-Disease Associations Based on the Combination of Multi-Head Graph Attention Network and Graph Convolutional Network. *Biomolecules* **2022**, *12*, 932. [[CrossRef](#)]
17. Xie, Z.; Zhu, R.; Zhao, K.; Liu, J.; Zhou, G.; Huang, J.X. Dual Gated Graph Attention Networks with Dynamic Iterative Training for Cross-Lingual Entity Alignment. *ACM Trans. Inf. Syst.* **2021**, *40*, 1165. [[CrossRef](#)]
18. Mo, X.; Huang, Z.; Xing, Y.; Lv, C. Multi-Agent Trajectory Prediction With Heterogeneous Edge-Enhanced Graph Attention Network. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 9554–9567. [[CrossRef](#)]
19. Hu, B.; Guo, K.; Wang, X.; Zhang, J.; Zhou, D. RRL-GAT: Graph Attention Network-Driven Multilabel Image Robust Representation Learning. *IEEE Internet Things J.* **2022**, *9*, 9167–9178. [[CrossRef](#)]
20. Li, Z.; Zhong, T.; Huang, D.; You, Z.H.; Nie, R. Hierarchical graph attention network for miRNA-disease association prediction. *Mol. Ther.* **2022**, *30*, 1775–1786. [[CrossRef](#)] [[PubMed](#)]
21. Yan, H.; Wang, J.; Chen, J.; Liu, Z.; Feng, Y. Virtual sensor-based imputed graph attention network for anomaly detection of equipment with incomplete data. *J. Manuf. Syst.* **2022**, *63*, 52–63. [[CrossRef](#)]
22. Chen, L.; Cao, J.; Wang, Y.; Liang, W.; Zhu, G. Multi-view Graph Attention Network for Travel Recommendation. *Expert Syst. Appl.* **2022**, *191*, 116234. [[CrossRef](#)]
23. Buterez, D.; Bica, I.; Tariq, I.; Andrés-Terré, H.; Liò, P. CellVGAE: An unsupervised scRNA-seq analysis workflow with graph attention networks. *Bioinformatics* **2021**, *38*, 1277–1286. [[CrossRef](#)]
24. Hu, J.; Cao, L.; Li, T.; Dong, S.; Li, P. GAT-LI: A graph attention network based learning and interpreting method for functional brain network classification. *BMC Bioinform.* **2021**, *22*, 379. [[CrossRef](#)]
25. Shi, C.; Chen, X.; Huang, K.; Xiao, J.; Lu, H.; Stachniss, C. Keypoint Matching for Point Cloud Registration Using Multiplex Dynamic Graph Attention Networks. *IEEE Robot. Autom. Lett.* **2021**, *6*, 8221–8228. [[CrossRef](#)]
26. Huang, J.; Guan, L.; Su, Y.; Yao, H.; Guo, M.; Zhong, Z. A topology adaptive high-speed transient stability assessment scheme based on multi-graph attention network with residual structure. *Int. J. Electr. Power Energy Syst.* **2021**, *130*, 106948. [[CrossRef](#)]
27. Ji, C.; Wang, Y.; Ni, J.; Zheng, C.; Su, Y. Predicting miRNA-Disease Associations Based on Heterogeneous Graph Attention Networks. *Front. Genet.* **2021**, *12*, 727744. [[CrossRef](#)]
28. Yang, T.; Hu, L.; Shi, C.; Ji, H.; Li, X.; Nie, L. HGAT: Heterogeneous Graph Attention Networks for Semi-Supervised Short Text Classification. *ACM Trans. Inf. Syst.* **2021**, *39*, 32. [[CrossRef](#)]
29. Wang, P.; Wu, Q.; Cao, J.; Shen, C.; Gao, L.; Hengel, A.v.d. Neighbourhood Watch: Referring Expression Comprehension via Language-Guided Graph Attention Networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 1960–1968. [[CrossRef](#)]
30. Huang, J.; Luo, K.; Cao, L.; Wen, Y.; Zhong, S. Learning Multiaspect Traffic Couplings by Multirelational Graph Attention Networks for Traffic Prediction. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 20681–20695. [[CrossRef](#)]
31. Li, Z.; Zhao, Y.; Zhang, Y.; Zhang, Z. Multi-relational graph attention networks for knowledge graph completion. *Knowl.-Based Syst.* **2022**, *251*, 109262. [[CrossRef](#)]
32. Wang, W.; Chen, H. Predicting miRNA-disease associations based on graph attention networks and dual Laplacian regularized least squares. *Briefings Bioinform.* **2022**, *23*, bbac292. [[CrossRef](#)]
33. Zhao, K.; Liu, J.; Xu, Z.; Liu, X.; Xue, L.; Xie, Z.; Zhou, Y.; Wang, X. Graph4Web: A relation-aware graph attention network for web service classification. *J. Syst. Softw.* **2022**, *190*, 111324. [[CrossRef](#)]
34. Yuan, J.; Cao, M.; Cheng, H.; Yu, H.; Xie, J.; Wang, C. A unified structure learning framework for graph attention networks. *Neurocomputing* **2022**, *495*, 194–204. [[CrossRef](#)]
35. Shimizu, R.; Matsutani, M.; Goto, M. An explainable recommendation framework based on an improved knowledge graph attention network with massive volumes of side information. *Knowl.-Based Syst.* **2022**, *239*, 107970. [[CrossRef](#)]
36. Jiang, L.; Sun, J.; Wang, Y.; Ning, Q.; Luo, N.; Yin, M. Identifying drug–target interactions via heterogeneous graph attention networks combined with cross-modal similarities. *Briefings Bioinform.* **2022**, *23*, bbac016. [[CrossRef](#)] [[PubMed](#)]
37. Safai, A.; Vakharia, N.; Prasad, S.; Saini, J.; Shah, A.; Lenka, A.; Pal, P.K.; Ingahlalikar, M. Multimodal Brain Connectomics-Based Prediction of Parkinson’s Disease Using Graph Attention Networks. *Front. Neurosci.* **2022**, *15*, 741489. [[CrossRef](#)] [[PubMed](#)]
38. Zhao, Z.; Yang, B.; Li, G.; Liu, H.; Jin, Z. Precise Learning of Source Code Contextual Semantics via Hierarchical Dependence Structure and Graph Attention Networks. *J. Syst. Softw.* **2022**, *184*, 111108. [[CrossRef](#)]

39. Long, Y.; Zhang, Y.; Wu, M.; Peng, S.; Kwoh, C.K.; Luo, J.; Li, X. Heterogeneous graph attention networks for drug virus association prediction. *Methods* **2022**, *198*, 11–18. [[CrossRef](#)]
40. Lan, W.; Wu, X.; Chen, Q.; Peng, W.; Wang, J.; Chen, Y.P. GANLDA: Graph attention network for lncRNA-disease associations prediction. *Neurocomputing* **2022**, *469*, 384–393. [[CrossRef](#)]
41. Wang, X.; Liu, X.; Wu, H.; Liu, J.; Chen, X.; Xu, Z. Jointly learning invocations and descriptions for context-aware mashup tagging with graph attention network. *World Wide Web* **2022**, *26*, 1295–1322. [[CrossRef](#)]
42. Long, J.; Zhang, R.; Yang, Z.; Huang, Y.; Liu, Y.; Li, C. Self-Adaptation Graph Attention Network via Meta-Learning for Machinery Fault Diagnosis With Few Labeled Data. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–11. [[CrossRef](#)]
43. Jiang, N.; Jie, W.; Li, J.; Liu, X.; Jin, D. GATrust: A Multi-Aspect Graph Attention Network Model for Trust Assessment in OSNs. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 5865–5878. [[CrossRef](#)]
44. Zhou, Y.; Shen, J.; Zhang, X.; Yang, W.; Han, T.; Chen, T. Automatic source code summarization with graph attention networks. *J. Syst. Softw.* **2022**, *188*, 111257. [[CrossRef](#)]
45. Feng, Y.Y.; Yu, H.; Feng, Y.H.; Shi, J.Y. Directed graph attention networks for predicting asymmetric drug–drug interactions. *Briefings Bioinform.* **2022**, *23*, bbac151. [[CrossRef](#)]
46. Xu, D.; Alameda-Pineda, X.; Ouyang, W.; Ricci, E.; Wang, X.; Sebe, N. Probabilistic Graph Attention Network With Conditional Kernels for Pixel-Wise Prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 2673–2688. [[CrossRef](#)] [[PubMed](#)]
47. Zhang, Z.; Huang, J.; Tan, Q. Association Rules Enhanced Knowledge Graph Attention Network. *Knowl.-Based Syst.* **2022**, *239*, 108038. [[CrossRef](#)]
48. Lai, B.; Xu, J. Accurate protein function prediction via graph attention networks with predicted structure information. *Briefings Bioinform.* **2021**, *23*, bbab502. [[CrossRef](#)] [[PubMed](#)]
49. Lan, W.; Dong, Y.; Chen, Q.; Zheng, R.; Liu, J.; Pan, Y.; Chen, Y.P.P. KGANCDA: Predicting circRNA-disease associations based on knowledge graph attention network. *Briefings Bioinform.* **2021**, *23*, bbab494. [[CrossRef](#)]
50. Zhao, X.; Zhao, X.; Yin, M. Heterogeneous graph attention network based on meta-paths for lncRNA–disease association prediction. *Briefings Bioinform.* **2021**, *23*, bbab407. [[CrossRef](#)]
51. Zhao, Y.; Zhou, H.; Zhang, A.; Xie, R.; Li, Q.; Zhuang, F. Connecting Embeddings Based on Multiplex Relational Graph Attention Networks for Knowledge Graph Entity Typing. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 4608–4620. [[CrossRef](#)]
52. Yang, M.; Bai, X.; Wang, L.; Zhou, F. Mixed Loss Graph Attention Network for Few-Shot SAR Target Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [[CrossRef](#)]
53. Chen, H.; Mei, X.; Ma, Z.; Wu, X.; Wei, Y. Spatial–temporal graph attention network for video anomaly detection. *Image Vis. Comput.* **2023**, *131*, 104629. [[CrossRef](#)]
54. Gao, J.; Gao, J.; Ying, X.; Lu, M.; Wang, J. Higher-order Interaction Goes Neural: A Substructure Assembling Graph Attention Network for Graph Classification. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 1594–1608. [[CrossRef](#)]
55. Zhao, M.; Wang, L.; Jiang, Z.; Li, R.; Lu, X.; Hu, Z. Multi-task learning with graph attention networks for multi-domain task-oriented dialogue systems. *Knowl.-Based Syst.* **2023**, *259*, 110069. [[CrossRef](#)]
56. Guan, X.; Xing, W.; Li, J.; Wu, H. HGAT-VCA: Integrating high-order graph attention network with vector cellular automata for urban growth simulation. *Comput. Environ. Urban Syst.* **2023**, *99*, 101900. [[CrossRef](#)]
57. Yang, J.; Yang, L.T.; Wang, H.; Gao, Y. Multirelational Tensor Graph Attention Networks for Knowledge Fusion in Smart Enterprise Systems. *IEEE Trans. Ind. Inform.* **2023**, *19*, 616–625. [[CrossRef](#)]
58. Wang, L.; Zhong, C. gGATLDA: lncRNA-disease association prediction based on graph-level graph attention network. *BMC Bioinform.* **2022**, *23*, 11. [[CrossRef](#)] [[PubMed](#)]
59. Zhao, C.; Song, A.; Du, Y.; Yang, B. TrajGAT: A map-embedded graph attention network for real-time vehicle trajectory imputation of roadside perception. *Transp. Res. Part C: Emerg. Technol.* **2022**, *142*, 103787. [[CrossRef](#)]
60. He, J.; Cui, J.; Zhang, G.; Xue, M.; Chu, D.; Zhao, Y. Spatial–temporal seizure detection with graph attention network and bi-directional LSTM architecture. *Biomed. Signal Process. Control* **2022**, *78*, 103908. [[CrossRef](#)]
61. Zhang, X.; Xu, Y.; Shao, Y. Forecasting Traffic Flow with Spatial–Temporal Convolutional Graph Attention Networks. *Neural Comput. Appl.* **2022**, *34*, 15457–15479. [[CrossRef](#)]
62. Wang, M.; Wu, L.; Li, M.; Wu, D.; Shi, X.; Ma, C. Meta-learning based spatial-temporal graph attention network for traffic signal control. *Knowl.-Based Syst.* **2022**, *250*, 109166. [[CrossRef](#)]
63. Wang, Y.; Jing, C.; Xu, S.; Guo, T. Attention based spatiotemporal graph attention networks for traffic flow forecasting. *Inf. Sci.* **2022**, *607*, 869–883. [[CrossRef](#)]
64. Tang, H.; Wei, P.; Li, J.; Zheng, N. EvoSTGAT: Evolving spatiotemporal graph attention networks for pedestrian trajectory prediction. *Neurocomputing* **2022**, *491*, 333–342. [[CrossRef](#)]
65. Fang, M.; Tang, L.; Yang, X.; Chen, Y.; Li, C.; Li, Q. FTPG: A Fine-Grained Traffic Prediction Method With Graph Attention Network Using Big Trace Data. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 5163–5175. [[CrossRef](#)]
66. Kong, X.; Zhang, J.; Wei, X.; Xing, W.; Lu, W. Adaptive spatial-temporal graph attention networks for traffic flow forecasting. *Appl. Intell.* **2022**, *52*, 4300–4316. [[CrossRef](#)]
67. Yang, J.; Sun, X.; Wang, R.G.; Xue, L.X. PTPGC: Pedestrian trajectory prediction by graph attention network with ConvLSTM. *Robot. Auton. Syst.* **2022**, *148*, 103931. [[CrossRef](#)]

68. Gao, H.; Xiao, J.; Yin, Y.; Liu, T.; Shi, J. A Mutually Supervised Graph Attention Network for Few-Shot Segmentation: The Perspective of Fully Utilizing Limited Samples. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *35*, 4826–4838. [[CrossRef](#)] [[PubMed](#)]
69. Wang, Z.; Liu, C.; Gombolay, M. Heterogeneous graph attention networks for scalable multi-robot scheduling with temporospatial constraints. *Auton. Robot.* **2022**, *46*, 249–268. [[CrossRef](#)]
70. Tang, J.; Zeng, J. Spatiotemporal gated graph attention network for urban traffic flow prediction based on license plate recognition data. *Comput.-Aided Civ. Infrastruct. Eng.* **2022**, *37*, 3–23. [[CrossRef](#)]
71. Tian, S.; Kang, L.; Xing, X.; Tian, J.; Fan, C.; Zhang, Y. A Relation-Augmented Embedded Graph Attention Network for Remote Sensing Object Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [[CrossRef](#)]
72. Zhao, Z.; Wang, H.; Yu, X. Spectral–Spatial Graph Attention Network for Semisupervised Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
73. Jiang, T.; Sun, J.; Liu, S.; Zhang, X.; Wu, Q.; Wang, Y. Hierarchical semantic segmentation of urban scene point clouds via group proposal and graph attention network. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102626. [[CrossRef](#)]
74. Wang, Y.; Wang, H.; He, J.; Lu, W.; Gao, S. TAGAT: Type-Aware Graph Attention neTworks for reasoning over knowledge graphs. *Knowl.-Based Syst.* **2021**, *233*, 107500. [[CrossRef](#)]
75. Yu, X.; Shi, S.; Xu, L. A spatial–temporal graph attention network approach for air temperature forecasting. *Appl. Soft Comput.* **2021**, *113*, 107888. [[CrossRef](#)]
76. Zhang, Z.; Zhang, H.; Liu, S. Person re-identification using heterogeneous local graph attention networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12136–12145.
77. Mi, L.; Chen, Z. Hierarchical graph attention network for visual relationship detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 13886–13895.
78. Zhang, R.; Yao, W.; Shi, Z.; Zeng, L.; Tang, Y.; Wen, J. A graph attention networks-based model to distinguish the transient rotor angle instability and short-term voltage instability in power systems. *Int. J. Electr. Power Energy Syst.* **2022**, *137*, 107783. [[CrossRef](#)]
79. Hao, J.; Liu, J.; Pereira, E.; Liu, R.; Zhang, J.; Zhang, Y.; Yan, K.; Gong, Y.; Zheng, J.; Zhang, J.; et al. Uncertainty-guided graph attention network for paraneoplastic effusion diagnosis. *Med Image Anal.* **2022**, *75*, 102217. [[CrossRef](#)] [[PubMed](#)]
80. Yan, P.; Li, L.; Zeng, D. Quantum Probability-Inspired Graph Attention Network for Modeling Complex Text Interaction. *Knowl.-Based Syst.* **2021**, *234*, 107557. [[CrossRef](#)]
81. Yang, C.; Wang, P.; Tan, J.; Liu, Q.; Li, X. Autism spectrum disorder diagnosis using graph attention network based on spatial-constrained sparse functional brain networks. *Comput. Biol. Med.* **2021**, *139*, 104963. [[CrossRef](#)]
82. Zhao, Y.; Zhang, G.; Dong, C.; Yuan, Q.; Xu, F.; Zheng, Y. Graph Attention Network with Focal Loss for Seizure Detection on Electroencephalography Signals. *Int. J. Neural Syst.* **2021**, *31*, 2150027. [[CrossRef](#)] [[PubMed](#)]
83. Xie, Y.; Zhang, Y.; Gong, M.; Tang, Z.; Han, C. MGAT: Multi-view Graph Attention Networks. *Neural Networks* **2020**, *132*, 180–189. [[CrossRef](#)]
84. Tao, Z.; Wei, Y.; Wang, X.; He, X.; Huang, X.; Chua, T.S. MGAT: Multimodal Graph Attention Network for Recommendation. *Inf. Process. Manag.* **2020**, *57*, 102277. [[CrossRef](#)]
85. Li, G.; Fang, T.; Zhang, Y.; Liang, C.; Xiao, Q.; Luo, J. Predicting miRNA-disease associations based on graph attention network with multi-source information. *BMC Bioinform.* **2022**, *23*, 244. [[CrossRef](#)]
86. Cai, P.; Wang, H.; Sun, Y.; Liu, M. DQ-GAT: Towards Safe and Efficient Autonomous Driving With Deep Q-Learning and Graph Attention Networks. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 21102–21112. [[CrossRef](#)]
87. Dai, G.; Wang, X.; Zou, X.; Liu, C.; Cen, S. MRGAT: Multi-Relational Graph Attention Network for knowledge graph completion. *Neural Networks* **2022**, *154*, 234–245. [[CrossRef](#)] [[PubMed](#)]
88. Wang, Y.; Lu, L.; Wu, Y.; Chen, Y. Polymorphic graph attention network for Chinese NER. *Expert Syst. Appl.* **2022**, *203*, 117467. [[CrossRef](#)]
89. Zhang, X.; Ma, H.; Gao, Z.; Li, Z.; Chang, L. Exploiting cross-session information for knowledge-aware session-based recommendation via graph attention networks. *Int. J. Intell. Syst.* **2022**, *37*, 7614–7637. [[CrossRef](#)]
90. Peng, S.; Nie, J.; Shu, X.; Ruan, Z.; Wang, L.; Sheng, Y.; Xuan, Q. A multi-view framework for BGP anomaly detection via graph attention network. *Comput. Networks* **2022**, *214*, 109129. [[CrossRef](#)]
91. Yang, M.; Huang, Z.A.; Gu, W.; Han, K.; Pan, W.; Yang, X.; Zhu, Z. Prediction of biomarker–disease associations based on graph attention network and text representation. *Briefings Bioinform.* **2022**, *23*, bbac298. [[CrossRef](#)]
92. Baul, S.; Ahmed, K.T.; Filipek, J.; Zhang, W. omicsGAT: Graph Attention Network for Cancer Subtype Analyses. *Int. J. Mol. Sci.* **2022**, *23*, 10220. [[CrossRef](#)]
93. Liu, Z.; Ma, Y.; Cheng, Q.; Liu, Z. Finding Asymptomatic Spreaders in a COVID-19 Transmission Network by Graph Attention Networks. *Viruses* **2022**, *14*, 1659. [[CrossRef](#)]
94. Jiang, J.; Wang, T.; Wang, B.; Ma, L.; Guan, Y. Gated Tree-based Graph Attention Network (GTGAT) for medical knowledge graph reasoning. *Artif. Intell. Med.* **2022**, *130*, 102329. [[CrossRef](#)]
95. Jiang, X.; Wang, Y.; Fan, A.; Ma, J. Learning for mismatch removal via graph attention networks. *ISPRS J. Photogramm. Remote Sens.* **2022**, *190*, 181–195. [[CrossRef](#)]
96. Zhou, L.; Zhao, Y.; Yang, D.; Liu, J. GCHGAT: Pedestrian trajectory prediction using group constrained hierarchical graph attention networks. *Appl. Intell.* **2022**, *52*, 11434–11447. [[CrossRef](#)]

97. Inan, E. ZoKa: A fake news detection method using edge-weighted graph attention network with transfer models. *Neural Comput. Appl.* **2022**, *34*, 11669–11677. [[CrossRef](#)]
98. Yang, J.; Yang, L.T.; Wang, H.; Gao, Y.; Liu, H.; Xie, X. Tensor Graph Attention Network for Knowledge Reasoning in Internet of Things. *IEEE Internet Things J.* **2022**, *9*, 9128–9137. [[CrossRef](#)]
99. Li, F.; Feng, J.; Yan, H.; Jin, D.; Li, Y. Crowd Flow Prediction for Irregular Regions with Semantic Graph Attention Network. *ACM Trans. Intell. Syst. Technol.* **2022**, *13*, 81. [[CrossRef](#)]
100. Shi, Y.; Zhou, K.; Li, S.; Zhou, M.; Liu, W. Heterogeneous graph attention network for food safety risk prediction. *J. Food Eng.* **2022**, *323*, 111005. [[CrossRef](#)]
101. Karbalayghareh, A.; Sahin, M.; Leslie, C.S. Chromatin interaction-aware gene regulatory modeling with graph attention networks. *Genome Res.* **2022**, *32*, 930–944. [[CrossRef](#)]
102. Peng, Y.; Tan, G.; Si, H.; Li, J. DRL-GAT-SA: Deep reinforcement learning for autonomous driving planning based on graph attention networks and simplex architecture. *J. Syst. Archit.* **2022**, *126*, 102505. [[CrossRef](#)]
103. Yang, X.; Deng, C.; Liu, T.; Tao, D. Heterogeneous Graph Attention Network for Unsupervised Multiple-Target Domain Adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 1992–2003. [[CrossRef](#)]
104. Dong, Z.; Li, Z.; Yan, Y.; Calinon, S.; Chen, F. Passive Bimanual Skills Learning From Demonstration With Motion Graph Attention Networks. *IEEE Robot. Autom. Lett.* **2022**, *7*, 4917–4923. [[CrossRef](#)]
105. Mahbub, S.; Bayzid, M.S. EGRET: Edge aggregated graph attention networks and transfer learning improve protein–protein interaction site prediction. *Briefings Bioinform.* **2022**, *23*, bbab578. [[CrossRef](#)]
106. Tekbiyik, K.; Yurduseven, O.; Kurt, G.K. Graph Attention Network-Based Single-Pixel Compressive Direction of Arrival Estimation. *IEEE Commun. Lett.* **2022**, *26*, 562–566. [[CrossRef](#)]
107. Zhou, H.; Yang, Y.; Luo, T.; Zhang, J.; Li, S. A unified deep sparse graph attention network for scene graph generation. *Pattern Recognit.* **2022**, *123*, 108367. [[CrossRef](#)]
108. Zhang, D.; Liu, Z.; Jia, W.; Liu, H.; Tan, J. Path Enhanced Bidirectional Graph Attention Network for Quality Prediction in Multistage Manufacturing Process. *IEEE Trans. Ind. Inform.* **2022**, *18*, 1018–1027. [[CrossRef](#)]
109. Wu, H.; Zhang, Z.; Shi, S.; Wu, Q.; Song, H. Phrase dependency relational graph attention network for Aspect-based Sentiment Analysis. *Knowl.-Based Syst.* **2022**, *236*, 107736. [[CrossRef](#)]
110. Wang, S.; Wang, F.; Qiao, S.; Zhuang, Y.; Zhang, K.; Pang, S.; Nowak, R.; Lv, Z. MSHGANMDA: Meta-Subgraphs Heterogeneous Graph Attention Network for miRNA-Disease Association Prediction. *IEEE J. Biomed. Health Inform.* **2022**, *27*, 4639–4648. [[CrossRef](#)]
111. Wei, Q.; Li, Y.; Zhang, J.; Wang, F.Y. VGN: Value Decomposition With Graph Attention Networks for Multiagent Reinforcement Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *35*, 182–195 [[CrossRef](#)] [[PubMed](#)]
112. Wang, H.; Wang, Z.; Chen, J.; Liu, W. Graph Attention Network Model with Defined Applicability Domains for Screening PBT Chemicals. *Environ. Sci. Technol.* **2022**, *56*, 6774–6785. [[CrossRef](#)] [[PubMed](#)]
113. Chen, Y.; Yan, J.; Jiang, M.; Zhang, T.; Zhao, Z.; Zhao, W.; Zheng, J.; Yao, D.; Zhang, R.; Kendrick, K.M.; et al. Adversarial Learning Based Node-Edge Graph Attention Networks for Autism Spectrum Disorder Identification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *35*, 7275–7286. [[CrossRef](#)]
114. Dong, Y.; Liu, Q.; Du, B.; Zhang, L. Weighted Feature Fusion of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2022**, *31*, 1559–1572. [[CrossRef](#)]
115. Li, X.; Tan, J.; Wang, P.; Liu, H.; Li, Z.; Wang, W. Anatomically constrained squeeze-and-excitation graph attention network for cortical surface parcellation. *Comput. Biol. Med.* **2022**, *140*, 105113. [[CrossRef](#)]
116. Liu, S.; Duan, L.; Zhang, Z.; Cao, X.; Durrani, T.S. Ground-Based Remote Sensing Cloud Classification via Context Graph Attention Network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [[CrossRef](#)]
117. Xu, C.; Cai, L.; Gao, J. An efficient scRNA-seq dropout imputation method using graph attention network. *BMC Bioinform.* **2021**, *22*, 582. [[CrossRef](#)] [[PubMed](#)]
118. Yang, Y.; Walker, T.M.; Kouchaki, S.; Wang, C.; Peto, T.E.A.; Crook, D.W.; Consortium, C.; Clifton, D.A. An end-to-end heterogeneous graph attention network for Mycobacterium tuberculosis drug-resistance prediction. *Briefings Bioinform.* **2021**, *22*, bbab299. [[CrossRef](#)] [[PubMed](#)]
119. Zeng, J.; Liu, T.; Jia, W.; Zhou, J. Fine-grained Question-Answer sentiment classification with hierarchical graph attention network. *Neurocomputing* **2021**, *457*, 214–224. [[CrossRef](#)]
120. Shao, Y.; Li, R.; Hu, B.; Wu, Y.; Zhao, Z.; Zhang, H. Graph Attention Network-Based Multi-Agent Reinforcement Learning for Slicing Resource Management in Dense Cellular Network. *IEEE Trans. Veh. Technol.* **2021**, *70*, 10792–10803. [[CrossRef](#)]
121. Ji, C.; Liu, Z.; Wang, Y.; Ni, J.; Zheng, C. GATNNCDA: A method based on graph attention network and multi-layer neural network for predicting circRNA-disease associations. *Int. J. Mol. Sci.* **2021**, *22*, 8505. [[CrossRef](#)] [[PubMed](#)]
122. Sacha, M.; Błaz, M.; Byrski, P.; Dabrowski-Tumanski, P.; Chrominski, M.; Loska, R.; Włodarczyk-Pruszyński, P.; Jastrzebski, S. Molecule edit graph attention network: Modeling chemical reactions as sequences of graph edits. *J. Chem. Inf. Model.* **2021**, *61*, 3273–3284. [[CrossRef](#)]
123. Wang, Z.; Gombolay, M. Learning Scheduling Policies for Multi-Robot Coordination With Graph Attention Networks. *IEEE Robot. Autom. Lett.* **2020**, *5*, 4509–4516. [[CrossRef](#)]

124. Li, Q.; Lin, W.; Liu, Z.; Prorok, A. Message-Aware Graph Attention Networks for Large-Scale Multi-Robot Path Planning. *IEEE Robot. Autom. Lett.* **2021**, *6*, 5533–5540. [CrossRef]
125. Sha, A.; Wang, B.; Wu, X.; Zhang, L. Semisupervised Classification for Hyperspectral Images Using Graph Attention Networks. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 157–161. [CrossRef]
126. Long, Y.; Luo, J.; Zhang, Y.; Xia, Y. Predicting human microbe–disease associations via graph attention networks with inductive matrix completion. *Briefings Bioinform.* **2020**, *22*, bbaa146. [CrossRef]
127. Hei, Y.; Yang, R.; Peng, H.; Wang, L.; Xu, X.; Liu, J.; Liu, H.; Xu, J.; Sun, L. Hawk: Rapid Android Malware Detection Through Heterogeneous Graph Attention Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *35*, 4703–4717. [CrossRef] [PubMed]
128. Yan, Y.; Ren, W.; Hu, X.; Li, K.; Shen, H.; Cao, X. SRGAT: Single Image Super-Resolution With Graph Attention Network. *IEEE Trans. Image Process.* **2021**, *30*, 4905–4918. [CrossRef]
129. Tang, Y.; Zhang, X.; Zhai, Y.; Qin, G.; Song, D.; Huang, S.; Long, Z. Rotating Machine Systems Fault Diagnosis Using Semisupervised Conditional Random Field-Based Graph Attention Network. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [CrossRef]
130. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2018**, arXiv:1710.10903.
131. Fey, M.; Lenssen, J.E. Fast Graph Representation Learning with PyTorch Geometric. *arXiv* **2019**, arXiv:1903.02428.
132. Varuna Jayasiri, N.W. labml.ai Annotated Paper Implementations. 2020. Available online: <https://nn.labml.ai/> (accessed on 29 August 2024).
133. Wang, M.; Zheng, D.; Ye, Z.; Gan, Q.; Li, M.; Song, X.; Zhou, J.; Ma, C.; Yu, L.; Gai, Y.; et al. Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks. *arXiv* **2019**, arXiv:1909.01315.
134. Li, M.; Zhou, J.; Hu, J.; Fan, W.; Zhang, Y.; Gu, Y.; Karypis, G. DGL-LifeSci: An Open-Source Toolkit for Deep Learning on Graphs in Life Science. *ACS Omega* **2021**, *6*, 27233–27238. [CrossRef]
135. Zheng, D.; Song, X.; Ma, C.; Tan, Z.; Ye, Z.; Dong, J.; Xiong, H.; Zhang, Z.; Karypis, G. DGL-KE: Training Knowledge Graph Embeddings at Scale. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA, 25 July 2020; SIGIR; pp. 739–748.
136. Dwivedi, V.P.; Joshi, C.K.; Luu, A.T.; Laurent, T.; Bengio, Y.; Bresson, X. Benchmarking Graph Neural Networks. *arXiv* **2020**, arXiv:2003.00982.
137. Wu, L.; Chen, Y.; Shen, K.; Guo, X.; Gao, H.; Li, S.; Pei, J.; Long, B. Graph Neural Networks for Natural Language Processing: A Survey. *arXiv* **2021**, arXiv:2106.06090.
138. Jin, Z.; Wang, Y.; Wang, Q.; Ming, Y.; Ma, T.; Qu, H. Gnnlens: A visual analytics approach for prediction error diagnosis of graph neural networks. *IEEE Trans. Vis. Comput. Graph.* **2022**, *29*, 3024–3038. [CrossRef]
139. Leontis, N.B.; Zirbel, C.L. Nonredundant 3D structure datasets for RNA knowledge extraction and benchmarking. *RNA 3D Struct. Anal. Predict.* **2012**, *27*, 281–298.
140. Han, H.; Zhao, T.; Yang, C.; Zhang, H.; Liu, Y.; Wang, X.; Shi, C. Openhgnn: An open source toolkit for heterogeneous graph neural network. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, 17–22 October 2022; pp. 3993–3997.
141. Zhou, H.; Zheng, D.; Nisa, I.; Ioannidis, V.; Song, X.; Karypis, G. TGL: A General Framework for Temporal GNN Training on Billion-Scale Graphs. *Proc. VLDB Endow.* **2022**, *15*. [CrossRef]
142. Sammut, C.; Webb, G.I. *Encyclopedia of Machine Learning*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011.
143. Gao, C.; Zheng, Y.; Li, N.; Li, Y.; Qin, Y.; Piao, J.; Quan, Y.; Chang, J.; Jin, D.; He, X.; et al. A Survey of Graph Neural Networks for Recommender Systems: Challenges, Methods, and Directions. *ACM Trans. Recomm. Syst.* **2023**, *1*, 3. [CrossRef]
144. Ameen, T.; Ali, A.A. Graph Attention Network for Movie Recommendation. *Int. J. Intell. Eng. Syst.* **2022**, *15*, 49. [CrossRef]
145. Xu, A.; Zhong, P.; Kang, Y.; Duan, J.; Wang, A.; Lu, M.; Shi, C. THAN: Multimodal Transportation Recommendation With Heterogeneous Graph Attention Networks. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 1533–1543. [CrossRef]
146. Wang, C.; Ren, J.; Liang, H. MSGraph: Modeling multi-scale K-line sequences with graph attention network for profitable indices recommendation. *Electron. Res. Arch.* **2023**, *31*, 2626–2650. [CrossRef]
147. Jin, Y.; Ji, W.; Shi, Y.; Wang, X.; Yang, X. Meta-path guided graph attention network for explainable herb recommendation. *Health Inf. Sci. Syst.* **2023**, *11*, 5. [CrossRef]
148. Li, X.; Zhang, X.; Wang, P.; Cao, Z. Web Services Recommendation Based on Metapath-Guided Graph Attention Network. *J. Supercomput.* **2022**, *78*, 12621–12647. [CrossRef]
149. Xie, F.; Xu, Y.; Zheng, A.; Chen, L.; Zheng, Z. Service recommendation through graph attention network in heterogeneous information networks. *Int. J. Comput. Sci. Eng.* **2022**, *25*, 643. [CrossRef]
150. Lu, W.; Jiang, N.; Jin, D.; Chen, H.; Liu, X. Learning Distinct Relationship in Package Recommendation With Graph Attention Networks. *IEEE Trans. Comput. Soc. Syst.* **2022**, *10*, 3308–3320. [CrossRef]
151. Song, T.; Guo, F.; Jiang, H.; Ma, W.; Feng, Z.; Guo, L. HGAT-BR: Hyperedge-based graph attention network for basket recommendation. *Appl. Intell.* **2022**, *53*, 1435–1451. [CrossRef]

152. Kouhsar, M.; Kashaninia, E.; Mardani, B.; Rabiee, H.R. CircWalk: A novel approach to predict CircRNA-disease association based on heterogeneous network representation learning. *BMC Bioinform.* **2022**, *23*, 331. [[CrossRef](#)]
153. Aznaourova, M.; Schmerer, N.; Schmeck, B.; Schulte, L.N. Disease-Causing Mutations and Rearrangements in Long Non-coding RNA Gene Loci. *Front Genet* **2020**, *11*, 527484. [[CrossRef](#)] [[PubMed](#)]
154. Bartel, D.P. Metazoan MicroRNAs. *Cell* **2018**, *173*, 20–51. [[CrossRef](#)] [[PubMed](#)]
155. Ozata, D.M.; Gainetdinov, I.; Zoch, A.; O'Carroll, D.; Zamore, P.D. PIWI-interacting RNAs: Small RNAs with big functions. *Nat. Rev. Genet.* **2019**, *20*, 89–108. [[CrossRef](#)] [[PubMed](#)]
156. Peng, L.; Yang, C.; Chen, Y.; Liu, W. Predicting CircRNA-Disease associations via feature convolution learning with heterogeneous graph attention network. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 3072–3082. [[CrossRef](#)] [[PubMed](#)]
157. Zhao, X.; Wu, J.; Zhao, X.; Yin, M. Multi-view contrastive heterogeneous graph attention network for lncRNA–disease association prediction. *Briefings Bioinform.* **2022**, *24*, bbac548. [[CrossRef](#)]
158. Zhao, H.; Li, Z.; You, Z.H.; Nie, R.; Zhong, T. Predicting Mirna-Disease Associations Based on Neighbor Selection Graph Attention Networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2023**, *20*, 1298–1307. [[CrossRef](#)]
159. Zheng, K.; Zhang, X.L.; Wang, L.; You, Z.H.; Zhan, Z.H.; Li, H.Y. Line graph attention networks for predicting disease-associated Piwi-interacting RNAs. *Briefings Bioinform.* **2022**, *23*, bbac393. [[CrossRef](#)]
160. Dayun, L.; Junyi, L.; Yi, L.; Qihua, H.; Deng, L. MGATMDA: Predicting microbe-disease associations via multi-component graph attention network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *19*, 3578–3585. [[CrossRef](#)]
161. Lu, J.; Shi, L.; Liu, G.; Zhan, X. Dual-Channel Edge-Featured Graph Attention Networks for Aspect-Based Sentiment Analysis. *Electronics* **2023**, *12*, 624. [[CrossRef](#)]
162. Miao, Y.; Luo, R.; Zhu, L.; Liu, T.; Zhang, W.; Cai, G.; Zhou, M. Contextual Graph Attention Network for Aspect-Level Sentiment Classification. *Mathematics* **2022**, *10*, 2473. [[CrossRef](#)]
163. Wang, P.; Zhao, Z. Improving context and syntactic dependency for aspect-based sentiment analysis using a fused graph attention network. *Evol. Intell.* **2023**, *17*, 589–598. [[CrossRef](#)]
164. Wang, Y.; Yang, N.; Miao, D.; Chen, Q. Dual-channel and multi-granularity gated graph attention network for aspect-based sentiment analysis. *Appl. Intell.* **2022**, *53*, 13145–13157. [[CrossRef](#)]
165. Yuan, L.; Wang, J.; Yu, L.C.; Zhang, X. syntactic Graph Attention Network for Aspect-Level Sentiment Analysis. *IEEE Trans. Artif. Intell.* **2022**, *5*, 140–153. [[CrossRef](#)]
166. Zhang, X.; Yu, L.; Tian, S. BGAT: Aspect-based sentiment analysis based on bidirectional GRU and graph attention network. *J. Intell. Fuzzy Syst.* **2023**, *44*, 3115–3126. [[CrossRef](#)]
167. Zhou, X.; Zhang, T.; Cheng, C.; Song, S. Dynamic multichannel fusion mechanism based on a graph attention network and BERT for aspect-based sentiment classification. *Appl. Intell.* **2022**, *53*, 6800–6813. [[CrossRef](#)]
168. Leng, J.; Tang, X. Graph Attention Networks for Multiple Pairs of Entities and Aspects Sentiment Analysis in Long Texts. *J. Syst. Sci. Inf.* **2022**, *10*, 203–215. [[CrossRef](#)]
169. Xu, K.; Zhao, Y.; Zhang, L.; Gao, C.; Huang, H. Spectral–Spatial Residual Graph Attention Network for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
170. Liu, C.; Dong, Y. CNN-Enhanced graph attention network for hyperspectral image super-resolution using non-local self-similarity. *Int. J. Remote Sens.* **2022**, *43*, 4810–4835. [[CrossRef](#)]
171. Shen, W. A Novel Conditional Generative Adversarial Network Based On Graph Attention Network For Moving Image Denoising. *J. Appl. Sci. Eng.* **2022**, *26*, 829–839. [[CrossRef](#)]
172. Shuai, W.; Jiang, F.; Zheng, H.; Li, J. MSGATN: A Superpixel-Based Multi-Scale Siamese Graph Attention Network for Change Detection in Remote Sensing Images. *Appl. Sci.* **2022**, *12*, 5158. [[CrossRef](#)]
173. Zhou, W.; Xia, Z.; Dou, P.; Su, T.; Hu, H. Double Attention Based on Graph Attention Network for Image Multi-Label Classification. *ACM Trans. Multimed. Comput. Commun. Appl.* **2023**, *19*, 1–23. [[CrossRef](#)]
174. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly Detection: A Survey. *ACM Comput. Surv.* **2009**, *41*, 15. [[CrossRef](#)]
175. Al-Saffar, A.; Guo, L.; Abbosh, A. Graph Attention Network in Microwave Imaging for Anomaly Localization. *IEEE J. Electromagn. RF Microwaves Med. Biol.* **2022**, *6*, 212–218. [[CrossRef](#)]
176. Ding, C.; Sun, S.; Zhao, J. MST-GAT: A multimodal spatial–temporal graph attention network for time series anomaly detection. *Inf. Fusion* **2023**, *89*, 527–536. [[CrossRef](#)]
177. Zhou, L.; Zeng, Q.; Li, B. Hybrid Anomaly Detection via Multihead Dynamic Graph Attention Networks for Multivariate Time Series. *IEEE Access* **2022**, *10*, 40967–40978. [[CrossRef](#)]
178. Li, D.; Ru, Y.; Liu, J. GATBoost: Mining graph attention networks-based important substructures of polymers for a better property prediction. *Mater. Today Commun.* **2024**, *38*, 107577. [[CrossRef](#)]
179. Wu, H.; Liu, J.; Jiang, T.; Zou, Q.; Qi, S.; Cui, Z.; Tiwari, P.; Ding, Y. AttentionMGT-DTA: A multi-modal drug-target affinity prediction using graph transformer and attention mechanism. *Neural Networks* **2024**, *169*, 623–636. [[CrossRef](#)]

-
180. Wang, C.; Wang, Y.; Ding, P.; Li, S.; Yu, X.; Yu, B. ML-FGAT: Identification of multi-label protein subcellular localization by interpretable graph attention networks and feature-generative adversarial networks. *Comput. Biol. Med.* **2024**, *170*, 107944. [[CrossRef](#)]
 181. Liao, Y.; Zhang, X.M.; Ferrie, C. Graph Neural Networks on Quantum Computers. *arXiv* **2024**, arXiv:2405.17060.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.