Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Review

A survey of large language models for data challenges in graphs

Mengran Li [a], Pengyu Zhang [b], Wenbin Xing [a], Yijia Zheng [b], Klim Zaporojets [c], Junzhou Chen [a], Ronghui Zhang [a,*], Yong Zhang [d], Siyuan Gong [e], Jia Hu [f], Xiaolei Ma [g], Zhiyuan Liu [h], Paul Groth [b], Marcel Worring [b]

[a] *Guangdong Key Laboratory of Intelligent Transportation System, School of Intelligent Systems Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen, 518107, Guangdong, China*
[b] *University of Amsterdam, Amsterdam, The Netherlands*
[c] *Aarhus University, Aarhus, Denmark*
[d] *Beijing Institute of Artificial Intelligence, Beijing University of Technology, Beijing, 100124, China*
[e] *School of Information and Engineering, Chang'an University, Xi'an, 710064, Shaanxi, China*
[f] *Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai, 201804, China*
[g] *Key Laboratory of Intelligent Transportation Technology and System, School of Transportation Science and Engineering, Beihang University, Beijing, 100191, China*
[h] *Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Nanjing, 210096, Jiangsu, China*

ARTICLE INFO

ABSTRACT

Graphs are a widely used paradigm for representing non-Euclidean data, with applications ranging from social network analysis to biomolecular prediction. While graph learning has achieved remarkable progress, real-world graph data presents a number of challenges that significantly hinder the learning process. In this survey, we focus on four fundamental data-centric challenges: **(1) Incompleteness**, real-world graphs have missing nodes, edges, or attributes; **(2) Imbalance**, the distribution of the labels of nodes or edges and their structures for real-world graphs are highly skewed; **(3) Cross-domain Heterogeneity**, graphs from different domains exhibit incompatible feature spaces or structural patterns; and **(4) Dynamic Instability**, graphs evolve over time in unpredictable ways. Recently, Large Language Models (LLMs) offer the potential to tackle these challenges by leveraging rich semantic reasoning and external knowledge. This survey focuses on how LLMs can address four fundamental data-centric challenges in graph-structured data, thereby improving the effectiveness of graph learning. For each challenge, we review both traditional solutions and modern LLM-driven approaches, highlighting how LLMs contribute unique advantages. Finally, we discuss open research questions and promising future directions in this emerging interdisciplinary field. To support further exploration, we have curated a repository of recent advances on graph learning challenges: https://github.com/limengran98/Awesome-Literature-Graph-Learning-Challenges.

1. Introduction

Graphs provide a flexible and expressive framework for modeling complex systems composed of interconnected entities. Unlike grid-structured data such as images or sequences, graphs encode non-Euclidean topologies with irregular neighborhood structures and varying connectivity patterns. To support learning on such structures, graph learning aims to derive low-dimensional node or graph-level representations that preserve both structural dependencies and semantic information from the original graph (Ju et al., 2024a; Khoshraftar & An, 2024; Li et al., 2025c; Xia et al., 2021). Due to these representational advantages, graph learning has been successfully applied across a wide range of domains, including social network analysis (Yang et al., 2024a), personalized recommendation (Wang et al., 2023b), transportation optimization (Rahmani et al., 2023), financial modeling (Motie & Raahemi, 2024), and bioinformatics (Xie et al., 2025; Yang et al., 2024b). As a result, it has become a foundational technique for extracting knowledge from relational data and enabling models to reason over structured domains.
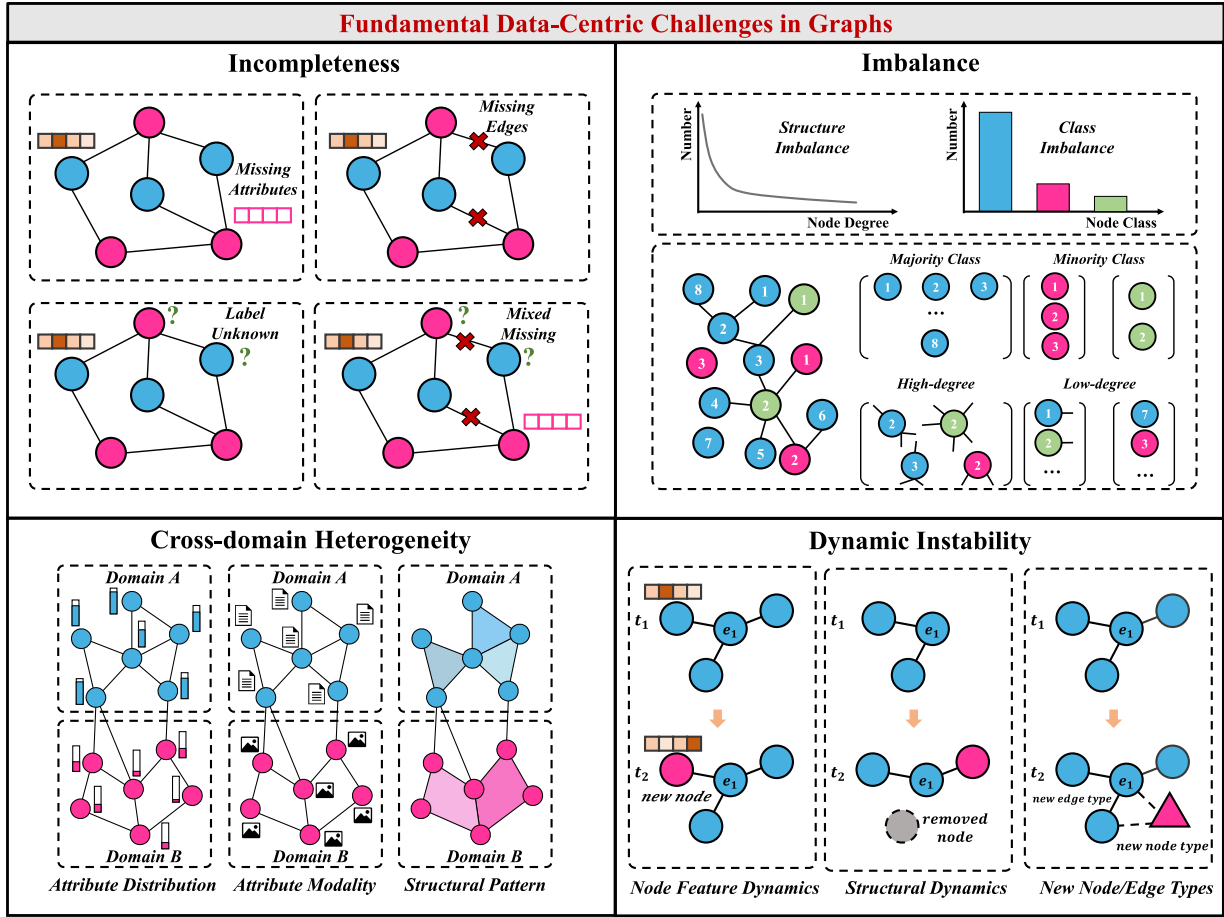

Fig. 1. The four fundamental challenges emerge of real-world graph complexity: (1) Incompleteness in graphs, where nodes, edges, or attributes are missing, (2) Imbalance in graphs, where the distribution of nodes, edges, or labels is highly skewed, (3) Cross-domain heterogeneity in graphs, where graph data from different domains exhibit semantic and structural discrepancies, and (4) Dynamic instability in graphs, where graphs undergo dynamic changes in topology, attributes, or interactions over time.

Despite this progress, real-world graph data often exhibits complexities, including sparsity, noise, domain divergence, and dynamic evolution, introducing practical difficulties for graph learning (Ju et al., 2024c). For instance, in social networks, cold-start users may lack sufficient links or profile features, making it difficult to learn reliable representations (Wang et al., 2023b). In fraud detection, the data is often highly imbalanced, where fraudulent behavior accounts for only a small fraction of all activities, and labels are sparse and noisy (Rajput & Singh, 2022). In smart cities, traffic systems involve heterogeneous and dynamically evolving road networks, posing difficulties for models to adapt to changing topologies in real time (Li et al., 2024b; Rahmani et al., 2023; Zhang et al., 2025b). As summarized in Fig. 1, these concrete scenarios highlight four **fundamental data-centric challenges** in real-world graphs:

(1) Incompleteness in Graphs An incomplete graph refers to a graph where some node or edge information is missing, resulting in an incomplete representation of the structure or attributes that fails to fully reflect the complexity of the real-world system (Chen et al., 2022). This challenge typically arises from incomplete data collection, missing link information, or outdated knowledge. For example, in social networks, missing user attributes-such as age, location, or interests-can obscure community boundaries, making it difficult to accurately detect user groups or recommend content. Similarly, missing connections between users may conceal important social ties, weakening the structural signals used for clustering or influence analysis. This incompleteness undermines the effectiveness of graph learning.

(2) Imbalance in Graphs A graph is considered imbalanced when certain categories or subgraphs contain significantly more nodes or edges than others (Liu et al., 2023d). This can lead to some categories dominating the graph, while others are left underrepresented (Wang et al., 2022). For example, financial transaction networks are characterized by containing a large number of legitimate transactions, and only a small fraction of fraudulent ones (Liu et al., 2021). Imbalanced training data may cause models to overfit to the majority class, thus impacting the model's ability to generalize and recognize the minority class effectively (Ma et al., 2023). This data imbalance challenge affects many graph learning methods, especially those relying on graph convolutional networks (Shi et al., 2020).

(3) Cross-Domain Heterogeneity in Graphs Cross-domain heterogeneity refers to significant disparities in attributes and structural patterns that arise when graph data are constructed from multiple source domains, especially when these domains have different data modalities or exhibit substantial distribution shifts. For instance, road network data collected from various cities may exhibit considerable heterogeneity, with some cities having grid-like structural patterns, while others exhibit radial configurations (Badhrudeen et al., 2022). This issue becomes even more critical in the rapidly growing field of graph foundation models (Liu et al., 2025), which aim to develop generalizable models capable of handling diverse graphs from different domains, such as molecular graphs and citation networks. Cross-domain heterogeneity poses additional challenges for graph learning, as diverse attributes and structural patterns cannot be easily processed in a unified manner

(Yan et al., 2024; Zhang et al., 2019c). Additionally, domain discrepancies may distort essential semantic and structural signals, complicating the identification of transferable features and limiting the effectiveness of graph learning methods (Hassani, 2022; Zhu et al., 2021).

(4) Dynamic Instability in Graphs Dynamic instability in graphs refers to continuous and unpredictable changes in node features and relationships over time. These may include updates to node attributes, alterations in graph structure, or the emergence of new node and edge types. Structural changes often involve adding or removing connections or reorganizing how nodes interact. For example, in a large-scale knowledge graph like Wikidata, an entity initially labeled as an "athlete" might later become classified as a "coach". This transition entails significant shifts in both node features and their connections within the graph (Zhang et al., 2024b,c). Furthermore, these continuous transformations introduce instability, making it difficult for graph learning models to adapt effectively. As a result, performance may degrade in tasks such as node classification, link prediction, and node representation learning (Kazemi et al., 2020). A deeper understanding of these forms of dynamic instability can facilitate the development of more robust and adaptive graph learning methods, crucial for accurately modeling the evolving real-world graphs.

A variety of graph learning methods have been developed to address these challenges, achieving notable success in specific contexts. For instance, graph completion algorithms (Cai et al., 2010) and graph autoencoders (Chen et al., 2022) have been employed to handle missing data; techniques like GraphSMOTE (Zhao et al., 2021) and re-weighted loss functions (Zhang et al., 2022) are commonly used to mitigate data imbalance; and graph domain adaptation methods, including adversarial regularization (Shen et al., 2020; Zhang et al., 2019c) and distribution alignment (Liu et al., 2023b; Zhu et al., 2021), have been explored to address cross-domain heterogeneity. Additionally, dynamic graph models such as EvolveGCN (Pareja et al., 2020) and DyRep (c)Trivedi, Farajtabar) have been proposed to model evolving graph structures. While these methods are effective within their respective scopes, they often rely on task-specific designs and require significant domain expertise.

In contrast, the emergence of large language models (LLMs) has opened up new opportunities for mitigating these fundamental challenges (Cui et al., 2025; Jeong et al., 2025; Ong et al., 2025). With their strong representational capacity, contextual reasoning abilities, and generalization potential, LLMs can extract rich semantic patterns from heterogeneous and noisy data. These capabilities enable LLMs to augment graph data by inferring missing information (Brasoveanu et al., 2023), synthesizing data for underrepresented classes (Yu et al., 2025), aligning heterogeneous attributes (Liu & Wu, 2023), and capturing temporal evolution or structural changes in dynamic graphs (Margatina et al., 2023). By addressing these issues, LLMs offer a promising complement to traditional graph learning approaches. To reflect this growing interest, Fig. 2 shows publication trends over the past decade (up to July 2025) related to "Graph Learning" and "LLM-based Graph Learning". The statistics indicate that not only has graph learning become an increasingly important topic, but the proportion of research exploring the integration of LLMs into graph learning has also been steadily rising.

This survey provides a comprehensive review of current research at the intersection of graph learning and large language models, centered on the four fundamental challenges introduced above. To ensure broad coverage, we performed an extensive literature search across major databases (arXiv, ACM Digital Library, IEEE Xplore, Elsevier, and Springer, etc.) using keyword combinations such as "LLMs", "graph learning", "incompleteness", "imbalance", "heterogeneity", "dynamic graphs". We focus on recent publications (primarily 2015-2025) in top machine learning, data mining, and NLP venues, as well as relevant preprints. This survey includes papers that meet at least one of the following criteria: (1) the work explicitly addresses one of the four fundamental challenges in graph learning, or (2) the proposed solution incorporates LLMs as a central component. Based on these criteria, we initially collected and reviewed over 1000 relevant papers. We then
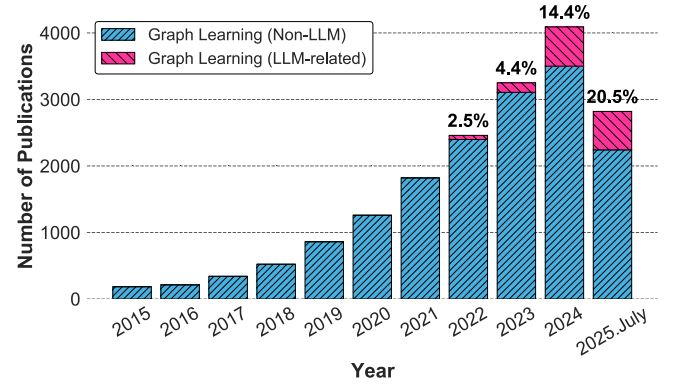

Fig. 2. Annual publication trends in Graph Learning from January 2015 to July 2025, with a highlighted subset of works related to LLMs. The blue bars represent Graph Learning (non-LLM) publications and the pink segments represent LLM-related publications; the stacked height indicates the total per year. The percentages above each bar denote the proportion of LLM-related papers within the total for that year. Data sourced from Google Scholar (https://scholar.google.com).

categorized them according to the type of challenge addressed and the strategies employed, ultimately selecting a curated set of over 380 representative works. For each challenge, we briefly discuss conventional (non-LLM) techniques to provide background context, while our primary focus is to highlight the novel contributions and advantages introduced by LLM-based approaches.

The structure of the paper is organized as follows: Section 2 provides an analysis and summary of traditional methods and LLM-based approaches for handling graph learning challenges. Section 3, building on the existing body of research, proposes new technical paths and solutions. Section 4 discusses the potential future directions for graph learning research. Finally, Section 5 concludes the paper.

2. Overview of graph learning and LLMs

2.1. Graph notation definition

We denote a graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X}_V, \mathbf{X}_E)$, where $\mathcal{V}$ is the set of n nodes, $\mathcal{E}$ is the set of edges, $\mathbf{X}_V \in \mathbb{R}^{|\mathcal{V}| \times d_v}$ is the node feature matrix, $\mathbf{X}_E \in \mathbb{R}^{|\mathcal{E}| \times d_e}$ is the edge feature matrix, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the adjacency matrix, where $\mathbf{A}_{ij} = 1$ if $(v_i, v_j) \in \mathcal{E}$. The degree matrix is $\mathbf{D}$ with $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$. The normalized graph Laplacian is defined as: $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$.

The types of graph data involved in this work can be broadly categorized as follows. Homogeneous graph refers to a graph in which all nodes and edges belong to the same type. Heterogeneous graph refers to a graph where nodes and/or edges belong to multiple types, such as user/item nodes or different edge relations. Attributed graph refers to a graph where nodes and/or edges are associated with real-valued features, *i.e.*, $\mathbf{X}_V$ and/or $\mathbf{X}_E$ are non-empty. Knowledge graph refers to a directed, multi-relational graph represented as triplets (h, r, o), where $h \in \mathcal{V}$ is the head entity, $o \in \mathcal{V}$ is the object entity, and $r \in \mathcal{R}$ is a relation type. Dynamic graph refers to a sequence of time-evolving graphs $\{\mathcal{G}_t\}_{t=1}^T$ in which the node set, edge set, or features may change over time.

2.2. Graph learning methods

Graph learning refers to a family of machine learning techniques designed specifically to analyze and learn from graph-structured data (Khoshraftar & An, 2024; Li et al., 2023c; Xia et al., 2021). Unlike data in Euclidean domains, graphs encode relationships between entities, and thus graph learning algorithms must leverage both node features and
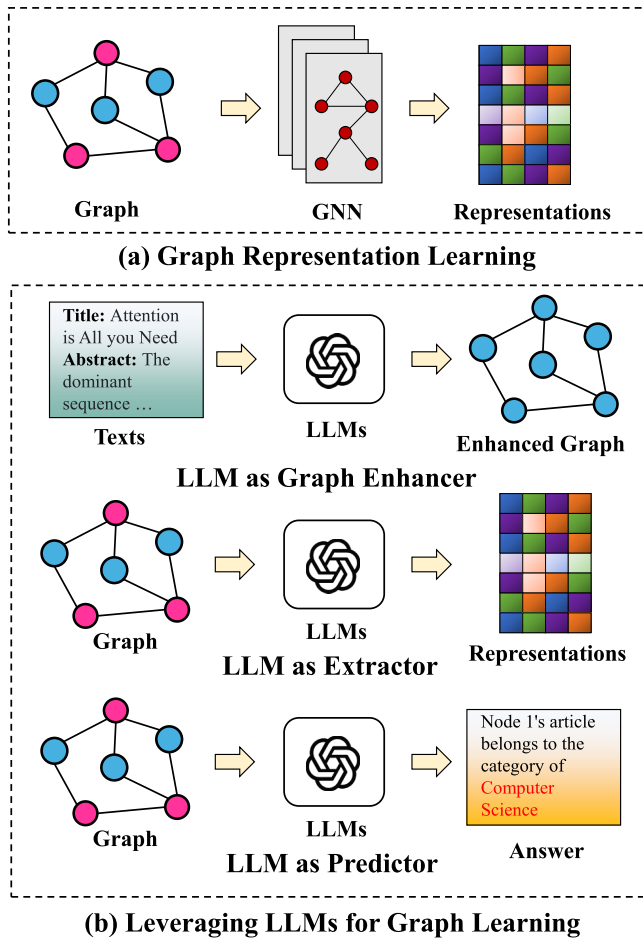
Fig. 3. Comparison between traditional graph representation learning and emerging paradigms that leverage LLMs. (a) Conventional graph learning relies on GNNs to encode structural information into node or graph-level representations. (b) In LLM-based approaches, LLMs can enhance graph structures using auxiliary text (top), extract semantic representations directly from graph inputs (middle), or generate task-specific answers through reasoning over graph structures (bottom).

the graph topology. The goal is often to learn vector representations for nodes, edges, or entire subgraphs/graphs that capture their structural role and feature information, so that these embeddings can be used for downstream tasks, as shown in Fig. 3(a).

Traditional Graph Learning Methods: A wide range of traditional techniques have been developed to learn graph representations. Early methods such as spectral clustering (Ng et al., 2001) and LINE (Tang et al., 2015) learn embeddings by factorizing adjacency or similarity matrices, capturing global connectivity patterns. Random-walk-based approaches like DeepWalk (Perozzi et al., 2014), Node2Vec (Grover & Leskovec, 2016), and metapath2vec (Dong et al., 2017) model local contexts through node sequences generated by truncated walks, offering better scalability and flexibility. In parallel, autoencoder-based models aim to reconstruct graph structures from latent representations. Notable examples include Graph Autoencoder (GAE) and Variational Graph Autoencoder (VGAE) (Kipf & Welling, 2016), which adopt graph convolutional encoders with deterministic or probabilistic decoders. Further developments such as Adversarially Regularized Graph Autoencoders (ARGA) (Pan et al., 2018) and Graph Attention Autoencoders (GATE) (Salehi & Davulcu, 2019) introduce adversarial training and attention mechanisms to improve expressiveness and robustness. Despite their foundational role, traditional methods often fall short in handling the complexity of real-world graph data.

Graph Neural Networks: The advent of graph neural networks (GNNs) marked a significant change by enabling end-to-end learning directly on graph structures. Spectral GNNs, starting with Bruna et al. (2014), were initially limited by high computational costs, which were mitigated by subsequent works using polynomial approximations (Defferrard et al., 2016). The introduction of Graph Convolutional Networks (GCNs) by Kipf and Welling (2017) popularized neighborhood aggregation in semi-supervised settings. Further developments such as Graph-SAGE (Hamilton et al., 2017), GAT (Veličković et al., 2018), SGC (Wu et al., 2019), and GIN (Xu et al., 2019) improved the expressiveness, scalability, and adaptability of GNNs.

More recently, two powerful approaches have emerged in graph learning: contrastive learning and Transformer-based architectures. Graph contrastive learning methods like DGI (Velickovic et al., 2019), GraphCL (You et al., 2020), and HeCo (Wang et al., 2021c) focus on learning robust representations by maximizing consistency across multiple views of the same graph. Transformer-inspired models, such as GT (Shi et al., 2021) and GraphGPS (Rampášek et al., 2022), introduce global self-attention and hybrid message-passing to model long-range dependencies and scale across graph types.

While graph learning techniques have shown remarkable success in modeling relational structures and topological dependencies, they often fall short in capturing high-level semantics and leveraging external unstructured data (Ren et al., 2024b; Yang et al., 2024d). These limitations become more pronounced when dealing with real-world graphs that are frequently incomplete, imbalanced, heterogeneous across domains, or dynamically evolving over time. These challenges have motivated a growing interest in integrating graph models with powerful language-based representations, such as LLMs, which offer strong semantic reasoning capabilities and the flexibility to incorporate external knowledge into the graph learning pipeline.

2.3. LLMs for graph learning

LLMs are a class of deep neural networks pretrained on massive text corpora to capture the statistical and semantic structure of natural language (Chang et al., 2024; Naveed et al., 2023). By learning distributed token representations and contextual relationships, LLMs achieve impressive performance across a broad spectrum of NLP tasks, including text generation (Li et al., 2024a), question answering (Singhal et al., 2025), and machine translation (Moslem et al., 2023).

The development of LLMs reflects a paradigm shift from task-specific pipelines to general-purpose language understanding. Early NLP relied on symbolic or statistical models (*e.g.*, HMMs (Rabiner, 1989), CRFs (Lafferty et al., 2001)), which were constrained by domain specificity and limited context. The introduction of the Transformer architecture (Vaswani et al., 2017) enabled global attention and deep sequence modeling, which laid the foundation for modern LLMs. Subsequent milestones such as BERT (Devlin et al., 2019), GPT-3 (Brown et al., 2020), and ChatGPT (Ouyang et al., 2022) progressively advanced context modeling, few-shot reasoning, and conversational alignment. Recent systems like GPT-4 (Achiam et al., 2023) and DeepSeek-MoE (Dai et al., 2024) further scale capacity and introduce modular expert routing, enabling more efficient and multimodal learning.

Despite being originally designed for natural language, LLMs exhibit emergent capabilities-such as abstract reasoning, flexible generalization, and heterogeneous input integration-that extend far beyond textual tasks. As shown in Fig. 3(b), LLMs can complement graph learning by generating structural hypotheses (Zhang et al., 2024g), enriching semantic content (Brasoveanu et al., 2023), and aligning cross-domain knowledge (Liu & Wu, 2023). Viewed through the lens of graph learning, these abilities make LLMs well-suited for enhancing data quality and robustness in graph-based applications:

Semantic reasoning: By leveraging rich contextual understanding, LLMs can infer latent or missing information from incomplete inputs. In the graph domain, this enables the imputation of missing node or edge
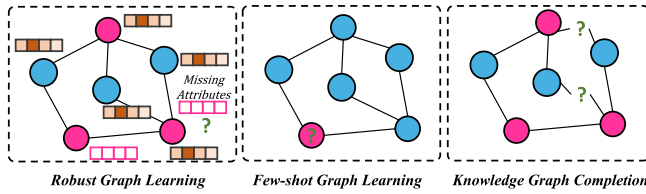
Fig. 4. Illustrations of three representative tasks addressing graph incompleteness. Robust Graph Learning (left) handles noisy or missing attributes to ensure stable performance. Few-shot Graph Learning (middle) aims to generalize from limited labeled nodes or subgraphs. Knowledge Graph Completion (right) focuses on inferring missing links between entities based on observed relationships.

attributes, the prediction of potential links, and the interpretation of sparsely connected substructures. Such capabilities directly address the challenge of **incompleteness** in real-world graphs, where data sparsity or noise is common (Brasoveanu et al., 2023).

Cross-domain and cross-modal alignment: LLMs are trained on diverse sources of knowledge and can serve as translators between incompatible domains or feature spaces. This makes them particularly useful in **cross-domain heterogeneous** graphs (Liu & Wu, 2023)-for example, aligning product graphs across e-commerce platforms that differ in attribute schemas, category taxonomies, and language conventions. LLMs can bridge such gaps by reasoning over textual descriptions and contextual cues to identify semantically equivalent items across domains.

Generative augmentation: As generative models, LLMs can synthesize new graph components-such as nodes, edges, or subgraphs-conditioned on textual or structural prompts. This ability can be used to create balanced training samples for underrepresented classes, thereby mitigating **data imbalance** (Yu et al., 2025), or to simulate hypothetical future graph states for handling **dynamic instability** (Zhang et al., 2024g).

In the following sections, we examine each of the four fundamental data-centric challenges in turn, analyzing how LLMs have been used to mitigate them, and what open questions remain at this intersection.

3. Current techniques for graph challenges

In this section, we discuss each of the four core challenges in detail. For each challenge, we first formalize the problem. We then review representative traditional and LLM-enhanced solutions to provide context.

3.1. Incompleteness in graphs

In the real world, graph data is often incomplete due to missing node attributes, edges or labels, which significantly impacts the accuracy and generalization ability of graph models. By leveraging LLMs for graph data completion and the inference of missing information, the performance of graph learning models can be significantly improved, even in scenarios with incomplete or missing information. We categorize current approaches to incomplete graphs into three major directions (as shown in Fig. 4): (1) Robust Graph Learning, which focuses on making models resilient to missing data; (2) Few-Shot Graph Learning, which addresses scenarios with extremely limited labeled data or structure by transferring knowledge; and (3) Knowledge Graph Completion, which is a prominent sub-area dealing with inferring missing facts in large knowledge graphs. The relevant references and categorization are presented in Table 1.

3.1.1. Robust graph learning

Robust graph learning methods aim to maintain model performance even when the input graph is noisy or partially observed. Traditional GNNs often degrade in accuracy if key attributes are missing or if the graph is sparsely connected, because the iterative message passing has

less information to propagate. Therefore, researchers have proposed specialized techniques to make graph learning robust to incompleteness.

(1) Traditional Methods for Robust Graph Learning Early efforts in robust graph learning focused on imputing missing values using global interpolation techniques under low-rank assumptions. For example, matrix completion methods (Cai et al., 2010; Chen et al., 2019a; Huang et al., 2019; Şimşek & Jensen, 2008) estimate missing attributes based on the global structure of the data. GraphRNA (Huang et al., 2019) introduces Attribute Random Walk (AttriWalk), modeling attributes as a bipartite graph to capture attribute-structure interactions, thereby enhancing local adaptability and robustness in heterogeneous environments. ARWMF (Chen et al., 2019a) combines random walks with matrix factorization and mutual information to improve unsupervised node embedding under noisy conditions.

With the rise of deep learning, more sophisticated generative models were developed to overcome the linear assumptions of early methods (Chen et al., 2025, 2022; Li et al., 2025a, 2024c,d; Morales-Alvarez et al., 2022; Peng et al., 2024b; Tu et al., 2022; Xia et al., 2024a; Yoo et al., 2022). SAT (Chen et al., 2022) enhances robustness by decoupling structure and attribute signals via shared latent spaces and distribution alignment. SVGA (Yoo et al., 2022) imposes strong probabilistic regularization on latent variables to prevent overfitting when features are sparse or unreliable. Models like ITR (Tu et al., 2022) and VISL (Morales-Alvarez et al., 2022) progressively refine latent variables or inter-variable structures using topological cues, improving resilience against structural noise. More recently, methods like MATE (Peng et al., 2024b) and AIAE (Xia et al., 2024a) exploit multi-view and multi-scale generation strategies to stabilize representation learning under incomplete or corrupted input. CSAT (Li et al., 2024d) incorporates contrastive learning and Transformers to detect communities under noisy or weak supervision, further enhancing robustness through auxiliary signals.

Another line of work improves model robustness through dynamic inference mechanisms that rely on structural priors like homophily and community consistency (Li et al., 2025b; Rossi et al., 2022; Um et al., 2023). For instance, FP (Rossi et al., 2022) integrates Dirichlet energy minimization with graph diffusion to achieve stable feature recovery. PCFI (Um et al., 2023) assigns confidence scores to feature channels, allowing uncertainty-aware propagation. ARB (Li et al., 2025b) addresses the cold-start problem by introducing virtual edges and redefined boundary conditions to improve propagation in sparse or poorly connected graphs. A notable advantage of such propagation-based methods is their parameter-free nature, making them highly adaptable and easily integrated with LLM-based pipelines for downstream reasoning tasks. To tackle the cold-start problem in node representation learning, Cold Brew (Zheng et al., 2021) introduced a feature contribution ratio metric to guide teacher-student distillation, uniquely encoding topological sparsity as a dynamic temperature coefficient. CTAug (Wu et al., 2024b) employs subgraph-aware contrastive learning to preserve dense subgraph priors, significantly enhancing representation learning in dense graphs.

These methods significantly enhance the predictive capabilities of graph data in scenarios with missing attributes by jointly modeling, employing variational inference, and optimizing multi-source information. Nevertheless, three challenges remain: (1) coupling optimization of completion and prediction dramatically increases training complexity; (2) latent-space completion methods are sensitive to prior assumptions; (3) efficient incremental mechanisms are lacking for dynamic attribute completion.

(2) LLM-enhanced methods Recently, researchers have begun exploring the deep integration of semantic reasoning and incomplete graph learning. Methods based on LLMs (Chai et al., 2023; Chen et al., 2023c; Fang et al., 2025; Sun et al., 2023; Wang et al., 2024c; Yu et al., 2025) pioneer a new paradigm that combines semantic completion and structural refinement, leveraging the strong generative and reasoning capabilities of LLMs.

Table 1

LLM-based methods for handling **incompleteness** in graphs, grouped by domain and incomplete type, with representative methods, datasets, metrics, and downstream tasks.

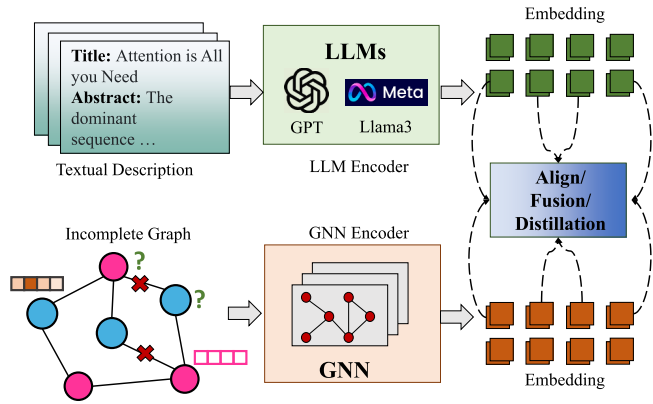| Domain	Incompleteness	Method	Typical Datasets	Common Metrics	Downstream Tasks
Robust Graph Learning	Node	LLM4NG (Yu et al., 2025)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), ogbn-arxiv (Hu et al., 2020)	Accuracy	Node Classification
		LLM-TAG (Sun et al., 2023)	Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), PubMed (Prithviraj et al., 2008), Arxiv-2023 (He et al., 2024a)	Accuracy	Node Classification
		SPLLM (Fang et al., 2025)	PeMS03 (Zhang et al., 2021), PeMS04 (Zhang et al., 2021), PeMS07 (Zhang et al., 2021)	MAE, RMSE, MAPE	Spatiotemporal Forecasting
	Label	LLMGNN (Chen et al., 2023c)	Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), PubMed (Prithviraj et al., 2008), ogbn-arxiv (Hu et al., 2020), ogbn-products (Hu et al., 2020), Wikics (Mernyei & Cangea, 2020)	Accuracy	Node Classification
	Mixed	GraphLLM (Chai et al., 2023)	Synthetic Data (Wang et al., 2023a)	Exact Match Accuracy	Graph Reasoning
		PRO-LINK (Wang et al., 2024c)	FB15k237 (Toutanova & Chen, 2015), Wikidata68K (Gesese et al., 2022), NELL-995 (Xiong et al., 2017)	MRR, Hits@N	Knowledge Graph Completion
		UnIMP (Wang et al., 2025)	BG (Asuncion et al., 2007), ZO (Asuncion et al., 2007), PK (Asuncion et al., 2007), BK (Asuncion et al., 2007), CS (Asuncion et al., 2007), ST (Asuncion et al., 2007), PW (Asuncion et al., 2007), BY (Asuncion et al., 2007), RR (Asuncion et al., 2007), WM (Asuncion et al., 2007)	RMSE, MAE	Data Imputation
Few-Shot Graph Learning	Structure	LinkGPT (He et al., 2024b)	AmazonSports (McAuley et al., 2015), Amazon-Clothing (McAuley et al., 2015), MAG-Geology (Sinha et al., 2015), MAG-Math (Sinha et al., 2015)	MRR, Hits@N	Link Prediction
		Anoma-lyLLM (Liu et al., 2024d)	UCI Messages (Opsahl & Panzarasa, 2009), Blogcatalog (Tang & Liu, 2009), T-Finance (Tang et al., 2022), T-Social (Tang et al., 2022)	AUC	Anomaly Detection
	Mixed	LLMDGCN (Li et al., 2024h)	Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), PubMed (Prithviraj et al., 2008), Religion (Rosenthal & McKeown, 2016)	Accuracy	Node Classification
		HeGTa (Jin et al., 2024)	IM-TQA (Zheng et al., 2023a), WCC (Ghasemi-Gol & Szekely, 2018), HiTab (Cheng et al., 2021), WTQ (Pasupat & Liang, 2015), TabFact (Chen et al., 2019b)	Macro-F1, Accuracy	Table Understanding
		FlexKBQA (Li et al., 2024i)	GrailQA (Gu et al., 2021), WebQSP (Yih et al., 2016), KQA Pro(Cao et al., 2022)	Exact Match, F1, Accuracy	Knowledge Graph Question Answering
Knowledge Graph Completion	Node	KGQG (Zhao et al., 2024a)	WebQuestions (Kumar et al., 2019), PathQuestions (Zhou et al., 2018)	BLEU-4, ROUGE-L, Hits@N	Knowledge Graph Question Answering
		LLM-KGC (b)Sehwag, Papasotiriou)	ILPC (Galkin et al., 2022)	MRR, Hits@N	Knowledge Graph Completion
		GS-KGC (Yang et al., 2025)	WN18RR (Dettmers et al., 2018), FB15k-237 (Toutanova & Chen, 2015), FB15k-237N (Lv et al., 2022), ICEWS14 (García-Durán et al., 2018), ICEWS05-15 (Li et al., 2021b)	Hits@N	Knowledge Graph Completion
		GLTW (Luo et al., 2025)	FB15k-237 (Toutanova & Chen, 2015), WN18RR (Dettmers et al., 2018), Wikidata5M (Vrandečić & Krötzsch, 2014)	MRR, Hits@N	Link Prediction
	Label	KGs-LLM (Carta et al., 2023)	Wikipedia Carta et al. (2023)	F1, Precision, Recall	Knowledge Graph Generation
	Mixed	FSKG (Brasoveanu et al., 2023)	WN18RR Dettmers et al. (2018), FB15k-237 (Toutanova & Chen, 2015)	MRR, Hits@N	Knowledge Graph Completion
		KGLLM (Yao et al., 2023)	WN11 (Socher et al., 2013), FB13 (Socher et al., 2013), WN18RR (Dettmers et al., 2018), YAGO3-10 (Dettmers et al., 2018)	Accuracy, MRR, Hits@N	Link Prediction, Knowledge Graph Completion
		KICGPT (Wei et al., 2023)	FB15k-237 (Toutanova & Chen, 2015), WN18RR (Dettmers et al., 2018)	MRR, Hits@N	Link Prediction
		RL-LLM (Chen et al., 2023a)	Electronics, Instacart (Chen et al., 2023a)	Precision, Recall, Accuracy	Knowledge Graph Completion
		GoG (Xu et al., 2024b)	Synthetic Data (Xu et al., 2024b)	Hits@N	Knowledge Graph Question Answering
		KoPA (Zhang et al., 2024f)	UMLS (Yao et al., 2019), CoDeX-S (Lv et al., 2022), FB15K-237N (Lv et al., 2022)	F1, Precision, Recall, Accuracy	Knowledge Graph Completion
		LLMKG (Iga & Silaghi, 2024)	Templates Easy (Iga & Silaghi, 2024), Templates Hard (Iga & Silaghi, 2024)	Strict Metrics, Flexible Metrics	Knowledge Graph Completion
		DIFT (Liu et al., 2024e)	WN18RR (Dettmers et al., 2018), FB15k-237 (Toutanova & Chen, 2015)	MRR, Hits@N	Link Prediction, Knowledge Graph Completion
		CP-KGC (Yang et al., 2024c)	WN18RR (Dettmers et al., 2018), FB15k-237 (Toutanova & Chen, 2015), UMLS (Yao et al., 2019)	MRR, Hits@N	Knowledge Graph Completion
		MuKDC (Li et al., 2024f)	NELL (Mitchell et al., 2018), Wiki (Vrandečić & Krötzsch, 2014)	MRR, Hits@N	Knowledge Graph Completion

Fig. 5. The framework addresses incomplete graph structures by integrating structural signals from GNNs and semantic priors from LLMs. The "Align / Fusion / Distillation" module enables the model to compensate for missing graph information by leveraging textual context, resulting in enriched and unified node attribute/edge/label representations.

Fig. 5 illustrates how LLMs can extract domain-specific knowledge and use it to compensate for missing nodes, edges, or attributes. For example, Sun et al. (2023) demonstrate that LLMs can infer missing nodes and edges by recognizing semantic similarities between entities. In this way, the model does not just "guess" a connection statistically, but proposes links that are meaningful in the real world. Building on this idea, LLM4NG (Yu et al., 2025) generates new, contextually relevant nodes, which is especially helpful in few-shot settings where the graph is very sparse. Both methods show that LLMs can enrich incomplete graphs more effectively than traditional imputation. Meanwhile, SPLLM (Fang et al., 2025) focuses on traffic sensor networks, where missing values break both spatial and temporal dependencies. By combining spatiotemporal GCNs with LLM fine-tuning, and incorporating external knowledge such as road maps or weather data, it achieves more reliable predictions. Similarly, UnIMP (Wang et al., 2025) works on tabular data by treating it as a hypergraph and using LLMs to fill in missing values, even when the data is semantically or structurally heterogeneous.

To further close the semantic gap introduced by graph incompleteness, several works propose end-to-end frameworks combining LLMs and graph models. For instance, GraphLLM (Chai et al., 2023) introduces a unified architecture where a graph encoder maps node features into the LLM's semantic space, enabling it to reason over incomplete or noisy node attributes using attention-gated fusion of structure and context. This helps, for example, in citation networks or recommendation systems where node metadata may be sparse or outdated. PRO-LINK (Wang et al., 2024c) enhances inductive reasoning on low-resource knowledge graphs by generating structural prompts and using LLMs to fill in gaps from limited graph and text data - particularly valuable in biomedical or emerging domains with sparse curation. Lastly, LLMGNN (Chen et al., 2023c) proposes a semi-supervised approach where LLMs annotate a small subset of nodes, helping GNNs generalize from minimal labeled data - addressing the classic label incompleteness challenge in large-scale graphs.

3.1.2. Few-shot graph learning

Few-shot learning (FSL) deals with scenarios where very few labeled examples are available for training, or where a graph has only a handful of nodes/edges in a particular class or subset of interest. This is related to incompleteness in that labels can be viewed as a type of missing information.

(1) Traditional Methods for Few-Shot Learning

FSL in graph domains aims to generalize effectively from limited labeled samples, addressing not only label scarcity but also structural issues such as missing links and long-tail relations (Qian et al., 2021; Sheng et al., 2020; Xiong et al., 2018; Yao et al., 2020; Zhang et al.,

2020). Early approaches primarily focused on explicitly modeling local graph structures and relational semantics under low-resource settings. For instance, Gmatching (Xiong et al., 2018) pioneered the integration of single-hop neighborhood information with relation prototype matching, providing a foundation for rare relation prediction. Building on this, GFL (Yao et al., 2020) established a transferable metric space by projecting auxiliary knowledge into the target domain, improving cross-graph adaptability.

Subsequent works moved toward more sophisticated meta-learning strategies. FSRL (Zhang et al., 2020) introduced a heterogeneous relation meta-learning framework that decouples relation semantics through a reference representation generator, enabling better handling of heterogeneous relation types. Complementing this semantic focus, FAAN (Sheng et al., 2020) designed an adaptive encoder-aggregator mechanism to dynamically weight entity-reference contributions, thereby producing fine-grained semantic representations that significantly enhance few-shot knowledge graph completion. Extending FSL techniques to real-world applications, MetaHG (Qian et al., 2021) demonstrated the utility of meta-learning in heterogeneous environments by detecting drug trafficking on Instagram.

Despite these advances, traditional FSL methods remain constrained by hand-crafted architecture designs, often lacking efficiency when dealing with large, complex graph structures and exhibiting limited capability for structured and interpretable reasoning.

(2) Few-Shot Learning in LLMs Few-shot learning has emerged as a natural strength of LLMs, enabling them to handle graph-related tasks with limited labeled data by leveraging rich prior knowledge from pre-training. Recent work demonstrates their effectiveness in diverse scenarios such as graph classification (Li et al., 2024h), link prediction (He et al., 2024b), anomaly detection (Liu et al., 2024d), table understanding (Jin et al., 2024), and knowledge graph question answering (KGQA) (Li et al., 2024i; Zhao et al., 2024a). For instance, LLMDGCN (Li et al., 2024h) incorporates degree-aware prompt tuning with graph-encoded positional embeddings, enabling both node classification and edge recovery in low-label settings. In link prediction, LinkGPT (He et al., 2024b) combines instruction tuning with retrieval-based re-ranking to enhance reasoning under sparse supervision. AnomalyLLM (Liu et al., 2024d) integrates dynamic-aware encoding and prototype-based edge reprogramming for improved detection in dynamic graphs. Beyond pure graph tasks, HeGTa (Jin et al., 2024) aligns table semantics with LLM knowledge via heterogeneous graph-enhanced soft prompts, while FlexKBQA (Li et al., 2024i) and a zero-shot KGQA approach (Zhao et al., 2024a) reformulate KB queries into natural language to generate synthetic training data or reduce reliance on full supervision. Collectively, these results indicate that LLMs can generalize to novel graph tasks with minimal examples, acting as strong priors that bridge missing or noisy information.

In summary, while few-shot graph learning remains challenging, LLMs act as few-shot learners by nature - they can leverage their pretrained knowledge to make sense of novel tasks with limited data. Results show that even with minimal graph information, an LLM can propose connections or classifications that align with human knowledge, essentially providing a powerful prior to the graph model.

3.1.3. Knowledge graph completion

Incompleteness is inherent in knowledge graphs: no knowledge base is ever complete. The development of knowledge graph completion (KGC) revolves around the core challenge of structure-semantics fusion, forming a progressive innovation trajectory from fundamental approach exploration to multimodal collaboration. LLMs, with their strong language understanding and reasoning capabilities, demonstrate great potential in addressing long-tail challenges, reducing annotation burdens, and handling incomplete graph data (Brasoveanu et al., 2023; Carta et al., 2023; Chen et al., 2023a; Iga & Silaghi, 2024; Li et al., 2024f; Liu et al., 2024e; b)Sehwag, Papasotiriou; Wei et al., 2023; Xu et al., 2024b; Yang et al., 2024c, 2025; Yao et al., 2023; Zhang et al., 2024f).

Early explorations of LLM-driven KGC primarily focused on converting triples into natural language sequences to verify whether generative LLMs could address incompleteness effectively. FSKG (Brasoveanu et al., 2023) exemplifies this trend by introducing a generation-refinement pipeline that mitigates long-tail sparsity through staged generation. Building on this concept, MuKDC incorporated multi-stage knowledge distillation to generate coherent supplementary triples, further improving the coverage of long-tail relations. Extending beyond generation quality, KGs-LLM (Carta et al., 2023) demonstrated that zero-shot prompting, even without external knowledge, can iteratively extract graph components, thereby reducing annotation costs and scaling efficiently to domain-specific datasets. While these methods validated the feasibility of text-based KGC, KGLLM (Yao et al., 2023) further showed that such reformulation enables lightweight fine-tuning of smaller LLMs like LLaMA-7B, achieving competitive triple classification and relation prediction. Addressing the persistent challenge of long-tail entities, KICGPT (Wei et al., 2023) integrated a triple-based retriever with contextual prompting, directly encoding structural cues into LLM inputs. This approach avoids additional training while yielding strong few-shot performance, thus highlighting the efficiency of retrieval-augmented prompting in KGC.

To mitigate the structural information loss inherent in purely text-based approaches, researchers have developed prompt encoding strategies that embed richer graph context. RL-LLM (Chen et al., 2023a) applied few-shot learning with multi-prompt optimization for relation prediction, showing that even minimal labeled data can yield competitive results in e-commerce KGs. Expanding the reasoning process, GoG (Xu et al., 2024b) proposed a "think-search-generate" pipeline, enabling the synthesis of new triples without additional training, which strengthens reasoning over incomplete graphs. Complementing this, LLM-KGC (b)Sehwag, Papasotiriou) fused ontology and graph structure directly into prompts, aligning topological and semantic information to improve inductive reasoning. Further enhancing structure-aware inference, KoPA (Zhang et al., 2024f) introduced a knowledge prefix adapter to inject structural signals during LLM reasoning, thereby improving logical consistency in KGC predictions. These works collectively highlight a shift from simple text reformulation toward structurally enriched prompting, aiming to bridge the gap between LLM semantic priors and graph topology.

Building upon these prompting and structural integration advances, recent research has optimized LLM-graph fusion for few-shot KGC, entity alignment, and robustness in heterogeneous settings. MuKDC (Li et al., 2024f) extended its earlier generation framework by incorporating multimodal knowledge and consistency evaluation, boosting performance on long-tail few-shot tasks. In parallel, LLMKG (Iga & Silaghi, 2024) demonstrated that carefully engineered prompts allow both Mixtral-8x7b and GPT-4 to perform competitively in zero- and one-shot KGC across diverse metrics. Addressing entity alignment, DIFT (Liu et al., 2024e) combined lightweight models with discriminative prompts to prevent alignment drift while improving overall KGC accuracy. In handling semantic ambiguity, CP-KGC (Yang et al., 2024c) leveraged contextually adaptive prompts to disambiguate polysemous entities, particularly in quantized LLMs. Moving toward deeper structural integration, GLTW (Luo et al., 2025) fused Graph Transformers with LLMs to jointly capture local and global patterns, while GS-KGC (Yang et al., 2025) employed subgraph-based question answering with negative sampling to enhance missing triple identification. Together, these frameworks signal a transition from isolated prompt design to holistic, multimodal, and structure-aware LLM-graph systems.

Overall, LLMs bring in a semantic understanding of the KG elements, which is valuable because KGs often have textual labels for entities/relations that carry meaning. LLMs leverage this to make more informed predictions, especially when data is sparse or the pattern is not purely structural but semantic.

3.1.4. Evaluation for handling incompleteness in graphs

We summarize the existing evaluation pipeline of LLM-based methods for incomplete graph learning, covering benchmark datasets, evaluation metrics, and downstream tasks (see Table 1).

Across the reviewed methods, commonly used benchmark datasets include Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), PubMed (Prithviraj et al., 2008), and ogbn-arxiv (Hu et al., 2020) for node classification tasks. These are citation networks where each node is a document with associated textual attributes. For large-scale knowledge graph tasks, datasets such as FB15k-237 (Toutanova & Chen, 2015), WN18RR (Dettmers et al., 2018), Wikidata5M (Vrandečić & Krötzsch, 2014), NELL (Mitchell et al., 2018), and UMLS (Yao et al., 2019) are widely adopted for link prediction and knowledge graph completion. Additionally, domain-specific or task-specific datasets are used to evaluate generalizability, including Amazon-Clothing and Amazon-Sports (McAuley et al., 2015) for recommendation and link prediction, PeMS03/04/07 (Zhang et al., 2021) for spatiotemporal forecasting, TQA/WTQ/TabFact (Chen et al., 2019b; Pasupat & Liang, 2015; Zheng et al., 2023a) for table understanding, and GrailQA/WebQSP/KQA Pro (Cao et al., 2022; Gu et al., 2021; Yih et al., 2016) for knowledge graph question answering. Some works also utilize synthetic graphs (Wang et al., 2023a; Xu et al., 2024b) to test reasoning capabilities in controlled environments.

A variety of evaluation metrics are used depending on the task. For classification, Accuracy, Macro-F1, and Micro-F1 are most common. Knowledge graph-related tasks report Mean Reciprocal Rank (MRR) and Hits@N as standard ranking metrics. Spatiotemporal forecasting relies on MAE, RMSE, and MAPE, while generative tasks such as question answering are evaluated using Exact Match, BLEU-4, and ROUGE-L.

The surveyed models address a wide range of downstream tasks, including node classification, link prediction, knowledge graph completion, anomaly detection, graph reasoning, question answering over knowledge graphs, and spatiotemporal forecasting. This diversity highlights the versatility of LLMs in handling incomplete graph data across domains and tasks.

3.1.5. Summary of incompleteness

Across robust learning, few-shot learning, and KGC, the common thread is that LLMs serve as knowledge-infusers and intelligent guessers for what is missing in the graph. They help create a more complete picture by either filling in data directly or guiding the graph model on where to look. Empirically, studies report that incorporating LLM-generated features or suggestions leads to substantial gains in tasks like node classification with missing features, link prediction with sparse edges, and KG completion benchmarks. As LLMs continue to improve, especially in domain-specific knowledge, we expect their role in handling graph incompleteness to grow.

3.2. Imbalance in graphs

Graph data often exhibits imbalance that can severely affect the performance, particularly in tasks such as node classification (as shown in Fig. 6). This imbalance typically complicates the learning process as models tend to be biased toward the majority class or low-degree nodes, resulting in poor generalization for underrepresented elements. To address this, techniques such as resampling, re-weighting, and graph-based regularization have been proposed to alleviate the effects of data imbalance. However, these approaches often require domain-specific adjustments and may not always yield optimal results.

Recent advances in LLMs offer promising solutions to the challenges posed by imbalanced graph data. LLMs, with their ability to process and generate semantically rich representations, can be leveraged to enrich the graph's feature space, enabling more nuanced and balanced learning.
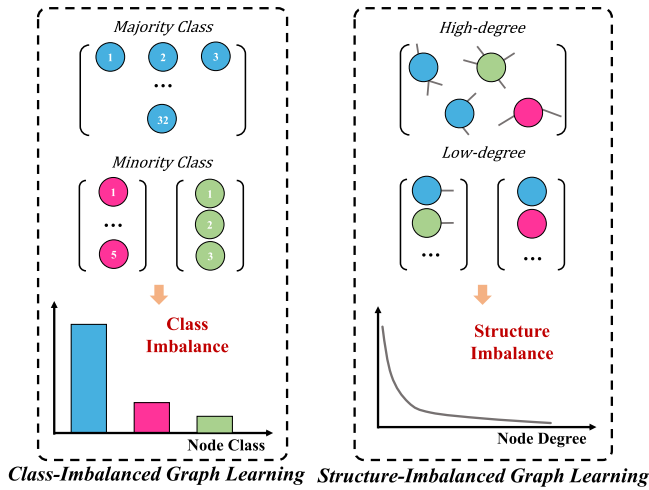
Fig. 6. Illustration of three representative tasks addressing graph imbalance. Class-imbalanced graph learning (left) refers to scenarios where certain node classes (minority classes) are underrepresented compared to others. Structure-imbalanced graph learning (right) highlights disparities in node connectivity, where high-degree nodes dominate information flow while low-degree nodes are underexposed.

This section reviews the application of traditional methods and LLMs for learning from imbalanced graph data. The literature is categorized into two key research areas: (1) Class Imbalance and (2) Structure Imbalance. The relevant references and categorization are presented in Table 2.

3.2.1. Class-imbalanced graph learning

Class imbalance refers to scenarios where the sample size of certain categories is significantly smaller than others. For instance, in social networks, ordinary users vastly outnumber influential "key opinion leaders", while in fraud detection systems, legitimate transactions overwhelmingly dominate over fraudulent ones. Traditional classifiers, including GNNs, often exhibit a bias toward predicting majority-class samples, resulting in suboptimal recognition accuracy for minority classes (Liu et al., 2023d; Ma et al., 2023). Furthermore, GNNs update node representations by aggregating neighborhood information. Minority-class nodes risk being structurally homogenized or misrepresented when their local neighborhoods are dominated by majority-class nodes, leading to a propagation of representation bias across the graph.

(1) Traditional Methods Traditional methods for addressing class imbalance in graph data primarily include clustering, contrastive learning, data augmentation, and innovative loss function. Most articles combine one or more of the above techniques to solve the class imbalance problem.

ECGN (Thapaliya et al., 2024) proposes a cluster-aware graph neural network framework that explicitly models the graph cluster structure to balance class representations, utilizing intra-cluster contrastive loss to enhance the distinguishability of minority class nodes. Similarly, C³GNN (Ju et al., 2024b) clusters majority class nodes into subclasses and contrasts them with the minority class, using Mixup techniques to enhance the semantic substructure representation and alleviate the class imbalance problem. These methods share a key insight: clustering helps balance the class distribution by dividing majority class samples into multiple subclasses, and contrastive learning can improve the separability of minority nodes by constructing positive and negative sample pairs. Building on this, CCGC (Yang et al., 2023) designs a clustering-based positive sample selection strategy and optimizes graph representations through hierarchical contrastive learning, reducing overfitting to the minority class. ImGCL (Zeng et al., 2023) introduces a progressive balanced sampling strategy, dynamically adjusting contrastive weights based on node centrality measures, and utilizes pseudo-labeling to address topo-

logical imbalance. These works highlight how clustering and contrastive learning can be synergistically combined to tackle both structural and label imbalance.

Another line of work focuses on data augmentation, which involves generating diverse topologies or attribute changes for minority classes to balance the data distribution. GraphSMOTE (Zhao et al., 2021), a pioneering method, interpolates minority nodes in the embedding space and predicts edges to preserve topological consistency. However, such interpolation-based methods may struggle with semantic coherence. To address this, GraphSHA (Li et al., 2023d) generates hard samples via feature mixing and employs a semi-mixed connection strategy to prevent neighbor-class encroachment, enhancing boundary discrimination. Augmentation is often paired with adaptive loss functions. For example, TAM (Song et al., 2022) introduces topology-aware margin loss, dynamically adjusting classification boundaries based on local neighbor distributions of nodes, addressing decision bias caused by label node location shifts.

Beyond data generation, another research direction emphasizes directly strengthening tail-class representations and rebalancing classifiers, often through retrieval or expert-based designs. For instance, RAHNet (Mao et al., 2023) combines a retrieval module with a second-stage classifier that applies max-norm and weight decay regularization. This joint design improves representation quality while reducing head-class dominance. Building on the idea of specialized modules, CoMe (Yi et al., 2023) introduces a collaborative-expert framework, where a general expert captures global shared patterns and a specific expert focuses on tail-class features; a collaboration regularizer further enforces complementarity, leading to more balanced performance. Extending this paradigm, KDEX (Mao et al., 2025) trains knowledge-diverse experts, each specializing in different structural or semantic patterns, and integrates their outputs via an expert allocation and fusion mechanism. Collectively, these approaches highlight how retrieval and expert-based architectures can complement generation strategies by directly enhancing the representation and robustness of tail classes.

In addition, HGIF (Ren et al., 2024a) addresses the heterogeneous graph scenario by using invariance learning to separate semantically invariant features, enhancing the robustness of fraud detection in out-of-distribution settings. GraphSANN (Liu et al., 2023a) breaks the homophily assumption and designs a topological reconstruction framework to tackle the class imbalance problem in heterogeneous graphs.

(2) Methods Based on LLMs LLM-based techniques offer a novel approach to address class imbalance. Researchers can overcome the limitations of traditional methods by leveraging three aspects: semantic-driven data augmentation, context-aware text understanding, and external knowledge injection. Fig. 7 illustrates how LLMs can address class imbalance in graph learning by synthesizing semantically rich representations for minority class nodes. In this framework, graph neighborhoods, node attributes, and type constraints are linearized into natural language prompts, serving as the sole input modality for the LLM. This allows the model to generate new nodes or edge candidates with both semantic coherence and structural compatibility.

First, some methods utilize the text generation capabilities of LLMs for node attribute descriptions (such as entity descriptions and label semantic expansion) to synthesize more semantically coherent features for minority class nodes. LA-TAG (Wang et al., 2024d) proposes a framework for text attribute rewriting and synthesis based on LLMs. It generates minority class node descriptions compatible with graph topology through instruction tuning, solving the semantic disconnection problem in text-graph alignment of traditional methods. Similarly, LLM4NG (Yu et al., 2025) designs a node generation approach for few-shot scenarios. It utilizes the LLM decoder to generate virtual nodes with both semantic coherence and structural consistency, while applying contrastive regularization to constrain the embedding space distribution. The method in (Zhang et al., 2024e) also integrates graph structure understanding modules with LLM generators to achieve fine-grained synthesis of sparse data

Table 2
LLM-based methods for handling **imbalance** in graphs, grouped by domain and task, with representative methods, datasets, metrics, and downstream tasks.

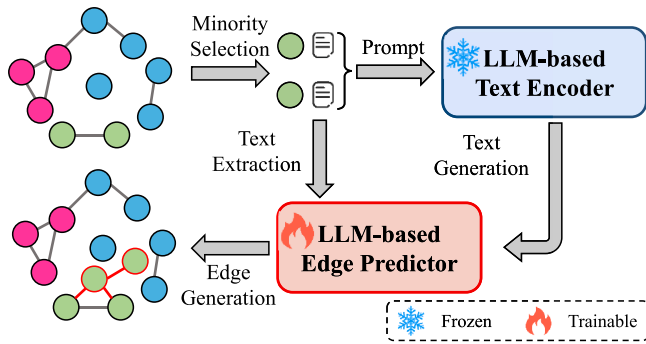| Domain	Tasks	Method	Typical Datasets	Common Metrics	Downstream Tasks
Class Imbalance Graph Learning	Node Classification	LLM4NG (Yu et al., 2025)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), ogbn-arxiv (Hu et al., 2020)	Accuracy	Few-shot Node Classification
		LLM-GNN (Chen et al., 2023c)	Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), PubMed (Prithviraj et al., 2008), ogbn-arxiv (Hu et al., 2020), ogbn-products (Hu et al., 2020), Wikics (Mernyei & Cangea, 2020)	Accuracy	Label-free Node Classification
		G2P2 (Wen & Fang, 2023)	Cora (McCallum et al., 2000), Art (Ni et al., 2019), Industrial (Ni et al., 2019) and Music Instruments (Ni et al., 2019)	Accuracy, Macro-F1	Zero- and Few-shot Low-resource Text Classification
		LA-TAG (Wang et al., 2024d)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), Photo (Yan et al., 2023), Computer (Yan et al., 2023), and Children (Yan et al., 2023)	Accuracy, Macro-F1	Zero- and Few-shot Low-resource Text Classification
		GSS-Net (Zhang et al., 2024e)	Magazine Subscriptions (Ni et al., 2019), Appliances (Ni et al., 2019), Gift Cards (Ni et al., 2019)	Accuracy, Precision, Recall, F1-score, MSE, RMSE, and MAE	Sentiment Classification on Streaming E-commerce Reviews
		TAGrader (Pan et al., 2024)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), ogbn-products (Hu et al., 2020), Arxiv-2023 (He et al., 2024a)	Accuracy	Node Classification on Text-attributed Graphs
		SEGA (Chen et al., 2024b)	DAIC-WOZ (Gratch et al., 2014), EATD (Shen et al., 2022)	Macro-F1	Depression Detection
		SocioHyperNet (Shu et al., 2024)	MBTI (Shu et al., 2024)	Accuracy, AUC, Macro-F1, Micro-F1, IMP	Examining Personality Traits
		Cella (Zhang et al., 2024d)	Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), PubMed (Prithviraj et al., 2008), Wiki-cs (Mernyei & Cangea, 2020)	Accuracy, NMI, ARI, F1-score	Label-free Node Classification
		LLM-TIKG (Hu et al., 2024)	threat-dataset (Hu et al., 2024)	Precision, Recall, F1-Score	Threat Intelligence Knowledge Graph Construction
		ANLM-assInNNER (Liao et al., 2025)	NE dataset (Liao et al., 2025)	Precision, Recall, F1-Score	Robotic Fault Diagnosis Knowledge Graph Construction
		LLM-HetGDT (Ma et al., 2025)	Twitter-HetDrug (Ma et al., 2025)	Macro-F1, GMean	Online Drug Trafficking Detection
	Prediction	LLM-SBCL (Ni et al., 2024)	biology (Ni et al., 2024), law (Ni et al., 2024), cardiff20102 (Ni et al., 2024), sydney19351 (Ni et al., 2024), and sydney23146 (Ni et al., 2024)	Binary-F1, Micro-F1, Macro-F1, Accuracy	Student Performance Prediction
		LKPNR (Runfeng et al., 2023)	MIND (Wu et al., 2020)	AUC, MRR, nDCG	Personalized News Recommendation
		LLM-DDA (Gu et al., 2024b)	BCFR-dataset (Gu et al., 2024b)	AUC, AUPR, F1-score, Precision	Computational Drug Repositioning
	Graph Completion	KICGPT (Wei et al., 2023)	FB15k-237 (Toutanova & Chen, 2015), WN18RR (Dettmers et al., 2018)	MRR, Hits@N	Link Completion
		KGCD (Hou et al., 2025)	WN18RR (Dettmers et al., 2018), YAGO3 -10 (Toutanova & Chen, 2015), WN18 (Bastos et al., 2021)	MRR, Hits@N	Low-resource Knowledge Graph Completion
	Foundation Model	GraphCLIP (Zhu et al., 2024c)	ogbn-arXiv (Hu et al., 2020), Arxiv-2023 (He et al., 2024a), PubMed (Prithviraj et al., 2008), ogbn-products (Hu et al., 2020), Reddit (Huang et al., 2024), Cora (McCallum et al., 2000), CiteSeer (Giles et al., 1998), Ele-Photo (Yan et al., 2023), Ele-Computers (Yan et al., 2023), Books-History (Yan et al., 2023), Wikics (Mernyei & Cangea, 2020), Instagram (Huang et al., 2024)	Accuracy	Transfer Learning on Text-attributed Graphs
Structure Imbalance Graph Learning	Node Classification	GraphEdit (Guo et al., 2024)	Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), PubMed (Prithviraj et al., 2008)	Accuracy	Refining Graph Topologies
	Graph Completion	SATKGC (Ko et al., 2024)	WN18RR (Dettmers et al., 2018), FB15k-237 (Toutanova & Chen, 2015), Wikidata5M (Wang et al., 2021b)	MRR, Hits@N	Knowledge Graph Completion
		MPIKGC (Xu et al., 2024a)	FB15k-237 (Toutanova & Chen, 2015), WN18RR (Dettmers et al., 2018), FB13 (Socher et al., 2013), WN11 (Socher et al., 2013)	MR, MRR, Hits@N, Accuracy	Knowledge Graph Completion
		LLM4RGNN (Zhang et al., 2024i)	Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), PubMed (Prithviraj et al., 2008), ogbn-arxiv (Hu et al., 2020), ogbn-products (Hu et al., 2020)	Accuracy	Improving the Adversarial Robustness

Fig. 7. Illustration of an LLM-based pipeline for addressing class imbalance in graphs. Minority class information is first extracted and linearized into textual prompts, which are processed by the LLM to generate semantically enriched nodes and edges. The newly synthesized elements (highlighted in red) are integrated into the original graph to enhance representation and balance class distributions.

streams, dynamically balancing class distributions in real-time graph data.

Another approach focuses on the contextual understanding and reasoning capabilities of LLMs, mining implicit relationships of minority classes through structure-aware reasoning. Both LLM-GNN (Chen et al., 2023c) and Cella (Zhang et al., 2024d) proposed a zero-shot classification framework based on LLM semantic reasoning. By parsing node attributes and neighbor context, they directly generate category prediction confidence and get rid of the dependence on labeled data. G2P2 (Wen & Fang, 2023) introduces a graph-guided pretraining prompt framework, using LLMs to generate augmented text related to graph structure, enhancing the discriminability of low-resource classes in heterogeneous spaces. Meanwhile, LLM-TIKG (Hu et al., 2024) builds a threat intelligence knowledge graph, using LLMs to extract attack pattern association rules from unstructured text, and enhancing the representation of minority class threat entities through heterogeneous graph attention.

To improve generalization on sparse or long-tail categories in graphs, another line of work focuses on injecting external knowledge-either from open-domain corpora or domain-specific sources-into the learning process. KICGPT (Wei et al., 2023) enriches long-tail relational reasoning by dynamically constructing structure-aware prompts with contextual knowledge, significantly improving completion performance for sparse relations. In the recommendation domain, LKPNR (Runfeng et al., 2023) integrates user preference reasoning from LLMs and semantic path mining to address cold-start scenarios in long-tail news recommendation. KGCD (Hou et al., 2025) augments knowledge graph completion with pseudo-triplets generated from logical rules, guiding the model to focus on underrepresented relations. Other methods, such as (Pan et al., 2024), distill open-domain knowledge from LLMs into graph encoders through hierarchical knowledge transfer, while GraphCLIP (Zhu et al., 2024c) leverages cross-modal alignment to construct graph-text contrastive learning objectives for tail-category enhancement. In biomedical applications, LLM-DDA (Gu et al., 2024b) infuses mechanistic knowledge into graph reconstruction, improving drug repurposing predictions for rare diseases. These works demonstrate that LLMs serve not just as predictors, but as external knowledge carriers that can guide graph learning systems beyond their native data distributions.

In addition to the aforementioned new technologies, the introduction of LLMs has accelerated innovation in cross-domain applications. SEGA (Chen et al., 2024b) uses LLMs to parse the structured elements of clinical conversation graphs, extracting depression semantic clues and social interaction patterns, enhancing the recognition of minority positive samples in mental health classification. The method in Shu et al. (2024) proposes an LLM-enhanced sociological analysis framework based on hypergraphs, addressing behavior class imbalance in social media data

through personality trait hyperedge modeling and LLM semantic deconstruction. ANLM-assInNNER (Liao et al., 2025) develops an automatic construction system for robot fault diagnosis knowledge graphs, using LLMs to generate fine-grained fault entity descriptions and balancing the data distribution of equipment status categories. LLM-SBCL (Ni et al., 2024) combines graph neural networks with LLM cognitive state modeling, improving the reliability of predicting questions on less common knowledge points through semantic enhancement of learner-question interaction graphs.

In summary, LLM-based approaches for class imbalance aim to even out the information content per class by generating additional data or highlighting distinguishing features, rather than just mathematically re-weighting or duplicating existing data.

3.2.2. Structure-imbalanced graph learning

Structural imbalance refers to skewness in graph topology that can hinder learning. A typical example is a hub node vs. peripheral node issue: Hub nodes with very high degree can dominate aggregation and also often have many more training signals, whereas low-degree nodes might be ignored or get a very noisy aggregate from a single neighbor. Similarly, certain substructures might be over-represented. For example, in a molecule graph dataset for drug discovery, maybe most molecules contain a benzene ring but only a few contain a rarer motif; a GNN might mostly learn features relevant to benzene rings and be less sensitive to the rare motif, which could be critical for certain properties.

(1) Traditional Methods The main techniques for addressing structural imbalance in graph data can be broadly divided into two lines: debiasing-based approaches and structural enhancement-based approaches. From the debiasing perspective, several methods focus on dynamically adjusting neighbor aggregation weights to reduce the over-dominance of high-centrality nodes. For example, DegFairGNN (Liu et al., 2023e) introduces a generalized degree fairness constraint that re-allocates neighborhood attention weights between high- and low-degree nodes, thereby mitigating bias in GNN message passing. Extending this idea to knowledge graphs, KG-Mixup (Shomer et al., 2023) analyzes degree bias in entity embeddings and proposes a degree-aware contrastive loss to balance geometric constraints between high- and low-degree entities. Together, these approaches highlight the importance of reweighting meachanisms for correcting degree-induced biases. On the other hand, structural enhancement approaches aim to enrich the representation capacity of underrepresented nodes or substructures. SOLT-GNN (Liu et al., 2022) improves graph classification by using a size-aware hierarchical pooling strategy that balances representation distributions through subgraph cropping and feature decoupling. Building on the idea of subgraph augmentation, SAILOR (Liao et al., 2023) strengthens the visibility of low-degree nodes in the global topology via k-hop subgraph expansion and adversarial edge generation. Both methods demonstrate that explicitly enhancing substructures can effectively alleviate long-tailed imbalance in graphs. A complementary line of work tackles imbalance from the perspective of global structure reconstruction and regularization. HiRe (Wu et al., 2022) develops a hierarchical relational meta-learning framework with a meta-path-guided negative sampling mechanism, balancing the structural coverage density between head and tail relations in knowledge graphs. Similarly, QTIAH-GNN (Liu et al., 2023c) introduces a heterogeneous GNN that jointly considers quantity imbalance and topological imbalance, using meta-relation-specific neighbor sampling together with topological entropy regularization. These methods illustrate how global structural modeling can address both numerical and topological imbalance.

Although existing methods have alleviated the problem of graph structural imbalance to a certain extent, they still have limitations. They often rely on manually designed constraint rules, which leads to limited generalization ability and difficulty in adapting to complex and changing graph structural deviations. At the same time, these methods tend to ignore the semantic associations of the global topology, and usually adopt static adjustment strategies, which cannot dynamically adapt to
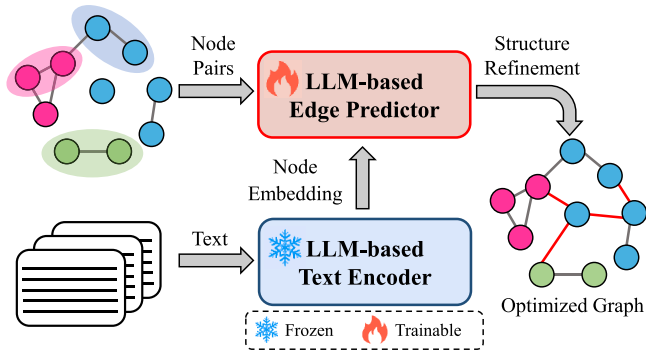
Fig. 8. Structural imbalance remediation via LLM-based text-topology co-optimization. The optimized graph (right) shows the rebalanced connectivity, with new connections indicated in red.

the structural evolution patterns implicit in the graph (Liu et al., 2023d; Ma et al., 2023).

(2) Methods Based on LLMs LLM-based methods offer a new paradigm for addressing structural imbalance in graphs by generating semantically meaningful edits-such as context-aware virtual edges or subgraph structures-that enhance connectivity for long-tail nodes (Fig. 8). Unlike traditional GNN-based approaches, which learn structural patterns primarily from in-graph statistical distributions, LLM-based frameworks integrate rich prior semantic knowledge obtained from large-scale pretraining with graph-specific structures.

For example, SATKGC (Ko et al., 2024) constructs subgraph-aware embeddings and uses contrastive losses to reinforce coherence among tail relations. MPIKGC (Xu et al., 2024a) integrates LLMs for semantic enhancement, structural correction, and logic-driven edge generation, effectively mitigating relational imbalance. LLM4RGNN (Zhang et al., 2024i) transfers GPT-4's reasoning capacity to detect and recover from malicious structural edits in a lightweight graph refinement framework. Complementing these, GraphEdit (Guo et al., 2024) proposes an iterative generate-and-validate loop, where LLMs suggest candidate edges which are then filtered by topological validators to ensure semantic and structural consistency. These approaches demonstrate that, when graphs are represented in text-compatible forms, LLMs can inject new connections that are not arbitrary, but semantically aligned-offering principled structure augmentation for underrepresented classes.

3.2.3. Evaluation for handling imbalance in graphs

We summarize the evaluation protocols for LLM-based methods addressing graph imbalance, covering key datasets, metrics, and downstream tasks (see Table 2). Commonly used datasets include Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), PubMed (Prithviraj et al., 2008), and ogbn-arxiv (Hu et al., 2020) for node classification under class imbalance, and FB15k-237, WN18RR (Dettmers et al., 2018; Toutanova & Chen, 2015) for knowledge graph completion in low-resource settings. Domain-specific benchmarks such as threat-dataset (Hu et al., 2024) and Twitter-HetDrug (Ma et al., 2025) are also adopted for specialized applications like fraud detection and mental health analysis.

Standard evaluation metrics vary by task: Accuracy and Macro-F1 are widely used for classification, while MRR and Hits@N are reported for link prediction and knowledge graph completion. For recommendation and retrieval tasks, AUC, MRR, and nDCG are common. Downstream tasks include few-shot node classification, sentiment analysis, knowledge graph construction, and drug repositioning, demonstrating the adaptability of LLMs in balancing both label and structural distributions. These evaluations highlight the effectiveness of LLM-based methods in enhancing model performance on underrepresented classes and nodes.

3.2.4. Summary of imbalance

Traditional graph imbalance methods mitigate bias via re-sampling, cost adjustment, or graph augmentation, but they lack external semantic context. LLM-integrated methods aim to generate new graph content or features that specifically bolster the minority classes or structures. As reported in recent studies, using LLMs in this way can significantly improve classification performance on long-tail classes and ensure that even structurally unique nodes are recognized. Essentially, LLMs function as an intelligent oversampling mechanism: instead of naive duplication, they produce novel yet relevant samples in data space. The result is often more balanced training data for the graph model and better generalization to minority cases.

3.3. Cross-domain heterogeneity in graphs

Real-world graph data is often collected from multiple source domains (Collarana et al., 2017; Li et al., 2023e; Petermann et al., 2014), which can exhibit significant heterogeneity, referring to extreme disparities in both attributes and structural patterns. This heterogeneity typically arises when the graph data is collected or integrated from domains with inconsistent data modalities or distributions. Cross-domain heterogeneity introduces significant challenges for graph analysis and modeling, as the severe disparities make it difficult to unify these data into a common representation space (Tsai et al., 2016; Zhang et al., 2019c) and even hinder the detection of valuable and transferable features that can generalize well throughout the graph (Hassani, 2022; Zhu et al., 2021). Moreover, addressing cross-domain heterogeneity is a key prerequisite for building graph foundation models for more generalizable and scalable graph learning across diverse real-world applications (Liu et al., 2024a).

While traditional graph domain adaptation methods (Wu et al., 2024a) and multimodal graph learning methods (Peng et al., 2024a) can help bridge the gap between different distributions or modalities from different graph domains, they typically require large-scale training datasets from different domains and cannot generalize to some unseen domains during training. In contrast, LLMs, with their superior comprehension and generalization abilities, can extract valuable and transferable semantic and structural features within a unified representation space without training from scratch, effectively addressing the challenges posed by cross-domain heterogeneity in graph learning (Liu et al., 2024c).

In this section, as illustrated in Fig. 9, we categorize cross-domain heterogeneity in graph data into three types: within-modality attribute heterogeneity, cross-modality attribute heterogeneity, and structural heterogeneity. Since LLMs are specifically designed for modeling textual modality, our focus on within-modality attribute heterogeneity primarily centers around textual attributes. For each type, we first review traditional graph learning methods, analyzing their strengths and limitations, and then discuss recent research leveraging LLMs to unify graph data and address the corresponding challenges. The relevant references and categorization are summarized in Table 3.

3.3.1. Text-attributed graph learning

Textual attributes are common in real-world graph data, such as the abstract of each paper in a co-citation network (Yang et al., 2021) or item descriptions in a recommendation network (Wei et al., 2024a). Although these textual attributes appear in the same modality, they can demonstrate significant distribution heterogeneity due to different sources. For example, textual attributes such as clinical notes could come from different healthcare providers, each with their own writing style, vocabulary, and context (Mashima et al., 2024). Similarly, clinical notes might be informal and easy to understand for patients, while medical codes should be more formal and standardized. Moreover, the textual attributes can appear in different languages (Moreo et al., 2022), further increasing the heterogeneity. The distribution heterogeneity in textual attributes can lead to inconsistent semantic representations, making it

difficult for graph learning methods to effectively capture transferable features.

(1) Traditional Methods for Textual Attribute Modeling To handle the heterogeneity in textual data, early data integration methods extract structured information from unstructured textual inputs, thereby reducing variability across sources and producing unified features for downstream tasks (Dong & Rekatsinas, 2018; Ford et al., 2016; Sheikhalishahi et al., 2019). However, these approaches require manual design of data schemes and training extraction models tailored to specific applications. In contrast, statistical methods, such as Bag of Words and TF-IDF

Table 3
LLM-based methods for handling **cross-domain heterogeneity** in graphs, grouped by domains and tasks, with representative methods, datasets, metrics, and downstream tasks.

Domains	Tasks	Methods	Typical Datasets	Common Metrics	Downstream Tasks
Text-Attributed Graph Learning	Textual Attribute Alignment	**TAPE** (He et al., 2024a)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), Arxiv-2023 (He et al., 2024a), ogbn-arxiv (Hu et al., 2020), ogbn-products (Hu et al., 2020)	Accuracy	Node Classification
		LLMRec (Wei et al., 2024a)	MovieLens (Harper & Konstan, 2015), Netflix (Netflix, Inc., 2025)	Recall, NDCG, Precision	Item Recommendation
		MINGLE (Cui et al., 2024)	MIMIC-III (Johnson et al., 2016), CRADLE (Cui et al., 2024)	Accuracy, AUC, AUPR, F1	Node Classification
		GHGRL (Gao et al., 2024a)	IMDB (Zhang et al., 2019a), DBLP (Zhang et al., 2019a), ACM (Zhang et al., 2019a), Wiki-CS (Mernyei & Cangea, 2020), IMDB-RIR (Gao et al., 2024a), DBLP-RID (Gao et al., 2024a)	Macro-F1, Micro-F1	Node Classification
	Graph Foundation Model	**OFA** (Liu et al., 2024a)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), ogbn-arxiv (Hu et al., 2020), Wiki-CS (Mernyei & Cangea, 2020), MOLHIV (Wu et al., 2018), MOLPCBA (Wu et al., 2018), FB15K237 (Toutanova & Chen, 2015), WN18RR (Dettmers et al., 2018), ChEMBL (Gaulton et al., 2012)	Accuracy, AUC, AUPR	Node Classification, Link Prediction, Graph Classification
		UniGraph (He & Hooi, 2024)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), ogbn-arxiv (Hu et al., 2020), ogbn-products (Hu et al., 2020), Wiki-CS (Mernyei & Cangea, 2020), FB15K237 (Toutanova & Chen, 2015), WN18RR (Dettmers et al., 2018), MOLHIV (Wu et al., 2018), MOLPCBA (Wu et al., 2018)	AUC	Node Classification, Link Prediction, Graph Classification
		BooG (Cheng et al., 2024)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), ogbn-arxiv (Hu et al., 2020), Wiki-CS (Mernyei & Cangea, 2020), MOLHIV (Wu et al., 2018), MOLPCBA (Wu et al., 2018)	AUC	Node Classification, Graph Classification
		Hyper-FM (Feng et al., 2025a)	Cora-CA-Text (Feng et al., 2025a), Cora-CC-Text (Feng et al., 2025a), Pubmed-CA-Text (Feng et al., 2025a), Pubmed-CC-Text (Feng et al., 2025a), AminerText (Feng et al., 2025a), Arxiv-Text (Feng et al., 2025a), Movielens-Text (Feng et al., 2025a), IMDB-Text (Feng et al., 2025a), GoodBook-Text (Feng et al., 2025a), PPI-Text (Feng et al., 2025a)	Accuracy	Node Classification
Multimodal Attributed Graph Learning	MLLM-based Multimodal Alignment	**LLMRec** (Wei et al., 2024a)	MovieLens (Harper & Konstan, 2015), Netflix (Netflix, Inc., 2025)	Recall, NDCG, Precision	Item Recommendation
		MAGB (Yan et al., 2024)	Cora (McCallum et al., 2000), Wiki-CS (Mernyei & Cangea, 2020), Ele-Photo (Yan et al., 2023), Flickr (Zeng et al., 2020), Movies (Yan et al., 2024), Toys (Yan et al., 2024), Grocery (Yan et al., 2024), Reddit-S (Yan et al., 2024), Reddit-M (Yan et al., 2024)	Accuracy, F1	Node Classification
	Graph-Enhanced Multimodal Alignment	**MMGL** (Yoon et al., 2023)	WikiWeb2M (Burns et al., 2023)	BLEU-4, ROUGE-L, CIDEr	Section Summarization
		GraphAdapter (Li et al., 2024g)	ImageNet (Deng et al., 2009), StandfordCars (Krause et al., 2013), UCF101 (Soomro et al., 2012), Caltech101 (Fei-Fei et al., 2004), Flowers102 (Nilsback & Zisserman, 2008), SUN397 (Xiao et al., 2010), DTD (Cimpoi et al., 2014), EuroSAT (Helber et al., 2019), FGVCAircraft (Maji et al., 2013), OxfordPets (Parkhi et al., 2012), Food101 (Bossard et al., 2014)	Accuracy	Image Classification
		TouchUp-G (Zhu et al., 2024a)	ogbn-arxiv (Hu et al., 2020), ogbn-products (Hu et al., 2020), Books (Zhu et al., 2024a), Amazon-CP (Zhu et al., 2024a)	MRR, Hits@N, Accuracy	link prediction, node classification
		UniGraph2 (He et al., 2025)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), ogbn-arxiv (Hu et al., 2020), ogbn-papers100M (Hu et al., 2020), ogbn-products (Hu et al., 2020), Wiki-CS (Mernyei & Cangea, 2020), FB15K237 (Toutanova & Chen, 2015), WN18RR (Dettmers et al., 2018), Amazon-Sports (Zhu et al., 2024b), Amazon-Cloth (Zhu et al., 2024b), Goodreads-LP (Zhu et al., 2024b), Goodreads-NC (Zhu et al., 2024b), Ele-Fashion (Zhu et al., 2024b), WikiWeb2M (Burns et al., 2023)	Accuracy, BLEU-4, ROUGE-L, CIDEr	Node Classification, Edge Classification, Section Summarization

Table 3
Continued

Domains	Tasks	Methods	Typical Datasets	Common Metrics	Downstream Tasks
Structural Heterogeneous Graph Learning	Topological Graph Textualization	**LLMto-Graph** (Liu & Wu, 2023)	synthetic graph data (Liu & Wu, 2023)	Accuracy, Positive Response Ratio	Node Classification, Path Finding, Pattern Matching
		NLGraph (Wang et al., 2023a)	NLGraph (Wang et al., 2023a)	Accuracy, Partial Credit Score, Relative Error	Path Finding, Pattern Matching, Topological Sort
		Talk-like-a-Graph (Fatemi et al., 2024)	GraphQA (Fatemi et al., 2024)	Accuracy	Link Prediction, Pattern Matching
		GPT4Graph (Guo et al., 2023)	ogbn-arxiv (Hu et al., 2020), MOLHIV (Wu et al., 2018), MOLPCBA (Wu et al., 2018), MetaQA (Zhang et al., 2018)	Accuracy	Node Classification, Graph Classification, Graph Query Language Generation
		GITA (Wei et al., 2024b)	GVLQA (Wei et al., 2024b)	Accuracy	Link Prediction, Pattern Matching, Path Finding, Topological Sort
		LLM4-Hypergraph (Feng et al., 2025b)	LLM4Hypergraph (Feng et al., 2025b)	Accuracy	Isomorphism Recognition, Structure Classification, Link Prediction, Path Finding
Structural Heterogeneous Graph Learning	Attributed Graph Textualization	**Graph-Text** (Zhao et al., 2023)	Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), Texas (Pei et al., 2020), Wisconsin (Pei et al., 2020), Cornell (Pei et al., 2020)	Accuracy	Node Classification
		WalkLM (Tan et al., 2024b)	PubMed (Prithviraj et al., 2008), MIMIC-III (Johnson et al., 2016)	Macro-F1, Micro-F1, AUC, MRR	Node Classification, Link Prediction
		Path-LLM (Shang et al., 2024)	Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), PubMed (Prithviraj et al., 2008), ogbn-arxiv (Hu et al., 2020)	Macro-F1, Micro-F1, AUC, Accuracy	Node Classification, Link Prediction
		Instruct-GLM (Ye et al., 2024)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), ogbn-arxiv (Hu et al., 2020)	Accuracy	Node Classification, Link Prediction
		MuseG-raph (Tan et al., 2024a)	Cora (McCallum et al., 2000), ogbn-arxiv (Hu et al., 2020), MIMIC-III (Johnson et al., 2016), AGENDA (Koncel-Kedziorski et al., 2019), WebNLG (Gardent et al., 2017)	Macro-F1, Micro-F1, Weighted-F1, BLEU-4, METEOR, ROUGE-L, CHRF + +	Node Classification, Graph-to-Text Generation
		Graph-LLM (Chen et al., 2024c)	Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), PubMed (Prithviraj et al., 2008), ogbn-arxiv (Hu et al., 2020), ogbn-products (Hu et al., 2020)	Accuracy	Node Classification
	Graph Token Learning	**GNP** (Tian et al., 2024)	OBQA (Mihaylov et al., 2018), ARC (Clark et al., 2018), PIQA (Bisk et al., 2020), Riddle (Lin et al., 2021), PQA (Jin et al., 2019), BioASQ (Tsatsaronis et al., 2015)	Accuracy	Question Answering
		GraphTo-ken (Perozzi et al., 2024)	GraphQA (Fatemi et al., 2024)	Accuracy	Link Prediction, Pattern Matching
		GraphGPT (Tang et al., 2024a)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), ogbn-arxiv Hu et al. (2020)	Accuracy, Macro-F1, AUC	Node Classification, Link Prediction
		LLaGA (Chen et al., 2024a)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), ogbn-arxiv Hu et al. (2020), ogbn-products (Hu et al., 2020)	Accuracy	Node Classification, Link Prediction
		TEA-GLM (Wang et al., 2024a)	Cora (McCallum et al., 2000), PubMed (Prithviraj et al., 2008), ogbn-arxiv Hu et al. (2020), TAG benchmark (Yan et al., 2023)	Accuracy, AUC	Node Classification, Link Prediction
		HiGPT (Tang et al., 2024b)	IMDB (Zhang et al., 2019a), DBLP (Zhang et al., 2019a), ACM Zhang et al. (2019a)	Macro-F1, Micro-F1, AUC	Node Classification

(Sparck Jones, 1972), have been introduced to automatically generate unified feature vectors from text without relying on domain-specific design. These feature vectors are often used as inputs for graph learning methods, such as GNNs (Hamilton et al., 2017; Kipf & Welling, 2017; Veličković et al., 2018), which further incorporate structural information from the graph to obtain more effective node representations for addressing downstream tasks. For instance, TADW (Yang et al., 2015) approximates DeepWalk (Perozzi et al., 2014) using matrix factorization, where the TF-IDF feature vectors of node textual attributes serve as the initial feature matrix. Paper2vec (Ganguly & Pudi, 2017) utilizes learnable Word2vec (Mikolov et al., 2013) text embeddings as initial

node features, which are then trained by predicting whether two nodes belong to the same neighborhood in the graph. While these early statistical methods or shallow models offer solutions for mapping heterogeneous textual attributes into a unified space, their limited expressiveness cannot effectively capture complex features in these textual attributes. Moreover, these methods adopt text embeddings as fixed initial node features rather than integrating them with structural learning, limiting the potential of leveraging graph topology to enhance the textual attributes representation learning (Yang et al., 2021).

Several advanced methods explore approaches for modeling textual attributes that can better integrate with graph structures, aim-
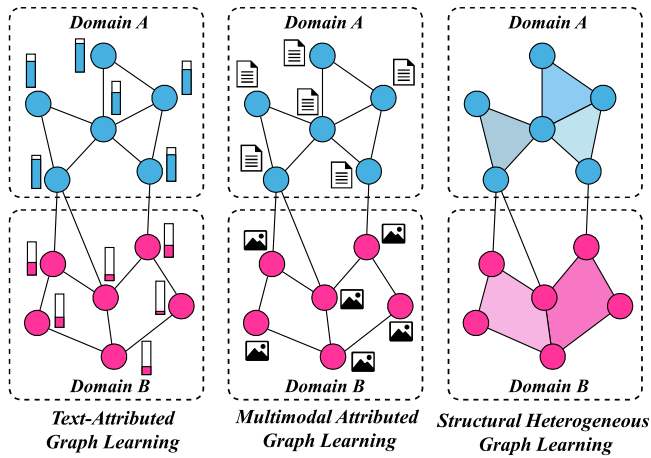
Fig. 9. Illustration of three representative tasks addressing cross-domain heterogeneity in graphs. Text-attributed graph learning (left) involves graphs where nodes or edges are associated with textual content, requiring joint modeling of structure and language. Multimodal attributed graph learning (middle) integrates diverse data types across nodes or relations. Structural heterogeneous graph learning (right) with complex cross-modal dependencies and heterogeneous attribute spaces.

Fig. 10. Handling cross-domain textual attribute heterogeneity with LLMs. Attributes from different domains are first transformed into unified textual descriptions by an LLM, then converted into unified embeddings using a trainable small LM to generate initial features for graph learning methods.

ing to improve the representation learning of both textual attributes and graph structure data simultaneously. Some approaches, like GI-ANT (Chien et al., 2022) and its extension E2EG (Dinh et al., 2023), leverage self-supervised learning to force text embeddings to encode graph-dependent information, which is trained by predicting a node's neighbors from its text. Other methods focus on creating joint architectures. For example, GraphFormers (Yang et al., 2021) proposes a GNN-nested architecture where the transformer-based textual embedding modules and the GNN modules are nested and trained together. Heterformer (Jin et al., 2023b) introduces virtual neighbor tokens that capture information from both text-rich and textless neighbors, allowing the model to simultaneously consider textual semantic information and structural information to embed the textual node attributes. To capture relation-specific signals in heterogeneous graphs, METAG (Jin et al., 2023a) employs a single pretrained language model to learn multiplex text representations, which can effectively handle the diverse semantic relations while maintaining high parameter efficiency.

Although these advanced methods can embed heterogeneous textual node attributes into a unified representation space, their dependence on complex deep architectures requires a sufficient amount of training data. In applications with limited data availability, the model may fail to capture the valuable features from excessive heterogeneous textual attributes, and the learned features might fail to generalize effectively to textual attributes from other sources.

(2) LLM-based Methods for Textual Attribute Modeling LLMs, as language models, are naturally suited for modeling attributes represented in textual modality. With advanced language comprehension and generalization capabilities, LLMs can effectively capture the semantic meanings from heterogeneous textual attributes across diverse source domains and project them into a unified representation space that preserves the semantic information.

A number of methods leverage the use of LLMs to generate unified textual descriptions, which can be fed into a trainable smaller language model to extract task-specific embeddings, as depicted in Fig. 10. TAPE (He et al., 2024a) leverages the powerful language comprehension capabilities of LLMs to predict the category of a node based on its textual attributes and generate explanations for the prediction, which can be regarded as enhanced and aligned textual outputs. These aligned textual outputs are fed into a smaller language model to generate node feature vectors. As shown in Liu et al. (2024c), the domain shift is reduced in
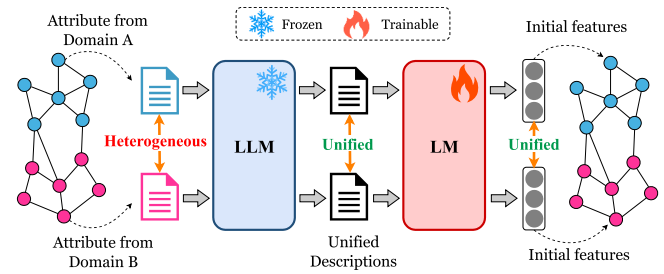
the LLM-enhanced text compared to the original textual attributes. Similarly, LLMRec (Wei et al., 2024a) leverages LLMs to summarize user profiles and item attributes within a recommendation network, reducing the heterogeneity of the original attributes and producing aligned textual representations. These representations are then reprocessed by LLMs to generate unified embedding vectors for both users and items. Instead of generating unified textual descriptions, MINGLE (Cui et al., 2024) utilizes LLMs to map clinical notes and medical codes into a unified embedding space and uses these unified embedding vectors as initial node features for training a hypergraph neural network. GHGRL (Gao et al., 2024a) utilizes LLMs to unify textual attributes on heterogeneous graphs based on format types and content types. Specifically, it first summarizes types information of all nodes using LLMs, which are then used to generate attribute summaries for each node based on the predicted types. These generated attribute summaries are then fed into a language model to generate unified feature vectors for training a graph neural network.

By transforming categorical and numerical features into textual attributes with language-based descriptions, LLMs can extend their ability to handle the heterogeneity in these structured attributes, which do not naturally belong to the textual modality. For example, in molecular graph modeling, node attributes such as atom types or properties are typically represented as categorical or numerical values. By converting these attributes into descriptive text and enriching with domain-specific explanations, LLMs can effectively understand their semantic meaning and unify them within a common representation space. Both OFA (Liu et al., 2024a) and UniGraph (He & Hooi, 2024) address attribute heterogeneity by constructing text-attributed graphs. These models enhance textual attributes by incorporating additional semantically rich contextual descriptions. Furthermore, they leverage domain knowledge to enrich non-textual attributes with textual representations, enabling LLMs to process diverse attributes in a unified manner and generate consistent embedding vectors. BooG (Cheng et al., 2024) follows a similar approach but further employs a contrastive learning-based pretraining objective, which enhances the ability to learn expressive representations and generalize across different domains and downstream tasks. Similarly, Hyper-FM (Feng et al., 2025a) leverages a language model to extract semantic features from cross-domain textual attributes on hypergraphs and integrates structural information through hierarchical high-order neighbor prediction.

These LLM-based methods utilize LLMs to comprehend semantic information from heterogeneous textual attributes and generate unified textual descriptions or vector embeddings for downstream graph learning methods. The unified and high-quality outputs from LLMs greatly simplify the downstream learning process and enhance performance by addressing the challenges posed by cross-domain heterogeneity in textual attributes.

3.3.2. Multimodal attributed graph learning

Compared to attributes within the same modality, multimodal attributes encompass information from various modalities, such as text,

images, audio, and videos. These multimodal attributes offer a more comprehensive and enriched representation of the underlying data. By capturing different aspects of a node through diverse inputs from different sources, multimodal attributes enhance the contextual understanding and provide complementary information that single-modal data may lack (Hu et al., 2021; Wang et al., 2020; Wei et al., 2019). However, this diversity also introduces additional challenges, as attribute heterogeneity across different domains manifests not only in data distribution but also in data formats. This variability complicates the unified processing of these attributes, requiring effective fusion methods to integrate information across different modalities (Wilcke et al., 2020; Yan et al., 2024).

(1) Traditional Methods for Multimodal Attribute Modeling Traditional methods aim to learn alignment or fusion patterns across different modalities by training models to capture the underlying relationships between heterogeneous multimodal attributes. MMGCN (Wei et al., 2019) learns a multimodal graph convolution network on a user-item bipartite graph to learn modal-specific representations of users and micro-videos to better capture user preferences in different modalities. scMoGNN (Wen et al., 2022) utilizes different GNNs to learn representations of each cell-feature bipartite graph in different modalities, and finally concatenates these representations to fuse information from different modalities. Cai et al. (2022) and He and Wang (2023) introduce an attention mechanism in graph neural networks to dynamically capture the importance of information from different modalities. To reduce the need for large amounts of labeled data and improve robustness, contrastive learning methods are proposed for modeling multimodal graph data. Joyful (Li et al., 2023a) designs a global contextual fusion module and a specific modalities fusion module to capture information at different scales, and then concatenates the representations from both modules to generate a unified representation vector for each node. These modules are trained by comparing positive and negative pairs in the corrupted graphs through edge perturbation and random masking. FormNetV2 (Lee et al., 2023a) proposes a centralized multimodal graph contrastive learning strategy to learn fused representations from different modalities in one loss. HGraph-CL (Lin et al., 2022) introduces a hierarchical graph contrastive learning framework that builds intra-modal and inter-modal graphs, leveraging graph augmentations and contrastive learning to capture complex sentiment relations within and across different modalities.

While these traditional methods can align or fuse multimodal attributes in graph data, they still rely on abundant data to train the model and cannot easily generalize to an unseen domain. Compared to LLMs, these traditional methods are less adaptable to diverse data distributions and struggle to leverage pretrained knowledge for more efficient generalization.

(2) LLM-based Methods for Multimodal Attribute Modeling Although LLMs are specifically designed for understanding natural languages and not naturally suited for handling multimodal attributes, they can be combined or aligned with models for other modalities. Leveraging their superior generalization ability, LLM-based methods eliminate the need for large amounts of application-specific training data, opening up new possibilities for addressing attribute heterogeneity across multimodal source domains.

Recent research on MLLMs focuses on developing advanced methods to align LLMs with models from other modalities (Li et al., 2023b; Liu et al., 2024b; Lyu et al., 2024; Sun et al., 2024; Wang et al., 2024f). Although these methods are not specifically designed for graph data, they can be used as powerful preprocessing tools for aligning multimodal attributes and generating unified representations. For example, by unifying textual and visual side information through a pretrained model Clip-ViT (Radford et al., 2021), LLMRec (Wei et al., 2024a) effectively mitigates the multimodal heterogeneity and enhances the node features in the recommendation network, leading to significant performance improvements. MAGB (Yan et al., 2024) conducts experiments on a set of large multimodal attributed graph datasets, which demonstrate

that MLLMs can effectively alleviate the biases from cross-domain multimodal heterogeneity. Recent approaches align encoders for different modalities by leveraging graph structures, making them more naturally suited for graph learning tasks. MMGL (Yoon et al., 2023) utilizes LLMs and image encoders with adapter layers to embed text and image attributes, respectively. These embeddings are then combined with graph positional encodings to capture graph structure information and finally fed into LLMs to generate the corresponding outputs. GraphAdapter (Li et al., 2024g) introduces GNN-based adaptors for encoders of different modalities, which can better align these encoders based on the graph structure information. TouchUp-G (Zhu et al., 2024a) improves node features obtained from pretrained models of different modalities by adapting them to the graph structure, using a new metric called feature homophily to quantify the correlation between the graph and node features, which enhances GNN performance across different tasks and data modalities. Following this direction, UniGraph2 (He et al., 2025) leverages modality-specific encoders alongside a GNN and an MoE module to effectively unify multimodal features while preserving the underlying graph structure.

By aligning LLMs with models for different modalities, these methods can simultaneously understand the heterogeneity across modalities and map the multimodal attributes into a unified embedding space. These unified attribute representations simplify the downstream graph learning process and improve the performance of multi-modal graph learning tasks.

3.3.3. Structural heterogeneous graph learning

Graph structures capture essential connectivity patterns that are crucial for real-world applications. However, graphs constructed from heterogeneous source domains can exhibit excessive heterogeneity in structural patterns. The structural heterogeneity stems from the inherent biases of distinct structural patterns across different source domains, which cannot generalize throughout the graph and may obscure truly valuable structural information. For instance, road network data collected from various cities always exhibit structural heterogeneity, where data from some cities may exhibit grid-like structural patterns and data from others have radial configurations (Badhrudeen et al., 2022). Such heterogeneity hinders the models from capturing underlying generalizable structural features and therefore limits their performance on real-world applications like traffic flow prediction (Zhang et al., 2024a, 2025a) or route planning (Zhuang et al., 2019). Traditional graph domain adaptation methods rely on training data from different domains and cannot generalize to unseen domains. In contrast, LLMs, with their superior semantic understanding and generalization abilities, can understand different graph structural patterns in a zero-shot manner (Wang et al., 2024a), providing new opportunities for mitigating structural heterogeneity in different graph data from unseen domains.

(1) Traditional Methods for Heterogeneous Structure Modeling While many graph learning methods struggle to generalize across structures from different source domains with different data distributions, recent research investigates how to mitigate this issue by adapting a learned model from a source domain to a target domain. These graph adaptation methods provide solutions for eliminating the biases across different domains and help capture unified and generalizable embeddings for graph structures with excessive heterogeneity.

Earlier methods for graph domain adaptation, including DANE (Zhang et al., 2019c) and ACDNE (Shen et al., 2020), employ shared-weight GNNs to align the embedding spaces of different graphs and utilize a least squares generative adversarial network to regularize the distribution alignment, ensuring that the learned unified embeddings do not include domain-specific information. Different from these adversarial regularization methods, some approaches focus on directly aligning distributions from different domains using different metrics. For example, SR-GNN (Zhu et al., 2021) addresses the domain shifts by regularizing the hidden layer distributions using central moment discrepancy. GDA-SpecReg (You et al., 2023) combines optimal transport theory and
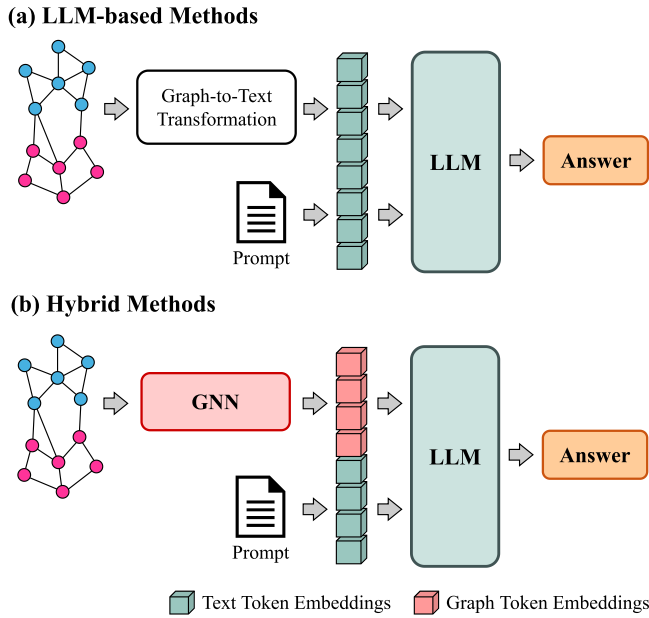
(a) LLM-based Methods

(b) Hybrid Methods

☐ Text Token Embeddings   ☐ Graph Token Embeddings

Fig. 11. Two different approachs for handling cross-domain structural heterogeneity with LLMs. (a) LLM-based methods convert graph structures into textual descriptions, enabling direct input to LLMs. (b) Hybrid methods align GNN-generated graph token embeddings with the native text token embeddings of LLMs in a unified representation space, thereby enhancing the structural comprehension capabilities of LLMs.

graph filter theory to derive a theoretical bound for graph domain adaptation. Based on this bound, the method utilizes Wasserstein-1 distance to regularize the node representation distributions. StruRW (Liu et al., 2023b) proposes an effective approach to reduce conditional structural shifts by re-weighting the edges in the source graph. While most methods align distributions for spatial representations, DASGA (Pilancı & Vural, 2020) introduces a spectral-based approach by aligning the Fourier bases of the source and target graphs, which ensures that the label functions in both domains have similar coefficients in their respective bases.

While these traditional methods demonstrate superiority in mitigating cross-domain structural heterogeneity, they require separate source and target domain data for training, which limits their application in scenarios without such training data or with a large number of domains. Additionally, traditional GNNs typically rely on fixed k-hop neighborhoods, limiting their ability to generalize across highly diverse connectivity patterns from different source domains with varying hop distances. Furthermore, these methods are inherently dependent on training data, which constrains their ability to handle structures that deviate from the patterns observed during training (Zheng et al., 2023b).

(2) LLM-based Methods for Heterogeneous Structure Modeling
Different from traditional graph learning methods, LLM-based methods provide a more flexible alternative by representing graph structures through unified text-based descriptions, which are not restricted by a fixed k-hop structural neighborhood assumption. Instead of relying on predefined graph processing mechanisms, LLMs, with superior comprehension and generalization abilities, can understand unseen structural patterns in different domains. Since LLMs are designed to process natural language inputs, a direct way to bridge the gap between graph structures and the input format of LLMs is to transform structural information into textual descriptions, as depicted in Fig. 11(a). Here we survey recent advances in structure-to-text transformation methods that enable LLMs to understand heterogeneous structural patterns. Although some methods are not specifically designed for mitigating cross-domain structural heterogeneity, they provide strong potential for addressing the associated challenges.

Recent works have explored how to effectively transform graph structures into text to enable better understanding and reasoning by LLMs. Early methods (Fatemi et al., 2024; Guo et al., 2023; Liu & Wu, 2023; Wang et al., 2023a) experimented with existing graph textualization techniques, such as structure description (like node sequences and edge sequences) and formal languages (like graph markup language (Brandes et al., 2010)). While LLMs can comprehend graph structures from these natural language descriptions, their results demonstrate that the choice of different textualization techniques can significantly affect the performance of LLMs on different graph tasks. Different from previous methods that only consider natural language format for representing graph structures, GITA (Wei et al., 2024b) transforms graph structures into both text and images to obtain a better understanding of structural information from different modalities, which are then fed into a Vision-Language Model (VLM) for addressing graph tasks. LLM4Hypergraph (Feng et al., 2025b) designs low-order and high-order structure languages to transform hypergraph structures into natural languages. These studies have significantly advanced the exploration of structure-to-text approaches for LLMs to comprehend graph topological structures.

Beyond merely understanding topological structures, textual descriptions can also incorporate the attributes of nodes and edges, providing a richer context that enables LLMs to capture the intricate features of the graph structure. GraphText (Zhao et al., 2023) constructs a graph-syntax tree to preserve the hierarchical structure and traverses the tree structure to generate a graph text sequence, where the node attributes are incorporated in the description of leaf nodes. WalkLM (Tan et al., 2024b) generates paths by attributed random walks and textualizes the paths into natural languages for processing in LLMs. To effectively capture cross-group connections while minimizing noisy nodes, Path-LLM (Shang et al., 2024) employs shortest paths to generate structural sequences. InstructGLM (Ye et al., 2024) describes graph structures using neighbors in different scales and utilizes instruction tuning to finetune an LLM to better perform graph tasks. The textual node attributes are concatenated with the node index to better incorporate semantic information. MuseGraph (Tan et al., 2024a) incorporates neighbors and paths together to textualize graph structures, which can capture both local connectivity and complex path-based relationships between nodes. Graph-LLM (Chen et al., 2024c) transforms text-attributed graphs into natural languages and demonstrates that LLMs can understand the graph structures from language descriptions and demonstrate remarkable zero-shot performance on graph tasks.

By converting diverse graph structures into natural language descriptions, these structure-to-text transformation methods enable LLMs to interpret and reason across varying connectivity patterns in a unified manner. These approaches allow LLMs to leverage their pretrained linguistic knowledge to identify valuable relational patterns and facilitate more flexible reasoning across different graph topologies with excessive heterogeneity.

(3) Hybrid Methods for Heterogeneous Structure Modeling
While structure-to-text transformation methods offer a straightforward way to bridge the gap between graph structures and the input format of LLMs, they heavily depend on the language comprehension ability of LLMs to correctly interpret the input graph structures. Some recent works explore learning explicit structure representations using traditional graph learning methods like GNNs and aligning these representations with the token space of LLMs, as depicted in Fig. 11(b). These hybrid methods leverage the structural modeling capabilities of traditional graph learning methods alongside the comprehension and reasoning strengths of LLMs for more effective graph structure modeling.

Advanced methods bridge the gap between graph structures and LLMs by converting graph data into graph tokens that LLMs can understand. GNP (Tian et al., 2024) and GraphToken (Perozzi et al., 2024) both use a GNN to encode graph structure and then employ a projector to map these embeddings into the token space of the LLM, allowing them to be processed alongside regular text. LLaGA (Chen et al., 2024a) takes a different approach by reorganizing graph nodes into

structure-aware sequences before projecting them, which helps the model maintain its general-purpose nature. TEA-GLM (Wang et al., 2024a) refines this alignment with feature-wise contrastive learning, using principal components from the token space to precisely map GNN representations. GraphGPT (Tang et al., 2024a) introduces a dual-stage instruction tuning framework that directly aligns graph tokens with LLMs using both self-supervised and task-specific instruction tuning. To generate graph tokens for heterogeneous graphs, HiGPT (Tang et al., 2024b) introduces an in-context heterogeneous graph tokenizer that encodes diverse node and edge types by using a language-based parameterized heterogeneity projector, dynamically generating graph tokens that represent the heterogeneous semantic relationships.

These hybrid methods incorporate structure representations from traditional graph learning methods to facilitate the structure comprehension ability of LLMs, providing more effective solutions for mitigating the challenges from structural heterogeneity. Additionally, the use of compact graph tokens reduces input size and enables LLMs to process larger structures within their context window, which is crucial for understanding heterogeneous structural patterns across varying scales.

3.3.4. Evaluation for handling cross-domain heterogeneity in graphs

We summarize the existing evaluation settings for LLM-based methods for handling cross-domain heterogeneity in graphs, covering benchmark datasets, evaluation metrics, and downstream tasks (see Table 3).

Evaluations typically use well-established traditional graph benchmark datasets. These datasets span a wide range of domains, including citation networks (e.g., Cora (McCallum et al., 2000), Citeseer (Giles et al., 1998), and PubMed (Prithviraj et al., 2008)), biomedical graphs (e.g., MIMIC-III (Johnson et al., 2016), MOLHIV (Wu et al., 2018), and MOLPCBA (Wu et al., 2018)), commercial and recommendation networks (e.g., MovieLens (Harper & Konstan, 2015), Netflix (Netflix, Inc., 2025), and various Amazon product datasets (Zhu et al., 2024a)), as well as knowledge graphs (e.g., FB15K237 (Toutanova & Chen, 2015) and WN18RR (Dettmers et al., 2018)). While most of these datasets are standard in graph learning, cross-domain evaluations differ from conventional setups. Instead of training and testing within the same domain, LLM-based methods are often trained on source-domain datasets (or without further training) and tested on unseen target-domain datasets (Cheng et al., 2024; Wei et al., 2024b; Zhao et al., 2023). In some cases, few-shot settings are also utilized to assess generalization ability when only a limited number of labeled examples in the target domain are available (He & Hooi, 2024; Liu et al., 2024a).

The choice of evaluation metrics typically depends on the task type. Classification tasks adopt Accuracy, Macro-F1, Micro-F1, or AUC, ranking-oriented tasks like recommendation or link prediction use MRR, NDCG, or Hits@N, and generation tasks such as graph-to-text generation rely on BLEU-4, ROUGE-L, CIDEr, METEOR, or CHRF++. Metrics are usually evaluated under zero-shot or few-shot settings to reflect cross-domain generalization. However, traditional metrics may miss cases where outputs are semantically correct but inconsistent in domain-specific format or style. LLMs, in contrast, can capture nuanced correctness across domains based on their strong language comprehension ability, positioning LLM-as-a-Judge (Gu et al., 2024a) a promising direction for future evaluation. The downstream tasks range from node-level and edge-level tasks to graph-level tasks, including node classification, link prediction, and graph classification, as well as more complex settings like recommendation, question answering, section summarization, and graph-to-text generation. This broad coverage sufficiently evaluates the cross-domain generalization ability of LLM-based methods across diverse task types and application scenarios.

3.3.5. Summary of heterogeneity

LLMs, pretrained on diverse datasets, are inherently capable of integrating information from various sources into a unified semantic space. The reviewed studies show that LLM-integrated models can transfer
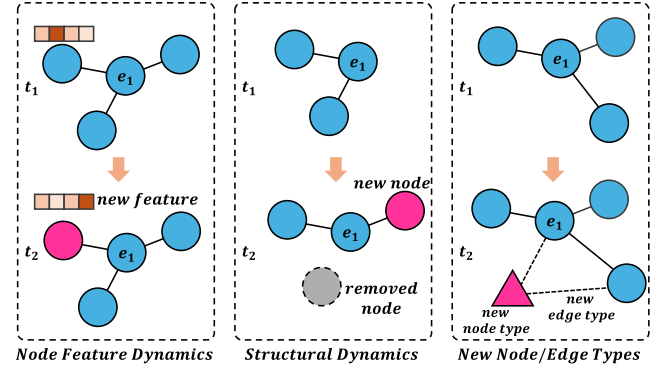

Fig. 12. Illustration of dynamic instability in graphs. The figure presents three categories of dynamic instability over two timestamps ($t_1 \rightarrow t_2$): (1) Node Feature Dynamics, where node attributes change over time; (2) Structural Dynamics, showing newly added node (solid dark purple circle) and newly removed node (gray dashed circle); (3) New Node/Edge Types, depicting the emergence of entirely new node and relationship types, represented by a solid dark purple triangle and dashed edges.

knowledge across different graph domains without the need for domain-specific retraining. By leveraging the generalizability of LLMs, these models are able to mitigate different forms of cross-domain heterogeneity, including textual attribute heterogeneity, multimodal attribute heterogeneity, and structural heterogeneity. As a result, they demonstrate superior performance in addressing cross-domain heterogeneity, particularly in scenarios where traditional methods are ineffective or inapplicable due to the lack of sufficient training data. Experiments in these papers show improved performance on cross-domain node classification, link prediction, and graph classification tasks when LLMs are used to either encode heterogeneous content or assist in aligning representations.

3.4. Dynamic instability in graphs

Many real-world systems, from social networks to knowledge bases, are best represented as graphs whose structure and attributes change over time (see Fig. 12). This constant evolution implies a shifting distribution, where patterns observed in the past may not hold true in the future. While gradual evolution poses modeling challenges, the situation becomes significant when the graph undergoes rapid, large-scale, or highly unpredictable changes that alter the graph's characteristics and create significant challenges for model training and inference (Han et al., 2022).

Dynamic graphs' instability is challenging to model due to their temporal and structural variability (Rossi et al., 2020). In static settings, models learn stable representations from collected graph data. In contrast, dynamic graphs experience frequent changes in nodes, edges, and attributes. Models struggle to generalize without adaptive mechanisms when faced with structural shifts or abrupt transformations (Sankar et al., 2020). Before the introduction of LLMs, researchers addressed dynamic graphs using snapshot-based and incremental methods. These approaches discretized the graph into time-specific snapshots for independent training (Pareja et al., 2020), or updated models as new nodes and edges appeared (Kazemi et al., 2020). While effective in some settings, such methods often rely on localized updates and struggle to capture long-range temporal dependencies, particularly in large-scale or rapidly changing graphs.

LLMs have recently emerged as a powerful new tool with the potential to address some challenges in dynamic graph learning (Fig. 13). By leveraging their strengths in natural language understanding, sequential data processing, few-shot learning, and complex reasoning, LLMs offer novel ways to interpret and model graph dynamics. They can process textual information associated with nodes or edges and understand

Table 4

LLM-based methods for handling **dynamic instability** in graphs, grouped by tasks, with representative methods, datasets, metrics, and downstream tasks.

Domain	Category	Method	Typical Datasets	Common Metrics	Downstream Tasks
Querying and Reasoning	Forecasting & Reasoning	**ICL** (Lee et al., 2023b)	WIKI (Leblay & Chekol, 2018), YAGO (Mahdisoltani et al., 2013), ICEWS14 (García-Durán et al., 2018), ICEWS18 (Jin et al., 2020b)	MRR, Hits@N	Link Prediction
		zrLLM (Ding et al., 2024)	ICEWS (Bossard et al., 2015), ACLED (Raleigh et al., 2010)	MRR, Hits@N	Link Prediction
		CoH (Xia et al., 2024b)	ICEWS14 (García-Durán et al., 2018), ICEWS18 (Jin et al., 2020b), ICEWS05-15 (García-Durán et al., 2018)	MRR, Hits@N	Link Prediction
		TG-LLM (Xiong et al., 2024a)	TGQA (Xiong et al., 2024b), TimeQA (Chen et al., 2021), TempReason (Tan et al., 2023)	F1, Accuracy, Exact Match	Temporal Reasoning
		LLM4DyG (Zhang et al., 2024h)	Enron (Shetty & Adibi, 2004), DBLP (Tang et al., 2008), Flights (Schäfer et al., 2014)	Accuracy, F1, Recall	Spatial-Temporal Reasoning, Graph Reasoning and Querying, Link Prediction
	QA & Interpretability	**TimeR4** (Qian et al., 2024)	MULTITQ (Chen et al., 2023b), TimeQuestions (Jia et al., 2021)	Hits@N	Temporal Knowledge Graph Question Answering
		Gen-TKGQA (Gao et al., 2024b)	CronQuestion (Saxena et al., 2021), TimeQuestions (Jia et al., 2021)	Hits@N	Temporal Knowledge Graph Question Answering
		Unveiling LLMs (Bronzini et al., 2024)	FEVER (Thorne et al., 2018), CLIMATE-FEVER (Diggelmann et al., 2021)	Precision, Recall, F1, ROC AUC, Accuracy	Claim Verification
Generating and Updating	Generating Structures	**FinDKG** (Li & Sanna Passino, 2024)	WIKI (Leblay & Chekol, 2018), YAGO (Mahdisoltani et al., 2013), ICEWS14 (García-Durán et al., 2018)	MRR, Hits@N	Link Prediction
		GenTKG (Liao et al., 2024)	ICEWS14 (García-Durán et al., 2018), ICEWS18 (Jin et al., 2020b), GDELT (Leetaru & Schrodt, 2013), YAGO (Mahdisoltani et al., 2013)	Hits@N	Link Prediction
		Up To Date (Hatem et al., 2024)	Wikidata (Vrandečić & Krötzsch, 2014)	Accuracy, Response Rate	Fact Validation, Question Answering
		PPT (Xu et al., 2023)	ICEWS14 (García-Durán et al., 2018), ICEWS18 (Jin et al., 2020b), ICEWS05-15 (García-Durán et al., 2018)	MRR, Hits@N	Link Prediction
		LLM-DA (Wang et al., 2024b)	ICEWS14 (García-Durán et al., 2018), ICEWS05-15 (García-Durán et al., 2018)	MRR, Hits@N	Link Prediction
	Generating Insights & Representations	**TimeLlama** (Yuan et al., 2024)	ICEWS14 (García-Durán et al., 2018), ICEWS18 (Jin et al., 2020b), ICEWS05-15 (García-Durán et al., 2018)	Precision, Recall, F1, BLEU, ROUGE	Event Forecasting, Explanation Generation
		RealTCD (Li et al., 2024e)	Simulation Datasets (Li et al., 2024e)	Structural Hamming Distance, Structural Interventional Distance	Temporal Causal Discovery, Anomaly Detection
		DynLLM (Zhao et al., 2024b)	Tmall (Tianchi, 2018b), Alibaba (Tianchi, 2018a)	Recall@K, NDCG@K	Dynamic Graph Recommendation, Top-K Recommendation
Evaluation and Application	Model Evaluation	**Dynamic-TempLAMA** (Margatina et al., 2023)	DYNAMICTEMPLAMA (Dhingra et al., 2022)	Accuracy, MRR, ROUGE, F1	Temporal Robustness Evaluation, Factual Knowledge Probing
		DARG (Zhang et al., 2024g)	GSM8K (Cobbe et al., 2021), BBQ (Parrish et al., 2022), BBH Navigate (Suzgun et al., 2023), BBH Dyck Language (Suzgun et al., 2023)	Accuracy, Complexity-Induced Accuracy Retention Rate, Exact Match, Accuracy	Mathematical Reasoning, Social Reasoning, Spatial Reasoning, Symbolic Reasoning
	Downstream Applications	**Anoma-lyLLM** (Liu et al., 2024d)	UCI Messages (Opsahl & Panzarasa, 2009), Blogcatalog (Tang & Liu, 2009)	AUC	Anomaly Detection
		MoMa-LLM (Honerkamp et al., 2024)	iGibson scenes (Li et al., 2021a)	AUC, Recall	Semantic Interactive Object Search
		TRR (Koa et al., 2024)	Reuters Financial News (Ding et al., 2014)	AUROC	Event Detection

the semantic context driving structural changes. This infusion of semantic reasoning capabilities opens promising research routes for creating more robust and adaptive models capable of navigating the complexities of dynamic instability. The relevant references and categorization are presented in Table 4.

3.4.1. Querying and reasoning in dynamic graphs

This category includes work that uses LLMs to query, reason about, or retrieve knowledge from dynamic graphs. These approaches analyze or validate existing graph structures rather than actively generating or modifying node and edge features. LLMs act as information retrieval or
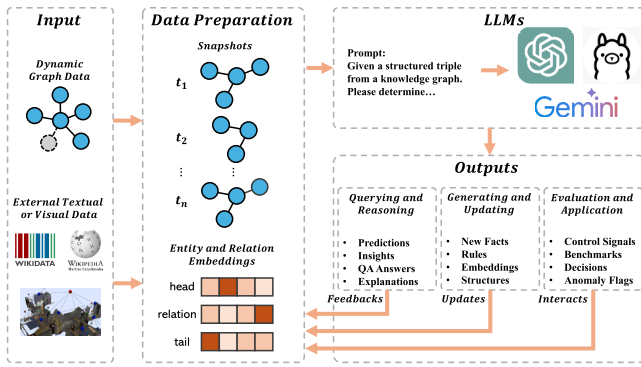
Fig. 13. A unified framework integrating LLMs for addressing dynamic instability in graphs. Dynamic changes in node features, structure, and new node or edge types are captured by temporal graph embeddings. These dynamic embeddings, combined with context (triples and historical facts), serve as inputs to the LLM. The LLM then performs reasoning tasks and outputs the final prediction or response.

reasoning tools, querying, analyzing, or inferring over existing graph structures and knowledge to discover new relationships or validate known ones. The key aspect is that LLMs assist in analyzing and understanding the graph's dynamic evolution.

(1) Traditional Methods Before the emergence of LLMs, querying and reasoning on dynamic graphs relied on specialized graph representation learning and temporal modeling methods. For example, Temporal Graph Networks (TGNs) (Rossi et al., 2020), and related methods use memory modules combined with graph neural networks to process continuous-time dynamic graphs and capture node evolution. Methods like DyRep (c)Trivedi, Farajtabar) and TGAT (Xu et al., 2020) focus on learning node representations that reflect temporal changes in neighborhood structures and interaction patterns. For reasoning on Temporal Knowledge Graphs (TKGs), such as link prediction, traditional approaches employed embedding techniques or rule-based systems to capture changes in entities and relations over time (Su et al., 2024). While effective for specific tasks, these methods struggle to fully use rich semantic information and face challenges in handling zero-shot relations or performing complex multi-step reasoning (Cai et al., 2023).

(2) Forecasting and Reasoning Current research now explores the semantic understanding and reasoning capabilities of LLMs to address querying and reasoning challenges in dynamic graphs, particularly for TKG forecasting and reasoning. One surprising finding is that LLMs, even without fine-tuning, can achieve performance comparable to specialized TKG models on forecasting tasks simply by using In-Context Learning (ICL) with historical facts converted to text. Their performance holds even when entity names are replaced with numerical IDs, suggesting LLMs can leverage structural and temporal patterns in the context (Lee et al., 2023b). To deepen LLM reasoning, CoH (Chain-of-History) (Xia et al., 2024b) proposes a method to explore high-order historical information step-by-step, overcoming the limitation of relying only on first-order history and improving temporal reasoning, especially as a plug-in module for graph-based models. TG-LLM (Xiong et al., 2024a) trains an LLM to translate text context into a latent temporal graph and then uses Chain-of-Thought (CoT) reasoning over this graph, enhancing generalizable temporal reasoning. To address the challenge of unseen relations in TKGs, zrLLM (Ding et al., 2024) uses LLMs to generate semantic representations from relation descriptions, enabling embedding models to recognize zero-shot relations via semantic similarity. Furthermore, LLM4DyG (Zhang et al., 2024h) introduces a systematic evaluation of LLMs' spatial-temporal understanding on general dynamic graphs. It introduces a benchmark and proposes the "Disentangled Spatial Temporal Thoughts" prompting method to improve performance on tasks like link prediction and node classification, although challenges remain with large or dense dynamic graphs.

(3) QA and Interpretability Another line of research focuses on using LLMs for more complex information retrieval from dynamic graphs, such as Temporal Knowledge Graph Question Answering (TKGQA), and understanding the LLMs' own reasoning processes. TKGQA requires understanding complex temporal constraints in questions and retrieving answers from dynamic knowledge, a task where traditional methods struggle with semantics. TimeR4 (Qian et al., 2024) introduces a time-aware Retrieve-Rewrite-Retrieve-Rerank framework. It uses LLMs and retrieved TKG facts to handle time constraints and reduce temporal hallucination by rewriting questions and reranking retrieved facts. Similarly, GenTKGQA (Gao et al., 2024b) employs a two-stage approach: the LLM first mines constraints to guide subgraph retrieval, then generates answers by fusing GNN signals with text representations via instruction tuning. These methods demonstrate the potential of retrieval-augmented LLMs for complex dynamic knowledge QA. Additionally, to understand how LLMs process factual knowledge, Unveiling LLMs (Bronzini et al., 2024) uses dynamic knowledge graphs as a tool to interpret LLM reasoning. It decodes internal token representations layer-wise during fact verification, revealing how factual representations evolve within the LLM.

3.4.2. Generating and updating in dynamic graphs

This category includes papers that use LLMs to actively generate new nodes, edges, or their features or to update dynamic graph structures. These approaches typically serve to complete, reconstruct, or adapt the graph structure to dynamic changes. LLMs function as generators or updaters, creating new node/edge attributes or descriptions to fill in missing information or updating the graph structure in real time based on changes. The focus is on utilizing LLMs' generative capabilities to actively create or modify the graph structure to cope with dynamics.

(1) Traditional Methods Before LLMs became widespread, addressing dynamic graph generation and updates relied mainly on graph generative models and dynamic embedding techniques. Models like GraphRNN (You et al., 2018) and GRAN (Liao et al., 2019) focused on generating static graph structures; extending them to dynamic settings proved challenging. For updating representations in dynamic graphs, methods like DyGEM (Goyal et al., 2018) adapted node embeddings at each time step to reflect graph changes. Knowledge graph updates often depended on manual editing, rule-based systems, or specific database maintenance procedures, which were difficult to automate and could not respond rapidly to real-world changes. These traditional methods generally lacked the ability to use external unstructured information (like text) to guide updates and had limitations in generating semantically consistent and structurally complex graph evolution patterns.

(2) Generating Structures Researchers now use LLMs' generative power to create or maintain evolving knowledge structures. The FinDKG (Li & Sanna Passino, 2024) demonstrates the potential of LLMs as dynamic knowledge graph (DKG) generators; their ICKG model builds a DKG directly from financial news text to capture market trends. To address the issue of outdated information in knowledge graphs, Up To Date (Hatem et al., 2024) proposes a method combining LLM reasoning and RAG. It automatically identifies potentially outdated facts in a KG and retrieves information from trusted sources to generate accurate corrections, enabling automated KG maintenance. For TKG completion and forecasting, PPT (Xu et al., 2023) converts TKG facts and time intervals into prompted natural language sequences, using a masked language modeling task for completion. GenTKG (Liao et al., 2024) also adopts a generative forecasting approach for TKGs. It uses a retrieval-augmented framework with parameter-efficient instruction tuning to generate future facts, achieving strong performance and generalization even with minimal training data. Furthermore, LLM-DA (Wang et al., 2024b) innovatively uses an LLM to generate temporal logical rules from historical TKG data to guide reasoning. A dynamic adaptation strategy updates these rules based on new events, allowing the model to adapt to knowledge evolution without retraining the LLM.

(3) Generating Insights and Representations Beyond directly generating graph structures or facts, LLMs generate higher-level

information like explanations, causal hypotheses, or enhanced node representations to aid dynamic graph analysis and applications. TimeLlama (Yuan et al., 2024) focuses on explainable temporal reasoning. It not only predicts future events but also uses an LLM to generate natural language explanations based on historical TKG paths, increasing model trustworthiness. To explore the underlying mechanisms of dynamic systems, RealTCD (Li et al., 2024e) employs LLMs to process textual information (*e.g.*, system logs) and integrate domain knowledge. Through LLM-guided meta-initialization, it improves the quality of temporal causal discovery, especially in industrial scenarios lacking intervention targets. For dynamic recommendation systems, DynLLM (Zhao et al., 2024b) uses LLMs to generate multi-dimensional user profiles (*e.g.*, interests, preferred brands) and their embeddings from the textual features of purchase histories. These LLM-generated, dynamically updated user profiles enrich user node information, improving the recommendation system's ability to adapt to changing user preferences. These studies show that LLMs can produce basic graph elements and high-level semantic information, offering new ways to understand and model the complex evolution of dynamic graphs.

3.4.3. Evaluation and application in dynamic graphs

This category includes work that uses LLMs to evaluate the effectiveness of dynamic graph learning models or apply LLMs in downstream tasks such as link prediction, node classification, or recommendation systems. LLMs serve as evaluation tools or modules within downstream applications, assessing the performance of dynamic graph models or providing additional guidance during training or inference. The emphasis is on using LLMs' comparative, assessment, or decision-support capabilities to handle tasks in dynamic environments.

(1) Traditional Methods Before using LLMs in these tasks, researchers relied on models that predicted future links or node states based on historical graph data (Jin et al., 2020b). However, standard evaluation metrics often fail to capture the predictions' semantic quality or real-world plausibility (Xiong et al., 2025). These metrics also struggled to assess how models handled unexpected structural changes or evolving concepts. In downstream applications, tasks such as dynamic anomaly detection typically use statistical methods to identify shifts in graph structure or connectivity patterns (Paranjape et al., 2017). Dynamic recommendation systems often relied on sequential models that analyzed user interaction histories (Hidasi et al., 2016) or time-aware collaborative filtering. Although these approaches were effective for specific problems, they lacked the ability to perform deeper semantic reasoning.

(2) Model Evaluation LLMs provide new perspectives for evaluating model performance and robustness in dynamic settings. DynamicTempLAMA (Margatina et al., 2023) presents a dynamic benchmarking framework specifically designed to assess how well pretrained language models (MLMs) handle temporal concept drift-the evolution of factual knowledge over time. The framework dynamically creates time-sensitive test sets from Wikidata and evaluates MLMs through multiple views (probing, generation, scoring) to determine if their internal knowledge is outdated. DARG (Zhang et al., 2024g) addresses the limitations of static benchmarks by proposing a method to dynamically generate new evaluation data. It introduces changes to the reasoning graphs of existing benchmark samples to create novel test data with controlled complexity and diversity, using a code-augmented LLM to ensure label correctness. This enables adaptive evaluation of LLMs' reasoning capabilities as they evolve. Both works highlight the importance of moving beyond static, snapshot-based evaluations toward more dynamic, adaptive approaches, especially as both models and world knowledge constantly change.

(3) Downstream Applications LLMs are also embedded directly into downstream applications that process dynamic graph data, serving as core reasoning or decision-making components. AnomalyLLM (Liu et al., 2024d) uses LLM knowledge for few-shot anomaly edge detection in dynamic graphs. By aligning edge representations with word embed-

ding prototypes and using in-context learning, the method effectively identifies emerging anomaly types with few labeled examples, demonstrating LLMs' potential for adapting to changing environments like cybersecurity. In robotics, MoMa-LLM (Honerkamp et al., 2024) integrates an LLM with dynamically updated, open-vocabulary scene graphs representing an explored environment. The LLM reasons over this evolving graph to guide a mobile robot in long-horizon, interactive object search tasks, showcasing how LLMs can integrate dynamic spatial-semantic information for high-level planning. The financial sector also leverages LLMs; TRR (Temporal Relational Reasoning) (Koa et al., 2024) uses an LLM-based framework mimicking human cognition (memory, attention) to detect potential stock portfolio crashes. It reasons over dynamically generated temporal relational information extracted from news to assess the aggregated impact of evolving events, which is useful for rare events lacking historical data. These applications demonstrate LLMs acting as powerful reasoning engines in complex real-world tasks that require understanding and responding to dynamically changing structured information.

3.4.4. Evaluation for handling dynamic instability in graphs

We review the evaluation pipeline of LLM-based methods for addressing *dynamic instability* in graphs, covering benchmark datasets, evaluation metrics, and downstream tasks (see Table 4). Commonly used datasets include temporal knowledge graph benchmarks such as ICEWS14 (García-Durán et al., 2018), ICEWS18 (Jin et al., 2020b), ICEWS05-15 (García-Durán et al., 2018), YAGO (Mahdisoltani et al., 2013), and WIKI (Leblay & Chekol, 2018) for link prediction and temporal reasoning. Event-based corpora such as ACLED (Raleigh et al., 2010), GDELT (Leetaru & Schrodt, 2013), and TempReason (Tan et al., 2023) are used for temporal forecasting, while domain-specific datasets like Enron (Shetty & Adibi, 2004), DBLP (Tang et al., 2008), Tmall (Tianchi, 2018b), and Alibaba (Tianchi, 2018a) support communication analysis and recommendation. Temporal KGQA relies on MULTITQ (Chen et al., 2023b), TimeQuestions (Jia et al., 2021), and CronQuestion (Saxena et al., 2021), and claim verification uses FEVER (Thorne et al., 2018) and CLIMATE-FEVER (Diggelmann et al., 2021). Anomaly and event detection adopt UCI Messages, Blogcatalog, iGibson scenes, and Reuters Financial News.

Evaluation metrics vary by task: MRR and Hits@N dominate ranking tasks; Accuracy, F1, Precision, Recall, and ROC-AUC are applied in classification; BLEU, ROUGE, and Exact Match assess generative tasks; Recall@K and NDCG@K measure recommendation; and SHD or SID evaluate temporal causal discovery. AUC or AUROC are common for anomaly detection.

The surveyed methods span diverse downstream tasks, including link prediction, temporal reasoning, temporal KGQA, claim verification, fact validation, dynamic graph recommendation, temporal causal discovery, event forecasting, and anomaly or event detection. This diversity highlights the capability of LLMs to integrate temporal reasoning and external knowledge for robust performance in evolving graph environments.

3.4.5. Summary of dynamic instability

The integration of LLMs with dynamic graph learning presents a promising direction for addressing the inherent challenges posed by evolving graph structures. Leveraging their powerful sequence modeling capabilities and the flexibility of natural language encoding, LLMs can effectively adapt to distribution shifts and temporal variations that traditional static graph models struggle to capture. Moreover, LLMs can incorporate external temporal knowledge into the graph reasoning process, enhancing predictive power and enabling anticipation of future changes rather than merely reacting to them. Early empirical evidence suggests that LLM-augmented dynamic graph models exhibit greater robustness and sustained performance over time. Nevertheless, these benefits come with increased computational costs and complexity, requiring careful architectural design and error mitigation strategies to prevent error propagation and maintain reliability in long-term deployment.

4. Limitations, challenges, and future directions

This section builds upon the preceding analysis to consolidate key challenges and limitations in graph-based learning with LLMs, and to outline potential directions for future research. The discussion is organized around several considerations, including **efficiency and scalability**, **explainability, fairness, and faithfulness**, and **security, robustness, and governance**, which together reflect both the current barriers and the broader goals for advancing LLM-graph integration.

4.1. Efficiency and scalability

Scalability is a key challenge in integrating LLMs with graph data, given the high memory and computation costs of large, heterogeneous, and dynamic graphs. We summarize efficiency challenges across scenarios and explore ways to improve tractability, compactness, and inference speed.

4.1.1. Incompleteness
LLM-based methods for graph completion show promise in generating missing nodes, edges, or attributes, but they also introduce risks of hallucination and structural inconsistency. To improve the fidelity of generated content, future research should explore structure-constrained decoding, confidence-aware integration, and post-hoc validation using GNNs or external knowledge bases (b)Sehwag, Papasotiriou). Incorporating structural priors from graphs directly into training objectives may further enhance the alignment between textual reasoning and graph topology (Zhang et al., 2024f). At the same time, reducing representational redundancy is critical for scalability, motivating the development of compact graph-text co-representations. For example, injecting structure tokens or refactoring sequence layouts can reduce the need for full-graph serialization during inference (Coppolillo, 2025; Xue et al., 2023). Besides, efficiency-oriented designs should also consider concrete resource metrics such as sequence length, memory usage, and per-sample or per-token computation cost (Wang et al., 2024e).

4.1.2. Imbalance
Imbalanced distributions are common in real-world graphs, where a few dominant classes overshadow long-tail entities that often lack sufficient structural and semantic support (Guo et al., 2024; Ko et al., 2024). Uniformly applying LLMs across all samples is inefficient and may overfit majority patterns while underperforming on minority nodes. Moreover, hallucinations in underrepresented regions pose reliability risks, especially when topological signals are weak (Zhang et al., 2024i). To mitigate these issues, future research should consider selective LLM invocation for hard or minority-class cases, while using GNNs for routine instances and transferring LLM knowledge into lightweight models via parameter-efficient tuning (Dettmers et al., 2023; Hu et al., 2022). Structure-aware validation modules that cross-check LLM outputs against graph-derived patterns can improve robustness (Wei et al., 2023), and a promising direction is to construct joint text-graph causal representations to help models infer the semantic origins of structural sparsity and synthesize logically grounded virtual connections (Xu et al., 2024a).

4.1.3. Cross-domain heterogeneity
Graphs that span multiple domains or modalities often exhibit heterogeneity in node types, attribute formats, and structural patterns, which increases prompt length and introduces misalignment between textual and topological inputs (Cui et al., 2024; He et al., 2024a). Existing methods struggle to handle these inconsistencies, as they typically rely on shallow normalization or isolated attribute encoding, and LLMs themselves lack permutation invariance when processing graph structures (Zhao et al., 2023). To address these limitations, future efforts should explore context-aware modeling that jointly encodes multimodal attributes, local connectivity, and task-specific signals. Symmetry-preserving techniques such as contrastive learning or explicit prompt designs can help models distinguish meaningful structural variations from order artifacts (Ye et al., 2024). Additionally, tighter LLM-GNN integration-via dual encoders or knowledge distillation from graph-level GNNs-may provide a more balanced trade-off between semantic richness and structural fidelity (Perozzi et al., 2024; Tian et al., 2024).

4.1.4. Dynamic instability
In dynamic graphs, frequent structural or attribute updates introduce substantial overhead for model recomputation and pose challenges for maintaining temporal consistency (Lee et al., 2023b). Current LLM-based approaches often rely on full re-encoding or static snapshots, which fail to reflect evolving semantics and result in stale or inaccurate reasoning (Xiong et al., 2024a). To overcome these limitations, future research should incorporate time-aware prompting, localized subgraph updates, and selective verification to reduce redundant computation (Wang et al., 2024e). Caching frequently accessed substructures or intermediate representations can further improve efficiency. More importantly, integrating GNNs with LLMs for temporal causal reasoning, along with pretraining strategies that embed temporal priors, may help models adapt to semantic drift and support robust inference over evolving graph states (Ding et al., 2024; Xia et al., 2024b).

4.2. Explainability, fairness, and faithfulness

LLM-based graph systems pose serious challenges in explainability, fairness, and faithful reasoning-especially in scenarios involving incomplete data, long-tail distributions, domain shifts, and temporal updates. While current approaches offer partial solutions, their limitations in interpretability, causal alignment, and robustness across scenarios call for more principled advancements.

4.2.1. Incompleteness
Incomplete graphs often result in LLM outputs that lack faithfulness, as the underlying reasoning is not grounded in observable evidence (Chen et al., 2023a; b)Sehwag, Papasotiriou; Xu et al., 2024b). This issue is exacerbated by decoding processes that fail to enforce topological constraints and by the absence of systematic post-generation verification. Future work should incorporate structure-constrained decoding objectives and apply graph consistency losses during training (Zhang et al., 2024f), while complementing generation with post-hoc validation using GNN-based scorers or confidence filters (Chai et al., 2023; Wang et al., 2024c). Furthermore, reasoning fidelity deteriorates significantly when only partial graph observations are available (Xu et al., 2024b). To address this, prompting strategies should explicitly encode subgraph structures to help LLMs reason over local topological contexts. In addition, iterative refinement mechanisms that alternate between retrieval, reasoning, and validation can simulate GNN-style aggregation while supporting auditable evidence chains.

4.2.2. Imbalance
Graph imbalance not only degrades predictive performance but also undermines explanation stability and reliability. Due to sparse semantics and weak connectivity, LLMs may struggle to generate faithful rationales for long-tail nodes. Moreover, existing interpretability tools rarely account for such imbalance, leading to inconsistent explanations under input perturbations. Future work should explore causal mechanisms underlying topological sparsity and use LLMs to generate logically consistent virtual subgraphs to restore class-level balance (Guo et al., 2024). At the same time, interpretability evaluation should emphasize stability: explanations for minority-class nodes should remain consistent across counterfactual edits or input augmentations, and validation modules should reject rationales that do not correspond to truly discriminative substructures.

4.2.3. Cross-domain heterogeneity

Inconsistencies in label semantics, graph patterns, and textual attributes across domains make it difficult for LLMs to produce generalizable explanations. These mismatches often introduce domain-specific biases that harm fairness and reasoning quality (Dong et al., 2023; Gallegos et al., 2024). Current models lack the ability to maintain explanation coherence when transferred across domains (Lucic et al., 2022; Ying et al., 2019). Future systems should develop domain-invariant rationale templates and establish multi-level auditing frameworks covering embedding spaces, output distributions, and textual justifications to detect and mitigate explanation drift. Additionally, aligning the structural reasoning paths produced by GNNs with the textual rationales generated by LLMs can bridge the gap between semantics and topology and enhance explanation faithfulness.

4.2.4. Dynamic instability

The evolving nature of real-world graphs introduces temporal variations that challenge the stability and trustworthiness of model explanations. As nodes and edges change over time, previously valid rationales may become outdated or misleading, yet current systems lack temporal auditing mechanisms. Future work should introduce rolling, time-aware auditing protocols that explicitly link model outputs with graph update events (Dong et al., 2023). Moreover, enabling executable explanations-where LLM-generated rationales are translated into symbolic constraints or verifiable subgraph traces-can improve the reliability and traceability of reasoning under temporal drift (Lyu et al., 2023).

4.3. Security, robustness, and governance

As graph-LLM systems become increasingly applied in high-stakes domains such as finance, healthcare, and recommendation, they face growing vulnerabilities arising from incomplete information, class imbalance, cross-domain inconsistency, and evolving graph structures. These challenges expose the system to risks including data poisoning, prompt injection, structure manipulation, and adversarial generalization, many of which are amplified by the flexible yet opaque behavior of LLMs.

4.3.1. Incompleteness

In settings where graph data is sparse or noisy, even minor perturbations to node or edge attributes can lead to significant shifts in completion outcomes, compromising the integrity of the generated graph (Zhang et al., 2019b; Zügner et al., 2018). Similarly, LLM-augmented retrieval systems are vulnerable to prompt-level attacks, where indirect injection or document poisoning misleads the model's response (De Stefano et al., 2024). To improve robustness, future work should investigate multi-layer defenses, including data-level sanitization and edge pruning, model-level protection through graph denoising techniques such as GNNGuard (Zhang & Zitnik, 2020), and certification-based methods like randomized smoothing to quantify the system's tolerance under adversarial conditions (Bojchevski & Günnemann, 2019; Wang et al., 2021a).

4.3.2. Imbalance

Nodes associated with rare categories or low connectivity often lack sufficient structural context, making them particularly susceptible to adversarial influence; subtle perturbations can alter neighborhood aggregation or trigger class boundary shifts (Jin et al., 2020a). Moreover, few-shot trigger constructions may induce undesirable distributional shifts in LLM outputs. To address these issues, future systems should adopt targeted defenses for long-tail nodes, including structural consistency checks and trigger detection via adversarial training regimes that expose the model to rare-case manipulations.

4.3.3. Cross-domain heterogeneity

When LLM-based models are deployed across domains with differing label semantics, structure formats, or template designs, inconsistencies

can be exploited to induce behavior shifts or prompt corruption. Existing systems often lack robust controls over prompt templates or graph encodings, making them vulnerable to injection and escalation attacks (De Stefano et al., 2024). Looking forward, stronger domain-level governance is needed, including adversarial evaluation prior to deployment, enforcement of signed prompt templates, and whitelist-based control over structural tokens to prevent unauthorized modifications.

4.3.4. Dynamic instability

Dynamic graphs introduce further complexity, as attackers may inject or remove connections during critical time windows or log adversarial traces to shape long-term model behavior (Jin et al., 2020a). These time-sensitive attacks are often stealthy and hard to reverse, leading to compounding damage (Wang et al., 2021a). Future directions should prioritize rolling-window detection and temporal rollback mechanisms, coupled with causal tracing of performance degradation to structural changes. Additional safeguards such as time-aware scoring and path-verifiable explanation generation can help separate legitimate evolution from malicious interference.

5. Conclusion

This survey has presented a comprehensive and systematic review of how recent advances in LLMs can be leveraged to address four fundamental, data-centric challenges in graph learning: incompleteness of structures or attributes, severe imbalance in node and edge distributions, cross-domain heterogeneity in semantics and structure, and dynamic instability arising from evolving topologies and interactions. To achieve this, we conducted an extensive literature collection and categorization, and organized representative methods. We further summarized benchmark datasets and evaluation metrics used in existing studies, assessed empirical trends. In addition, we identified open technical challenges and outlined promising future research directions.

We have synthesized a broad spectrum of LLM-graph integration strategies under this framework, demonstrating how LLMs bring distinctive capabilities that complement purely graph-based approaches. In the context of incompleteness, LLMs apply semantic reasoning and draw upon external knowledge to infer missing attributes and relationships, serving as intelligent imputers. For imbalanced graphs, they can generate synthetic samples and enrich feature spaces, thereby enhancing minority-class representations and mitigating bias. In heterogeneous graph scenarios, LLMs facilitate the unification of disparate modalities and domain-specific schemas into coherent embeddings, enabling effective cross-domain alignment. For dynamic graphs, their contextual and temporal reasoning allows for anticipating structural changes, explaining evolution, and supporting continuous adaptation.

Looking ahead, the integration of LLMs with graph learning still faces important open questions. Improving efficiency and scalability is essential for large-scale or real-time applications, while enhancing interpretability and trustworthiness will be critical for deployment in high-stakes domains such as healthcare and finance. Developing mechanisms for continual adaptation without catastrophic forgetting remains a significant challenge, as does bridging the gap between textual knowledge encoded in LLMs and structural graph signals. In addition, controlling hallucination in LLM-generated graph content is vital to ensure both semantic validity and structural consistency.

In conclusion, the convergence of LLMs and graph learning marks a promising new direction, combining the deep semantic understanding of natural language processing with the structured relational modeling of graph machine learning. This synergy has already yielded models that are more robust, knowledgeable, and adaptable than those based on either technology alone. As research advances, we anticipate rapid progress toward graph learning systems with greater generality and intelligence, capable of reasoning effectively over the rich, dynamic networks that underpin real-world data.

CRediT authorship contribution statement

Mengran Li: Conceptualization, Methodology, Formal analysis, Writing – original draft, Visualization; **Pengyu Zhang:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Visualization; **Wenbin Xing:** Conceptualization, Data curation, Investigation, Writing – original draft, Visualization; **Yijia Zheng:** Conceptualization, Data curation, Investigation, Writing – original draft, Visualization; **Klim Zaporojets:** Software, Validation, Writing – review & editing; **Junzhou Chen:** Supervision, Writing – review & editing; **Ronghui Zhang:** Supervision, Project administration, Funding acquisition, Writing – review & editing; **Yong Zhang:** Supervision, Resources; **Siyuan Gong:** Supervision; **Jia Hu:** Supervision, Resources; **Xiaolei Ma:** Supervision, Resources; **Zhiyuan Liu:** Supervision, Writing – review & editing; **Paul Groth:** Supervision, Writing – review & editing; **Marcel Worring:** Supervision, Writing – review & editing.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Acknowledgements

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S. et al. (2023). GPT-4 technical report. arXiv:2303.08774

Asuncion, A., Newman, D. et al. (2007). UCI machine learning repository. http://archive.ics.uci.edu/ml.

Badhrudeen, M., Derrible, S., Verma, T., Kermanshah, A., & Furno, A. (2022). A geometric classification of world urban road networks. *Urban Science* https://doi.org/{10.3390/urbansci6010011}.

Bastos, A., Singh, K., Nadgeri, A., Shekarpour, S., Mulang, I. O., & Hoffart, J. (2021). HopfE: Knowledge graph representation learning using inverse Hopf fibrations. *CIKM '21* (p. 89-99). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3459637.3482263

Bisk, Y., Zellers, R., Gao, J., Choi, Y. et al. (2020). PIQA: Reasoning about physical commonsense in natural language. In *Proceedings of the AAAI conference on artificial intelligence*. https://doi.org/10.1609/aaai.v34i05.6239

Bojchevski, A., & Günnemann, S. (2019). Certifiable robustness to graph perturbations. *Advances in Neural Information Processing Systems*, *32*, 8319–8330.

Bossard, L., Guillaumin, M., & Van Gool, L. (2014). Food-101–mining discriminative components with random forests. In *European conference on computer vision*. https://doi.org/10.1007/978-3-319-10599-4_29

Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., Starz, J. S., & Ward, M. (2015). In *ICEWS coded event data*. https://doi.org/10.7910/DVN/28075

Brandes, U., Eiglsperger, M., Lerner, J., & Pich, C. (2010). Graph markup language (graphML) http://nbn-resolving.de/urn:nbn:de:bsz:352-244261.

Brasoveanu, A. M. P., Nixon, L., Weichselbraun, A., & Scharl, A. (2023). Framing few-shot knowledge graph completion with large language models. In *Joint workshop proceedings of the 5th International workshop on Sem4Tra and SEMANTiCS*.

Bronzini, M., Nicolini, C., Lepri, B., Staiano, J., & Passerini, A. (2024). Unveiling LLMs: The evolution of latent representations in a dynamic knowledge graph. arXiv:2404.03623

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Bruna, J., Zaremba, W., Szlam, A., & LeCun, Y. (2014). Spectral networks and deep locally connected networks on graphs. In *Proceedings of the International conference on learning representations*. arXiv:1312.6203

Burns, A., Srinivasan, K., Ainslie, J., Brown, G., Plummer, B., Saenko, K., Ni, J., & Guo, M. (2023). A suite of generative tasks for multi-level multimodal webpage understanding. In *Proceedings of the 2023 conference on empirical methods in natural language processing*. https://aclanthology.org/2023.emnlp-main.119/

Cai, B., Xiang, Y., Gao, L., Zhang, H., Li, Y., & Li, J. (2023). Temporal knowledge graph completion: A survey. In *Proceedings of the thirty-second International joint conference on artificial intelligence, IJCAI '23* https://doi.org/10.24963/ijcai.2023/734

Cai, H., Gao, Y., & Liu, M. (2022). Graph transformer geometric learning of brain networks using multimodal MR images for brain age estimation. *IEEE Transactions on Medical Imaging* https://doi.org/10.1109/TMI.2022.3222093.

Cai, J.-F., Candès, E. J., & Shen, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, *20*(4), 1956–1982. https://doi.org/10.1137/080738970

Cao, S., Shi, J., Pan, L., Nie, L., Xiang, Y., Hou, L., Li, J., He, B., & Zhang, H. (2022). KQA Pro: A dataset with explicit compositional programs for complex question answering over knowledge base. In *Proceedings of the 60th annual meeting of the association for computational linguistics*. Association for Computational Linguistics. https://doi.org/10.18653/V1/2022.ACL-LONG.422

Carta, S., Giuliani, A., Piano, L., Podda, A. S., Pompianu, L., & Tiddia, S. G. (2023). Iterative zero-shot LLM prompting for knowledge graph construction. arXiv:2307.01128

Chai, Z., Zhang, T., Wu, L., Han, K., Hu, X., Huang, X., & Yang, Y. (2023). GraphLLM: Boosting graph reasoning ability of large language model. arXiv:2310.05845

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y. et al. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, *15*(3), 1–45. https://doi.org/10.1145/3641289

Chen, D., Zhang, L., Lai, X., Lu, W., & Li, Z. (2025). Temporal decomposition and attribute correlation differentiation at multiple scales: A graph imputation network for incomplete multivariate time series. *Knowledge-Based Systems*, (p. 113636). https://doi.org/10.1016/J.KNOSYS.2025.113636

Chen, J., Ma, L., Li, X., Thakurdesai, N., Xu, J., Cho, J. H. D., Nag, K., Korpeoglu, E., Kumar, S., & Achan, K. (2023a). Knowledge graph completion models are few-shot learners: An empirical study of relation labeling in e-commerce with LLMs. arXiv:2305.09858

Chen, L., Bruna, J., & Bronstein, M. (2019a). Attributed random walk as matrix factorization. In *Proceedings of the International conference on neural information processing systems, graph representation learning workshop*.

Chen, R., Zhao, T., JAISWAL, A. K., Shah, N., & Wang, Z. (2024a). LLaGA: Large language and graph assistant. In *International conference on machine learning*.

Chen, W., Wang, H., Chen, J., Zhang, Y., Wang, H., Li, S., Zhou, X., & Wang, W. Y. (2019b). TabFact: A large-scale dataset for table-based fact verification. arXiv:1909.02164

Chen, W., Wang, X., & Wang, W. Y. (2021). A dataset for answering time-sensitive questions. arXiv:2108.06314.

Chen, X., Chen, S., Yao, J., Zheng, H., Zhang, Y., & Tsang, I. W. (2022). Learning on attribute-missing graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(2), 740–757.

Chen, Z., Deng, J., Zhou, J., Wu, J., Qian, T., & Huang, M. (2024b). Depression detection in clinical interviews with LLM-empowered structural element graph. In *Proceedings of the 2024 conference of the North American chapter of the association for computational linguistics: Human language technologies, Long papers* (vol. *1*) (pp. 8181–8194).

Chen, Z., Liao, J., & Zhao, X. (2023b). Multi-granularity temporal question answering over knowledge graphs. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st annual meeting of the association for computational linguistics, Long papers* (vol. *1*) (pp. 11378–11392). Toronto, Canada: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.637

Chen, Z., Mao, H., Li, H., Jin, W., Wen, H., Wei, X., Wang, S., Yin, D., Fan, W., Liu, H. et al. (2024c). Exploring the potential of large language models (LLMs) in learning on graphs. *ACM SIGKDD Explorations Newsletter* https://doi.org/10.1145/3655103.3655110

Chen, Z., Mao, H., Wen, H., Han, H., Jin, W., Zhang, H., Liu, H., & Tang, J. (2023c). Label-free node classification on graphs with large language models (LLMs). arXiv:2310.04668

Cheng, Y., Zhao, Y., Yu, J., & Li, X. (2024). Boosting graph foundation model from structural perspective. arXiv:2407.19941

Cheng, Z., Dong, H., Wang, Z., Jia, R., Guo, J., Gao, Y., Han, S., Lou, J.-G., & Zhang, D. (2021). HiTab: A hierarchical table dataset for question answering and natural language generation. arXiv:2108.06712

Chien, E., Chang, W.-C., Hsieh, C.-J., Yu, H.-F., Zhang, J., Milenkovic, O., & Dhillon, I. S. (2022). Node feature extraction by self-supervised multi-scale neighborhood prediction. In *International conference on learning representations*. https://openreview.net/forum?id=KJggliHbs8

Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., & Vedaldi, A. (2014). Describing textures in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. https://doi.ieeecomputersociety.org/10.1109/CVPR.2014.461

Clark, P., Cowhey, I., Etzioni, O., Khot, T., Sabharwal, A., Schoenick, C., & Tafjord, O. (2018). Think you have solved question answering? Try ARC, the AI2 reasoning challenge. arXiv:1803.05457

Cobbe, K., Kosaraju, V., Bavarian, M., Chen, M., Jun, H., Kaiser, L., Plappert, M., Tworek, J., Hilton, J., Nakano, R. et al. (2021). Training verifiers to solve math word problems. arXiv:2110.14168

Collarana, D., Galkin, M., Traverso-Ribón, I., Lange, C., Vidal, M.-E., & Auer, S. (2017). Semantic data integration for knowledge graph construction at query time. In *International conference on semantic computing*. https://doi.org/10.1109/ICSC.2017.85

Coppolillo, E. (2025). Injecting knowledge graphs into large language models. arXiv:2505.07554

Cui, H., Fang, X., Xu, R., Kan, X., Ho, J. C., & Yang, C. (2024). Multimodal fusion of EHR in structures and semantics: Integrating clinical records and notes with hypergraph and LLM. arXiv:2403.08818

Cui, Y., Wang, K., Yu, H., Guo, X., & Cao, H. (2025). KLLMs4Rec: Knowledge graph-enhanced LLMs sentiment extraction for personalized recommendations. *Expert Systems with Applications*, *282*, 127430. https://doi.org/10.1016/J.ESWA.2025.127430

Dai, D., Deng, C., Zhao, C., Xu, R. X., Gao, H., Chen, D., Li, J., Zeng, W., Yu, X., Wu, Y. et al. (2024). DeepSeekMoE: Towards ultimate expert specialization in mixture-of-experts language models. arXiv:2401.06066

De Stefano, G., Schönherr, L., & Pellegrino, G. (2024). Rag and roll: An end-to-end evaluation of indirect prompt manipulations in LLM-based application frameworks. arXiv:2408.05025

Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. In *Proceedings of the neural information processing systems* (pp. 3844–3852). (vol. 29).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. https://doi.org/10.1109/CVPR.2009.5206848

Dettmers, T., Minervini, P., Stenetorp, P., & Riedel, S. (2018). Convolutional 2D knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*. (vol. 32). https://doi.org/10.1609/AAAI.V32I1.11573

Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). QLoRA: Efficient finetuning of quantized LLMs. *Advances in neural information processing systems*, *36*, 10088–10115.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171–4186).

Dhingra, B., Cole, J. R., Eisenschlos, J. M., Gillick, D., Eisenstein, J., & Cohen, W. W. (2022). Time-aware language models as temporal knowledge bases. *Transactions of the Association for Computational Linguistics*, *10*, 257–273.

Diggelmann, T., Boyd-Graber, J., Bulian, J., Ciaramita, M., & Leippold, M. (2021). Climate-fever: A dataset for verification of real-world climate claims. arXiv:2012.00614

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2014). Using structured events to predict stock price movement: An empirical investigation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1415–1425). ACL. https://doi.org/10.3115/V1/D14-1148

Ding, Z., Cai, H., Wu, J., Ma, Y., Liao, R., Xiong, B., & Tresp, V. (2024). zrLLM: Zero-shot relational learning on temporal knowledge graphs with large language models. In K. Duh, H. Gomez, & S. Bethard (Eds.), *Proceedings of the 2024 conference of the North American chapter of the association for computational linguistics: Human language technologies (volume 1: Long papers)* (pp. 1877–1895). Mexico City, Mexico: Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.104

Dinh, T. A., den Boef, J., Cornelisse, J., & Groth, P. (2023). E2EG: End-to-end node classification using graph topology and text-based node attributes. In *IEEE International conference on data mining workshops*. https://ieeexplore.ieee.org/abstract/document/10411709.

Dong, X. L., & Rekatsinas, T. (2018). Data integration and machine learning: A natural synergy. In *Proceedings of the 2018 International conference on management of data*. https://doi.org/10.1145/3183713.3197387

Dong, Y., Chawla, N. V., & Swami, A. (2017). metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD International conference on knowledge discovery and data mining* (pp. 135–144). https://doi.org/10.1145/3097983.3098036

Dong, Y., Ma, J., Wang, S., Chen, C., & Li, J. (2023). Fairness in graph mining: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *35*(10), 10583–10602.

Fang, L., Xiang, W., Pan, S., Salim, F. D., & Chen, Y.-P. P. (2025). Spatiotemporal pre-trained large language model for forecasting with missing values. *IEEE Internet of Things Journal* https://doi.org/10.1109/JIOT.2024.3524030

Fatemi, B., Halcrow, J., & Perozzi, B. (2024). Talk like a graph: Encoding graphs for large language models. In *International conference on learning representations*. https://openreview.net/forum?id=IuXR1CCrSi

Fei-Fei, L., Fergus, R., & Perona, P. (2004). Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. In *2004 conference on computer vision and pattern recognition workshop*. https://doi.org/10.1109/CVPR.2004.383

Feng, Y., Liu, S., Han, X., Du, S., Wu, Z., Hu, H., & Gao, Y. (2025a). Hypergraph foundation model. arXiv:2503.01203

Feng, Y., Yang, C., Hou, X., Du, S., Ying, S., Wu, Z., & Gao, Y. (2025b). Beyond graphs: Can large language models comprehend hypergraphs? In *International conference on learning representations*. https://openreview.net/forum?id=28qOQwjuma.

Ford, E., Carroll, J. A., Smith, H. E., Scott, D., & Cassell, J. A. (2016). Extracting information from the text of electronic medical records to improve case detection: A systematic review. *Journal of the American Medical Informatics Association* https://doi.org/10.1093/jamiaopen/ooae044

Galkin, M., Berrendorf, M., & Hoyt, C. T. (2022). An open challenge for inductive link prediction on knowledge graphs. arXiv:2203.01520

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics*, *50*(3), 1097–1179. https://doi.org/10.1162/COLI_A_00524

Ganguly, S., & Pudi, V. (2017). Paper2vec: Combining graph and text information for scientific paper representation. In *Advances in information retrieval*. https://doi.org/10.1007/978-3-319-56608-5_30

Gao, H., Zhang, C., Wu, F., Zhao, J., Zheng, C., & Liu, H. (2024a). Bootstrapping heterogeneous graph representation learning via large language models: A generalized approach. arXiv:2412.08038

Gao, Y., Qiao, L., Kan, Z., Wen, Z., He, Y., & Li, D. (2024b). Two-stage generative question answering on temporal knowledge graph using large language models. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), *Findings of the association for computational linguistics: ACL 2024* (pp. 6719–6734). Bangkok, Thailand: Association for Computational Linguistics. https://aclanthology.org/2024.findings-acl.401/. https://doi.org/10.18653/v1/2024.findings-acl.401

García-Durán, A., Dumančić, S., & Niepert, M. (2018). Learning sequence encoders for temporal knowledge graph completion. arXiv:1809.03202

Gardent, C., Shimorina, A., Narayan, S., & Perez-Beltrachini, L. (2017). Creating training corpora for NLG micro-planning. In *55th annual meeting of the association for computational linguistics*. https://aclanthology.org/P17-1017/.

Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B. et al. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Research* https://doi.org/{10.1093/nar/gkr777}.

Gesese, G. A., Sack, H., & Alam, M. (2022). RAILD: Towards leveraging relation features for inductive link prediction in knowledge graphs. In *Proceedings of the 11th International joint conference on knowledge graphs* (pp. 82–90). ACM. https://doi.org/10.1145/3579051.3579066

Ghasemi-Gol, M., & Szekely, P. (2018). TabVec: Table vectors for classification of web tables. arXiv:1802.06290

Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998). CiteSeer: An automatic citation indexing system. In *Proceedings of the third ACM conference on digital libraries*. https://doi.org/{10.1145/276675.276685}.

Goyal, P., Kamra, N., He, X., & Liu, Y. (2018). DynGEM: Deep embedding method for dynamic graphs. arXiv:1805.11273

Gratch, J., Artstein, R., Lucas, G. M., Stratou, G., Scherer, S., Nazarian, A., Wood, R., Boberg, J., DeVault, D., Marsella, S. et al. (2014). The distress analysis interview corpus of human and computer interviews. In *LREC* (pp. 3123–3128). Reykjavik (vol. 14). http://www.lrec-conf.org/proceedings/lrec2014/pdf/508_Paper.pdf.

Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International conference on knowledge discovery and data mining* (pp. 855–864). https://doi.org/10.1145/2939672.2939754

Gu, J., Jiang, X., Shi, Z., Tan, H., Zhai, X., Xu, C., Li, W., Shen, Y., Ma, S., Liu, H. et al. (2024a). A survey on LLM-as-a-judge. arXiv:2411.15594

Gu, Y., Kase, S., Vanni, M., Sadler, B., Liang, P., Yan, X., & Su, Y. (2021). Beyond I.I.D.: Three levels of generalization for question answering on knowledge bases. In *Proceedings of the web conference 2021* (pp. 3477–3488). https://doi.org/10.1145/3442381.3449992

Gu, Y., Xu, Z., & Yang, C. (2024b). Empowering graph neural network-based computational drug repositioning with large language model-inferred knowledge representation. *Interdisciplinary Sciences: Computational Life Sciences*, (pp. 1–18). https://doi.org/{10.1007/s12539-024-00654-7}.

Guo, J., Du, L., Liu, H., Zhou, M., He, X., & Han, S. (2023). GPT4Graph: Can large language models understand graph structured data? an empirical evaluation and benchmarking. arXiv:2305.15066

Guo, Z., Xia, L., Yu, Y., Wang, Y., Yang, Z., Wei, W., Pang, L., Chua, T.-S., & Huang, C. (2024). GraphEdit: Large language models for graph structure learning. arXiv:2402.15183

Hamilton, W. L., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. In *Proceedings of the neural information processing systems* (pp. 1024–1034). https://proceedings.neurips.cc/paper/2017/hash/5dd9db5e033da9c6fb5ba83c7a7ebea9-Abstract.html.

Han, Y., Huang, G., Song, S., Yang, L., Wang, H., & Wang, Y. (2022). Dynamic neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(11), 7436–7456. https://doi.org/10.1109/TPAMI.2021.3117837

Harper, F. M., & Konstan, J. A. (2015). The MovieLens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TIIS)* https://doi.org/{10.1145/2827872}.

Hassani, K. (2022). Cross-domain few-shot graph classification. In *Proceedings of the AAAI conference on artificial intelligence*. https://doi.org/{10.1609/aaai.v36i6.20642}.

Hatem, S., Khoriba, G., Gad-Elrab, M. H., & ElHelw, M. (2024). Up to date: Automatic updating knowledge graphs using LLMs. *Procedia Computer Science*, *244*, 327–334. 6th International Conference on AI in Computational Linguistics https://doi.org/10.1016/j.procs.2024.10.206

He, X., Bresson, X., Laurent, T., Perold, A., LeCun, Y., & Hooi, B. (2024a). Harnessing explanations: LLM-to-LM interpreter for enhanced text-attributed graph representation learning. In *International conference on learning representations*. https://openreview.net/forum?id=RXFVcynVe1.

He, X., & Wang, X. E. (2023). Multimodal graph transformer for multimodal question answering. In *Proceedings of the 17th conference of the European chapter of the association for computational linguistics*. https://doi.org/{10.18653/v1/2023.eacl-main.15}.

He, Y., & Hooi, B. (2024). UniGraph: Learning a cross-domain graph foundation model from natural language. arXiv:2402.13630

He, Y., Sui, Y., He, X., Liu, Y., Sun, Y., & Hooi, B. (2025). UniGraph2: Learning a unified embedding space to bind multimodal graphs. arXiv:2502.00806

He, Z., Zhu, J., Qian, S., Chai, J., & Koutra, D. (2024b). LinkGPT: Teaching large language models to predict missing links. arXiv:2406.04640

Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* https://doi.org/{10.1109/JSTARS.2019.2918242}.

Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. (2016). Session-based recommendations with recurrent neural networks. In Y. Bengio, & Y. LeCun (Eds.), *4th International conference on learning representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, conference track proceedings*. arXiv:1511.06939

Honerkamp, D., Büchner, M., Despinoy, F., Welschehold, T., & Valada, A. (2024). Language-grounded dynamic scene graphs for interactive object search with mobile manipulation. *IEEE Robotics and Automation Letters*, *9*(10), 8298–8305. https://doi.org/10.1109/LRA.2024.3441495

Hou, W., Zhao, W., Jia, N., & Liu, X. (2025). Low-resource knowledge graph completion based on knowledge distillation driven by large language models. *Applied Soft Computing*, *169*, 112622. https://doi.org/{10.1016/j.asoc.2024.112622}.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W. et al. (2022). LoRA: Low-rank adaptation of large language models. *ICLR*, *1*(2), 3. https://openreview.net/forum?id=nZeVKeeFYf9.

Hu, K., Wang, Z., Martens, K. A. E., Hagenbuchner, M., Bennamoun, M., Tsoi, A. C., & Lewis, S. J. G. (2021). Graph fusion network-based multimodal learning for freezing of gait detection. *IEEE Transactions on Neural Networks and Learning Systems* https://doi.org/{10.1109/TNNLS.2021.3105602}.

Hu, W., Fey, M., Zitnik, M., Dong, Y., Ren, H., Liu, B., Catasta, M., & Leskovec, J. (2020). Open graph benchmark: Datasets for machine learning on graphs. *Advances in Neural Information Processing Systems*, *33*, 22118–22133.

Hu, Y., Zou, F., Han, J., Sun, X., & Wang, Y. (2024). LLM-TIKG: Threat intelligence knowledge graph construction utilizing large language model. *Computers & Security*, *145*, 103999. https://doi.org/{10.1016/j.cose.2024.103999}.

Huang, X., Han, K., Yang, Y., Bao, D., Tao, Q., Chai, Z., & Zhu, Q. (2024). Can GNN be good adapter for LLMs? In *Proceedings of the ACM web conference 2024* (pp. 893–904). https://doi.org/10.1145/3589334.3645627

Huang, X., Song, Q., Li, Y., & Hu, X. (2019). Graph recurrent networks with attributed random walks. In *Proceedings of the ACM SIGKDD International conference on knowledge discovery & data mining* (pp. 732–740). https://doi.org/10.1145/3292500.3330941

Iga, V. I. R., & Silaghi, G. C. (2024). Assessing LLMs suitability for knowledge graph completion. In *International conference on neural-symbolic learning and reasoning* (pp. 277–290). Springer. https://doi.org/10.1007/978-3-031-71170-1_22

Jeong, S., Park, J., Choi, M., Kwon, Y., & Lim, S. (2025). LLM-powered scene graph representation learning for image retrieval via visual triplet-based graph transformation. *Expert Systems with Applications*, (p. 127926). https://doi.org/10.1016/J.ESWA.2025.127926

Jia, Z., Pramanik, S., Saha Roy, R., & Weikum, G. (2021). Complex temporal question answering on knowledge graphs. In *Proceedings of the 30th ACM International conference on information & knowledge management, CIKM '21* (p. 792-802). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3459637.3482416

Jin, B., Zhang, W., Zhang, Y., Meng, Y., Zhao, H., & Han, J. (2023a). Learning multiplex representations on text-attributed graphs with one language model encoder. arXiv:2310.06684

Jin, B., Zhang, Y., Zhu, Q., & Han, J. (2023b). HeterFormer: Transformer-based deep node representation learning on heterogeneous text-rich networks. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*. https://doi.org/10.1145/3580305.3599376

Jin, Q., Dhingra, B., Liu, Z., Cohen, W., & Lu, X. (2019). PubmedQA: A dataset for biomedical research question answering. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th International joint conference on natural language processing (EMNLP-IJCNLP)*. https://aclanthology.org/D19-1259/

Jin, R., Li, Y., Qi, G., Hu, N., Li, Y.-F., Chen, J., Wang, J., Chen, Y., Min, D., & Bi, S. (2024). HeGTa: Leveraging heterogeneous graph-enhanced large language models for few-shot complex table understanding. arXiv:2403.19723

Jin, W., Li, Y., Xu, H., Wang, Y., & Tang, J. (2020a). Adversarial attacks and defenses on graphs: A review and empirical study. arXiv:2003.00653

Jin, W., Qu, M., Jin, X., & Ren, X. (2020b). Recurrent event network: Autoregressive structure inferenceover temporal knowledge graphs. In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 6669–6683). Online: Association for Computational Linguistics. . https://doi.org/10.18653/v1/2020.emnlp-main.541

Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L.-w. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data* https://doi.org/{10.1038/sdata.2016.35}.

Ju, W., Fang, Z., Gu, Y., Liu, Z., Long, Q., Qiao, Z., Qin, Y., Shen, J., Sun, F., Xiao, Z. et al. (2024a). A comprehensive survey on deep graph representation learning. *Neural Networks*, *173*, 106207. https://doi.org/10.1016/J.NEUNET.2024.106207

Ju, W., Mao, Z., Yi, S., Qin, Y., Gu, Y., Xiao, Z., Shen, J., Qiao, Z., & Zhang, M. (2024b). Cluster-guided contrastive class-imbalanced graph classification. arXiv:2412.12984

Ju, W., Yi, S., Wang, Y., Xiao, Z., Mao, Z., Li, H., Gu, Y., Qin, Y., Yin, N., Wang, S. et al. (2024c). A survey of graph neural networks in real world: Imbalance, noise, privacy and OOD challenges. arXiv:2403.04468

Kazemi, S. M., Goel, R., Jain, K., Kobyzev, I., Sethi, A., Forsyth, P., & Poupart, P. (2020). Representation learning for dynamic graphs: A survey. *Journal of Machine Learning Research*, *21*(70), 1–73. http://jmlr.org/papers/v21/19-447.html.

Khoshraftar, S., & An, A. (2024). A survey on graph representation learning methods. *ACM Transactions on Intelligent Systems and Technology*, *15*(1), 1–55. https://doi.org/10.1145/3633518

Kipf, T. N., & Welling, M. (2016). Variational graph auto-encoders. arXiv:1611.07308

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *International conference on learning representations*.

Ko, Y., Yang, H., Kim, T., & Kim, H. (2024). Subgraph-aware training of language models for knowledge graph completion using structure-aware contrastive learning. arXiv:2407.12703

Koa, K. J. L., Ma, Y., Ng, R., Zheng, H., & Chua, T.-S. (2024). Temporal relational reasoning of large language models for detecting stock portfolio crashes. arXiv:2410.17266

Koncel-Kedziorski, R., Bekal, D., Luan, Y., Lapata, M., & Hajishirzi, H. (2019). Text generation from knowledge graphs with graph transformers. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics*. https://aclanthology.org/N19-1238/.

Krause, J., Stark, M., Deng, J., & Fei-Fei, L. (2013). 3D object representations for fine-grained categorization. In *Proceedings of the IEEE International conference on computer vision workshops*. https://doi.ieeecomputersociety.org/10.1109/ICCVW.2013.77.

Kumar, V., Hua, Y., Ramakrishnan, G., Qi, G., Gao, L., & Li, Y.-F. (2019). Difficulty-controllable multi-hop question generation from knowledge graphs. In *International semantic web conference* (pp. 382–398). Springer. https://doi.org/10.1007/978-3-030-30793-6_22

Lafferty, J., McCallum, A., Pereira, F. et al. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML* (p. 3). Williamstown, MA (vol. *1*). https://doi.org/10.5555/645530.655813

Leblay, J., & Chekol, M. W. (2018). Deriving validity time in knowledge graph. In *Companion proceedings of the web conference 2018, WWW '18* (p. 1771-1776). Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3184558.3191639

Lee, C.-Y., Li, C.-L., Zhang, H., Dozat, T., Perot, V., Su, G., Zhang, X., Sohn, K., Glushnev, N., Wang, R. et al. (2023a). FormNetV2: Multimodal graph contrastive learning for form document information extraction. In *Proceedings of the 61st annual meeting of the association for computational linguistics*. https://doi.org/{10.18653/v1/2023.acl-long.501}.

Lee, D.-H., Ahrabian, K., Jin, W., Morstatter, F., & Pujara, J. (2023b). Temporal knowledge graph forecasting without knowledge using in-context learning. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 conference on empirical methods in natural language processing* (pp. 544–557). Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.36

Leetaru, K., & Schrodt, P. A. (2013). GDELT: Global data on events, location, and tone, 1979–2012. In *ISA annual convention* (pp. 1–49). Citeseer (vol. *2*).

Li, C., Xia, F., Martín-Martín, R., Lingelbach, M., Srivastava, S., Shen, B., Vainio, K., Gokmen, C., Dharan, G., Jain, T. et al. (2021a). iGibson 2.0: Object-centric simulation for robot learning of everyday household tasks. arXiv:2108.03272

Li, D., Wang, Y., Funakoshi, K., & Okumura, M. (2023a). Joyful: Joint modality fusion and graph contrastive learning for multimoda emotion recognition. In *Proceedings of the 2023 conference on empirical methods in natural language processing*. https://doi.org/{10.18653/v1/2023.emnlp-main.996}.

Li, J., Li, D., Savarese, S., & Hoi, S. (2023b). BLIP-2: Bootstrapping language-image pretraining with frozen image encoders and large language models. In *International conference on machine learning*. https://proceedings.mlr.press/v202/li23q.

Li, J., Tang, T., Zhao, W. X., Nie, J.-Y., & Wen, J.-R. (2024a). Pre-trained language models for text generation: A survey. *ACM Computing Surveys*, *56*(9), 1–39. https://doi.org/{10.1145/3649449}.

Li, M., Chen, J., Li, B., Zhang, Y., Zhang, R., Gong, S., Ma, X., & Tian, Z. (2024b). TDG-Mamba: Advanced spatiotemporal embedding for temporal dynamic graph learning via bidirectional information propagation. *IEEE Transactions on Computational Social Systems*, (pp. 1–16). https://doi.org/10.1109/TCSS.2024.3509399

Li, M., Chen, J., Yu, C., Jiang, G., Zhang, R., Shen, Y., & Song, H. H. (2025a). Topology-driven attribute recovery for attribute missing graph learning in social internet of things. *IEEE Internet of Things Journal* arXiv:2501.10151

Li, M., Ding, C., Chen, J., Xing, W., Ye, C., Zhang, R., Zhuang, S., Hu, J., Qiu, T. Z., & Gao, H. (2025b). AttriReBoost: A gradient-free propagation optimization method for cold start mitigation in attribute missing graphs. arXiv:2501.00743

Li, M., Zhang, R., Zhang, Y., Piao, X., Zhao, S., & Yin, B. (2024c). SCAE: Structural contrastive auto-encoder for incomplete multi-view representation learning. *ACM Transactions on Multimedia Computing, Communications and Applications*, *20*(9), 1–24. https://doi.org/10.1145/3672078

Li, M., Zhang, Y., Li, X., Zhang, Y., & Yin, B. (2023c). Hypergraph transformer neural networks. *ACM Transactions on Knowledge Discovery from Data*, *17*(5). https://doi.org/10.1145/3565028

Li, M., Zhang, Y., Wang, S., Hu, Y., & Yin, B. (2025c). Redundancy is not what you need: An embedding fusion graph auto-encoder for self-supervised graph representation learning. *IEEE Transactions on Neural Networks and Learning Systems*, *36*(2), 3519–3533. https://doi.org/10.1109/TNNLS.2024.3357080

Li, M., Zhang, Y., Zhang, W., Zhao, S., Piao, X., & Yin, B. (2024d). CSAT: Contrastive sampling-aggregating transformer for community detection in attribute-missing networks. *IEEE Transactions on Computational Social Systems*, *11*(2), 2277–2290. https://doi.org/10.1109/TCSS.2023.3292145

Li, P., Wang, X., Zhang, Z., Meng, Y., Shen, F., Li, Y., Wang, J., Li, Y., & Zhu, W. (2024e). RealTCD: Temporal causal discovery from interventional data with large language model. In *Proceedings of the 33rd ACM International conference on information and knowledge management, CIKM '24* (p. 4669-4677). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3627673.3680042

Li, Q., Chen, Z., Ji, C., Jiang, S., & Li, J. (2024f). LLM-based multi-level knowledge generation for few-shot knowledge graph completion. In *Proceedings of the International joint conference on artificial intelligence* (pp. 2135–2143). https://www.ijcai.org/proceedings/2024/236.

Li, W.-Z., Wang, C.-D., Xiong, H., & Lai, J.-H. (2023d). GraphSHA: Synthesizing harder samples for class-imbalanced node classification. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 1328–1340).

Li, X., Lian, D., Lu, Z., Bai, J., Chen, Z., & Wang, X. (2024g). GraphAdapter: Tuning vision-language models with dual knowledge graph. In *Advances in neural information processing systems*. (vol. *36*). https://openreview.net/forum?id=YmEDnMynuO&noteId=0rFYtJNqHc.

Li, X. V., & Sanna Passino, F. (2024). FinDKG: Dynamic knowledge graphs with large language models for detecting global trends in financial markets. In *Proceedings of the*

5th ACM International conference on AI in finance, ICAIF '24 (p. 573-581). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3677052.3698603

Li, Y., Yang, Y., Zhu, J., Chen, H., & Wang, H. (2024h). LLM-empowered few-shot node classification on incomplete graphs with real node degrees. In *Proceedings of the ACM International conference on information and knowledge management* (pp. 1306–1315). https://doi.org/10.1145/3627673.3679861

Li, Z., Cheng, L., Zhang, C., Zhu, X., & Zhao, H. (2023e). Multi-source education knowledge graph construction and fusion for college curricula. In *International conference on advanced learning technologies.* https://doi.org/{10.1109/ICALT58122.2023.00111}.

Li, Z., Fan, S., Gu, Y., Li, X., Duan, Z., Dong, B., Liu, N., & Wang, J. (2024i). FlexKBQA: A flexible LLM-powered framework for few-shot knowledge base question answering. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 18608–18616). (vol. *38*). https://doi.org/10.1609/AAAI.V38I17.29823

Li, Z., Jin, X., Li, W., Guan, S., Guo, J., Shen, H., Wang, Y., & Cheng, X. (2021b). Temporal knowledge graph reasoning based on evolutional representation learning. In *Proceedings of the 44th International ACM SIGIR conference on research and development in information retrieval* (pp. 408–417). https://doi.org/10.1145/3404835.3462963

Liao, J., Li, J., Chen, L., Wu, B., Bian, Y., & Zheng, Z. (2023). SAILOR: Structural augmentation based tail node representation learning. In *Proceedings of the 32nd ACM International conference on information and knowledge management* (pp. 1389–1399).

Liao, R., Jia, X., Li, Y., Ma, Y., & Tresp, V. (2024). GenTKG: Generative forecasting on temporal knowledge graph with large language models. In K. Duh, H. Gomez, & S. Bethard (Eds.), *Findings of the association for computational linguistics: NAACL 2024* (pp. 4303–4317). Mexico City, Mexico: Association for Computational Linguistics. . https://doi.org/10.18653/v1/2024.findings-naacl.268

Liao, R., Li, Y., Song, Y., Wang, S., Hamilton, W. L., Duvenaud, D., Urtasun, R., & Zemel, R. (2019). Efficient graph generation with graph recurrent attention networks. In *Proceedings of the 33rd International conference on neural information processing systems* , pp. . Red Hook, NY, USA: Curran Associates Inc. https://proceedings.neurips.cc/paper/2019/hash/d0921d442ee91b896ad95059d13df618-Abstract.html.

Liao, X., Chen, C., Wang, Z., Liu, Y., Wang, T., & Cheng, L. (2025). Large language model assisted fine-grained knowledge graph construction for robotic fault diagnosis. *Advanced Engineering Informatics*, *65*, 103134. https://doi.org/{10.1016/j.aei.2025.103134}.

Lin, B. Y., Wu, Z., Yang, Y., Lee, D.-H., & Ren, X. (2021). RiddleSense: Reasoning about riddle questions featuring linguistic creativity and commonsense knowledge. In *Findings of the association for computational linguistics: ACL-IJCNLP 2021.* https://aclanthology.org/2021.findings-acl.131/.

Lin, Z., Liang, B., Long, Y., Dang, Y., Yang, M., Zhang, M., & Xu, R. (2022). Modeling intra- and inter-modal relations: Hierarchical graph contrastive learning for multimodal sentiment analysis. In *Proceedings of the 29th International conference on computational linguistics.* https://aclanthology.org/2022.coling-1.622/.

Liu, C., & Wu, B. (2023). Evaluating large language models on graphs: Performance insights and comparative analysis. arXiv:2308.11224

Liu, H., Feng, J., Kong, L., Liang, N., Tao, D., Chen, Y., & Zhang, M. (2024a). One for all: Towards training one graph model for all classification tasks. In *International conference on learning representations.* https://openreview.net/forum?id=4IT2pgc9v6.

Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2024b). Visual instruction tuning. In *Advances in neural information processing systems.* https://openreview.net/forum?id=w0H2xGHlkw.

Liu, J., He, M., Wang, G., Hung, N. Q. V., Shang, X., & Yin, H. (2023a). Imbalanced node classification beyond homophilic assumption. arXiv:2304.14635

Liu, J., Yang, C., Lu, Z., Chen, J., Li, Y., Zhang, M., Bai, T., Fang, Y., Sun, L., Yu, P. S., & Shi, C. (2025). Graph foundation models: Concepts, opportunities and challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence* https://doi.org/{10.1109/TPAMI.2025.3548729}.

Liu, M., Zhang, Z., Tang, J., Bu, J., He, B., & Zhou, S. (2024c). Revisiting, benchmarking and understanding unsupervised graph domain adaptation. In *Advances in neural information processing systems.*

Liu, S., Li, T., Feng, Y., Tran, N., Zhao, H., Qiu, Q., & Li, P. (2023b). Structural re-weighting improves graph domain adaptation. In *International conference on machine learning.* https://proceedings.mlr.press/v202/liu23u.html.

Liu, M., Zhang, Z., Tang, J., Bu, J., He, B., & Zhou, S., (2024d). *Proceedings of the 38th International Conference on Neural Information Processing Systems*, *37*, 89408–89436. http://papers.nips.cc/paper_files/paper/2024/hash/a2c548b2ed3141400d86a235eaa6fad0-Abstract-Datasets_and_Benchmarks_Track.html

Liu, Y., Ao, X., Qin, Z., Chi, J., Feng, J., Yang, H., & He, Q. (2021). Pick and choose: A GNN-based imbalanced learning approach for fraud detection. In *Proceedings of the web conference 2021* (pp. 3168–3177). https://doi.org/10.1145/3442381.3449989

Liu, Y., Gao, Z., Liu, X., Luo, P., Yang, Y., & Xiong, H. (2023c). QTIAH-GNN: Quantity and topology imbalance-aware heterogeneous graph neural network for bankruptcy prediction. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 1572–1582). https://doi.org/10.1145/3580305.3599479

Liu, Y., Tian, X., Sun, Z., & Hu, W. (2024e). Finetuning generative large language models with discrimination instructions for knowledge graph completion. In *International semantic web conference* (pp. 199–217). Springer. https://doi.org/10.1007/978-3-031-77844-5_11

Liu, Z., Li, Y., Chen, N., Wang, Q., Hooi, B., & He, B. (2023d). A survey of imbalanced learning on graphs: Problems, techniques, and future directions. arXiv:2308.13821

Liu, Z., Mao, Q., Liu, C., Fang, Y., & Sun, J. (2022). On size-oriented long-tailed graph classification of graph neural networks. In *Proceedings of the ACM web conference 2022* (pp. 1506–1516). https://doi.org/{10.1145/3485447.3512197}.

Liu, Z., Nguyen, T.-K., & Fang, Y. (2023e). On generalized degree fairness in graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 4525–4533). (vol. *37*). https://doi.org/10.1609/AAAI.V37I4.25574

Lucic, A., Ter Hoeve, M. A., Tolomei, G., De Rijke, M., & Silvestri, F. (2022). CF-GNNExplainer: Counterfactual explanations for graph neural networks. In *International conference on artificial intelligence and statistics* (pp. 4499–4511). PMLR. https://proceedings.mlr.press/v151/lucic22a.html.

Luo, K., Bai, Y., Gao, C., Si, S., Shen, Y., Liu, Z., Wang, Z., Kong, C., Li, W., Huang, Y. et al. (2025). GLTW: Joint improved graph transformer and llm via three-word language for knowledge graph completion. arXiv:2502.11471

Lv, X., Lin, Y., Cao, Y., Hou, L., Li, J., Liu, Z., Li, P., & Zhou, J. (2022). Do pre-trained models benefit knowledge graph completion? A reliable evaluation and a reasonable approach. In *Findings of the association for computational linguistics.* Association for Computational Linguistics. https://doi.org/10.18653/V1/2022.FINDINGS-ACL.282

Lyu, Q., Havaldar, S., Stein, A., Zhang, L., Rao, D., Wong, E., Apidianaki, M., & Callison-Burch, C. (2023). Faithful chain-of-thought reasoning. In *IJCNLP-AACL.* https://doi.org/10.18653/V1/2023.IJCNLP-MAIN.20

Lyu, Y., Zheng, X., Zhou, J., & Wang, L. (2024). UniBind: LLM-augmented unified and balanced representation space to bind them all. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* https://openaccess.thecvf.com/content/CVPR2024/html/Lyu_UniBind_LLM-Augmented_Unified_and_Balanced_Representation_Space_to_Bind_Them_CVPR_2024_paper.html.

Ma, T., Qian, Y., Wang, Z., Zhang, Z., Zhang, C., & Ye, Y. (2025). LLM-empowered class imbalanced graph prompt learning for online drug trafficking detection. arXiv:2503.01900

Ma, Y., Tian, Y., Moniz, N., & Chawla, N. V. (2023). Class-imbalanced learning on graphs: A survey. *ACM Computing Surveys* https://doi.org/{10.1145/3718734}.

Mahdisoltani, F., Biega, J., & Suchanek, F. M. (2013). YAGO3: A knowledge base from multilingual wikipedias. In *CIDR, Asilomar, United States.* https://imt.hal.science/hal-01699874.

Maji, S., Rahtu, E., Kannala, J., Blaschko, M., & Vedaldi, A. (2013). Fine-grained visual classification of aircraft. arXiv:1306.5151

Mao, Z., Ju, W., Qin, Y., Luo, X., & Zhang, M. (2023). RAHNet: Retrieval augmented hybrid network for long-tailed graph classification. In *Proceedings of the 31st ACM International conference on multimedia* (pp. 3817–3826). https://doi.org/10.1145/3581783.3612360

Mao, Z., Ju, W., Yi, S., Wang, Y., Xiao, Z., Long, Q., Yin, N., Liu, X., & Zhang, M. (2025). Learning knowledge-diverse experts for long-tailed graph classification. *ACM Transactions on Knowledge Discovery from Data*, *19*(2), 1–24. https://doi.org/10.1145/3705323

Margatina, K., Wang, S., Vyas, Y., Anna John, N., Benajiba, Y., & Ballesteros, M. (2023). Dynamic benchmarking of masked language models on temporal concept drift with multiple views. In A. Vlachos, & I. Augenstein (Eds.), *Proceedings of the 17th conference of the European chapter of the association for computational linguistics* (pp. 2881–2898). Dubrovnik, Croatia: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.eacl-main.211

Mashima, Y., Tanigawa, M., & Yokoi, H. (2024). Information heterogeneity between progress notes by physicians and nurses for inpatients with digestive system diseases. *Scientific Reports* https://doi.org/10.1038/s41598-024-56324-7}.

McAuley, J., Targett, C., Shi, Q., & Van Den Hengel, A. (2015). Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR conference on research and development in information retrieval* (pp. 43–52). https://doi.org/10.1145/2766462.2767755

McCallum, A. K., Nigam, K., Rennie, J., & Seymore, K. (2000). Automating the construction of internet portals with machine learning. *Information Retrieval* https://doi.org/{10.1023/A:1009953814988}.

Mernyei, P., & Cangea, C. (2020). Wiki-CS: A wikipedia-based benchmark for graph neural networks. arXiv:2007.02901

Mihaylov, T., Clark, P., Khot, T., & Sabharwal, A. (2018). Can a suit of armor conduct electricity? a new dataset for open book question answering. In *Proceedings of the 2018 conference on empirical methods in natural language processing.* https://aclanthology.org/D18-1260/.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems.* https://papers.nips.cc/paper_files/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html.

Mitchell, T., Cohen, W., Hruschka, E., Talukdar, P., Yang, B., Betteridge, J., Carlson, A., Dalvi, B., Gardner, M., Kisiel, B. et al. (2018). Never-ending learning. *Communications of the ACM*, *61*(5), 103–115. https://doi.org/10.1145/3191513

Morales-Alvarez, P., Gong, W., Lamb, A., Woodhead, S., Jones, S. P., Pawlowski, N., Allamanis, M., & Zhang, C. (2022). Simultaneous missing value imputation and structure learning with groups. In *Proceedings of the International conference on neural information processing systems* (pp. 20011–20024). http://papers.nips.cc/paper_files/paper/2022/hash/7e57131fdeb815764434b65162c88895-Abstract-Conference.html.

Moreo, A., Pedrotti, A., & Sebastiani, F. (2022). Generalized funnelling: Ensemble learning and heterogeneous document embeddings for cross-lingual text classification. *ACM Transactions on Information Systems* https://doi.org/{10.1145/3544104}.

Moslem, Y., Haque, R., Kelleher, J. D., & Way, A. (2023). Adaptive machine translation with large language models. arXiv:2301.13294

Motie, S., & Raahemi, B. (2024). Financial fraud detection using graph neural networks: A systematic review. *Expert Systems with Applications*, *240*, 122156. https://doi.org/10.1016/J.ESWA.2023.122156

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2023). A comprehensive overview of large language models. arXiv:2307.06435

Netflix, Inc. (2025). Netflix prize data. https://www.kaggle.com/datasets/netflix-inc/netflix-prize-data.

Ng, A., Jordan, M., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. *Proceedings of the International conference on neural information processing systems: natural and synthetic*, *14*. https://proceedings.neurips.cc/paper_files/paper/2001/file/801272ee79cfde7fa5960571fee36b9b-Paper.pdf.

Ni, J., Li, J., & McAuley, J. (2019). Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th International joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 188–197). https://aclanthology.org/D19-1018/. https://doi.org/10.18653/v1/D19-1018

Ni, L., Wang, S., Zhang, Z., Li, X., Zheng, X., Denny, P., & Liu, J. (2024). Enhancing student performance prediction on learnersourced questions with SGNN-LLM synergy. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 23232–23240). (vol. *38*). https://doi.org/{10.1609/aaai.v38i21.30370}.

Nilsback, M.-E., & Zisserman, A. (2008). Automated flower classification over a large number of classes. In *2008 sixth Indian conference on computer vision, graphics & image processing*. https://doi.org/{10.1109/ICVGIP.2008.47}.

Ong, R., Sun, J., Guo, Y.-K., & Serban, O. (2025). Dynamic link prediction: Using language models and graph structures for temporal knowledge graph completion with emerging entities and relations. *Expert Systems with Applications*, (p. 126648). https://doi.org/10.1016/J.ESWA.2025.126648

Opsahl, T., & Panzarasa, P. (2009). Clustering in weighted networks. *Social Networks*, *31*(2), 155–163. https://doi.org/10.1016/J.SOCNET.2009.02.002

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A. et al. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems, 35*, 27730–27744.

Pan, B., Zhang, Z., Zhang, Y., Hu, Y., & Zhao, L. (2024). Distilling large language models for text-attributed graph learning. In *Proceedings of the 33rd ACM International conference on information and knowledge management* (pp. 1836–1845). https://doi.org/{10.1145/3627673.3679830}.

Pan, S., Hu, R., Long, G., Jiang, J., Yao, L., & Zhang, C. (2018). Adversarially regularized graph autoencoder for graph embedding. arXiv:1802.04407

Paranjape, A., Benson, A. R., & Leskovec, J. (2017). Motifs in temporal networks. In *Proceedings of the tenth ACM International conference on web search and data mining, WSDM '17* (p. 601-610). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3018661.3018731

Pareja, A., Domeniconi, G., Chen, J., Ma, T., Suzumura, T., Kanezashi, H., Kaler, T., Schardl, T., & Leiserson, C. (2020). EvolveGCN: Evolving graph convolutional networks for dynamic graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(04), 5363–5370. https://doi.org/10.1609/aaai.v34i04.5984

Parkhi, O. M., Vedaldi, A., Zisserman, A., & Jawahar, C. V. (2012). Cats and dogs. In *2012 IEEE conference on computer vision and pattern recognition*. https://doi.org/{10.1109/CVPR.2012.6248092}.

Parrish, A., Chen, A., Nangia, N., Padmakumar, V., Phang, J., Thompson, J., Htut, P. M., & Bowman, S. R. (2022). BBQ: A hand-built bias benchmark for question answering. In *Findings of the association for computational linguistics* (pp. 2086–2105). https://doi.org/10.18653/V1/2022.FINDINGS-ACL.165

Pasupat, P., & Liang, P. (2015). Compositional semantic parsing on semi-structured tables. arXiv:1508.00305

Pei, H., Wei, B., Chang, K. C.-C., Lei, Y., & Yang, B. (2020). Geom-GCN: Geometric graph convolutional networks. In *International conference on learning representations*. https://openreview.net/forum?id=S1e2agrFvS.

Peng, C., He, J., & Xia, F. (2024a). Learning on multimodal graphs: A survey. arXiv:2402.05322

Peng, X., Cheng, J., Tang, X., Zhang, B., & Tu, W. (2024b). Multi-view graph imputation network. *Information Fusion*, *102*, 102024. https://doi.org/10.1016/J.INFFUS.2023.102024

Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International conference on knowledge discovery and data mining*. https://doi.org/{10.1145/2623330.2623732}.

Perozzi, B., Fatemi, B., Zelle, D., Tsitsulin, A., Kazemi, M., Al-Rfou, R., & Halcrow, J. (2024). Let your graph do the talking: Encoding structured data for LLMs. arXiv:2402.05862

Petermann, A., Junghanns, M., Müller, R., & Rahm, E. (2014). Graph-based data integration and business intelligence with BIIIG. *Proceedings of the VLDB Endowment* https://doi.org/10.14778/2733004.2733034}.

Pilancı, M., & Vural, E. (2020). Domain adaptation on graphs by learning aligned graph bases. *IEEE Transactions on Knowledge and Data Engineering* https://doi.org/{10.1109/TKDE.2020.2984212}.

Prithviraj, S., Namata, G., Bilgic, M., Getoor, L., Galligher, B., & Eliassi-Rad, T. (2008). Collective classification in network data. *AI Magazine* https://doi.org/{10.1609/aimag.v29i3.2157}.

Qian, X., Zhang, Y., Zhao, Y., Zhou, B., Sui, X., Zhang, L., & Song, K. (2024). Timer⁴ : Time-aware retrieval-augmented large language models for temporal knowledge graph question answering. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 conference on empirical methods in natural language processing* (pp. 6942–6952). Miami, Florida, USA: Association for Computational Linguistics. https://aclanthology.org/2024.emnlp-main.394/. https://doi.org/10.18653/v1/2024.emnlp-main.394

Qian, Y., Zhang, Y., Ye, Y., & Zhang, C. (2021). Distilling meta knowledge on heterogeneous graph for illicit drug trafficker detection on social media. In *Proceedings of the International conference on neural information processing systems* (pp. 26911–26923). https://proceedings.neurips.cc/paper/2021/hash/e234e195f3789f05483378c397db1cb5-Abstract.html.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257–286. https://doi.org/{10.1109/5.18626}.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J. et al. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning*. https://proceedings.mlr.press/v139/radford21a.

Rahmani, S., Baghbani, A., Bouguila, N., & Patterson, Z. (2023). Graph neural networks for intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems, 24*(8), 8846–8885. https://doi.org/10.1109/TITS.2023.3257759

Rajput, N., & Singh, K. (2022). Temporal graph learning for financial world: Algorithms, scalability, explainability & fairness. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 4818–4819). https://doi.org/10.1145/3534678.3542619

Raleigh, C., Linke, r., Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An armed conflict location and event dataset. *Journal of Peace Research, 47*(5), 651–660. . https://doi.org/10.1177/0022343310378914

Rampášek, L., Galkin, M., Dwivedi, V. P., Luu, A. T., Wolf, G., & Beaini, D. (2022). Recipe for a general, powerful, scalable graph transformer. In *Proceedings of the neural information processing systems* (pp. 14501–14515). (vol. *35*). http://papers.nips.cc/paper_files/paper/2022/hash/5d4834a159f1547b267a05a4e2b7cf5e-Abstract-Conference.html.

Ren, L., Hu, R., Wang, Z., Xiao, Y., Li, D., Wu, J., Zang, Y., Hu, J., & Huang, Z. (2024a). Heterophilic graph invariant learning for out-of-distribution of fraud detection. In *Proceedings of the 32nd ACM International conference on multimedia* (pp. 11032–11040). https://doi.org/{10.1145/3664647.3681312}.

Ren, X., Tang, J., Yin, D., Chawla, N., & Huang, C. (2024b). A survey of large language models for graphs. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 6616–6626). https://doi.org/10.1145/3637528.3671460

Rosenthal, S., & McKeown, K. (2016). Social proof: The impact of author traits on influence detection. In *Proceedings of the first workshop on NLP and computational social science* (pp. 27–36). https://doi.org/10.18653/V1/W16-5604

Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., & Bronstein, M. (2020). Temporal graph networks for deep learning on dynamic graphs. arXiv:2006.10637

Rossi, E., Kenlay, H., Gorinova, M. I., Chamberlain, B. P., Dong, X., & Bronstein, M. M. (2022). On the unreasonable effectiveness of feature propagation in learning on graphs with missing node features. In *Learning on graphs conference* (pp. 1–11). PMLR. https://proceedings.mlr.press/v198/rossi22a.html.

Runfeng, X., Xiangyang, C., Zhou, Y., Xin, W., Zhanwei, X., Kai, Z. et al. (2023). LKPNR: LLM and KG for personalized news recommendation framework. arXiv:2308.12028

Salehi, A., & Davulcu, H. (2019). Graph attention auto-encoders. arXiv:1905.10715

Sankar, A., Wu, Y., Gou, L., Zhang, W., & Yang, H. (2020). DySAT: Deep neural representation learning on dynamic graphs via self-attention networks. In *Proceedings of the 13th International conference on web search and data mining, WSDM '20* (p. 519-527). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3336191.3371845

Saxena, A., Chakrabarti, S., & Talukdar, P. (2021). Question answering over temporal knowledge graphs. arXiv:2106.01515

Schäfer, M., Strohmeier, M., Lenders, V., Martinovic, I., & Wilhelm, M. (2014). Bringing up opensky: A large-scale ADS-b sensor network for research. In *IPSN-14 proceedings of the 13th International symposium on information processing in sensor networks* (pp. 83–94). https://doi.org/10.1109/IPSN.2014.6846743

Sehwag, U. M., Papasotiriou, K., Vann, J., & Ganesh, S. (b). In-context learning with topological information for LLM-based knowledge graph completion. In *ICML 2024 workshop on structured probabilistic inference {\&} generative modeling*. https://openreview.net/forum?id=eUpH8AuVQa.

Shang, W., Zhu, X., & Huang, X. (2024). Path-LLM: A shortest-path-based LLM learning for unified graph representation. arXiv:2408.05456

Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., Osmani, V. et al. (2019). Natural language processing of clinical notes on chronic diseases: Systematic review. *JMIR Medical Informatics* https://doi.org/{10.2196/12239}.

Shen, X., Dai, Q., Chung, F.-l., Lu, W., & Choi, K.-S. (2020). Adversarial deep network embedding for cross-network node classification. In *Proceedings of the AAAI conference on artificial intelligence*. https://doi.org/{10.1609/aaai.v34i03.5692}.

Shen, Y., Yang, H., & Lin, L. (2022). Automatic depression detection: An emotional audio-textual corpus and a GRU/BiLSTM-based model. In *ICASSP 2022-2022 IEEE International conference on acoustics, speech and signal processing (icassp)* (pp. 6247–6251). IEEE. . https://doi.org/10.1109/ICASSP43922.2022.9746569

Sheng, J., Guo, S., Chen, Z., Yue, J., Wang, L., Liu, T., & Xu, H. (2020). Adaptive attentional network for few-shot knowledge graph completion. In *Proceedings of the conference on Empirical methods in natural language processing* (pp. 1681–1691). Association for Computational Linguistics. https://doi.org/10.18653/V1/2020.EMNLP-MAIN.131

Shetty, J., & Adibi, J. (2004). The enron email dataset database schema and brief statistical report. *Information Sciences Institute Technical Report, University of Southern California, 4*(1), 120–128.

Shi, M., Tang, Y., Zhu, X., Wilson, D., & Liu, J. (2020). Multi-class imbalanced graph convolutional network learning. In *Proceedings of the twenty-ninth International joint conference on artificial intelligence (IJCAI-20)*. https://doi.org/10.24963/IJCAI.2020/398

Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., & Sun, Y. (2021). Masked label prediction: Unified message passing model for semi-supervised classification. In *Proceedings of the thirtieth International joint conference on artificial intelligence* (pp. 1548–1554). ijcai.org. https://doi.org/10.24963/IJCAI.2021/214

Shomer, H., Jin, W., Wang, W., & Tang, J. (2023). Toward degree bias in embedding-based knowledge graph completion. In *Proceedings of the ACM web conference 2023* (pp. 705–715). http://dx.doi.org/10.1145/3543507.3583544.

Shu, Z., Sun, X., & Cheng, H. (2024). When LLM meets hypergraph: A sociological analysis on personality via online social networks. In *Proceedings of the 33rd ACM International conference on information and knowledge management* (pp. 2087–2096).

Şimşek, Ö., & Jensen, D. (2008). Navigating networks by using homophily and degree. *Proceedings of the National Academy of Sciences, 105*(35), 12758–12762. https://doi.org/10.1073/PNAS.0800497105

Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Amin, M., Hou, L., Clark, K., Pfohl, S. R., Cole-Lewis, H. et al. (2025). Toward expert-level medical question answering

with large language models. *Nature Medicine*, (pp. 1–8). https://doi.org/{10.1038/s41591-024-03423-7}.

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J., & Wang, K. (2015). An overview of microsoft academic service (mas) and applications. In *Proceedings of the 24th International conference on world wide web* (pp. 243–246). https://doi.org/10.1145/2740908.2742839

Socher, R., Chen, D., Manning, C. D., & Ng, A. (2013). Reasoning with neural tensor networks for knowledge base completion. *Advances in Neural Information Processing Systems, 26*. https://proceedings.neurips.cc/paper_files/paper/2013/file/b337e84de8752b27eda3a12363109e80-Paper.pdf.

Song, J., Park, J., & Yang, E. (2022). TAM: Topology-aware margin loss for class-imbalanced node classification. In *International conference on machine learning* (pp. 20369–20383). PMLR.

Soomro, K., Zamir, A. R., & Shah, M. (2012). UCF101: A dataset of 101 human actions classes from videos in the wild. arXiv:1212.0402

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* https://doi.org/{10.1108/eb026526}.

Su, M., Li, Z., Chen, Z., Bai, L., Jin, X., & Guo, J. (2024). Temporal knowledge graph question answering: A survey. arXiv:2406.14191

Sun, Q., Yu, Q., Cui, Y., Zhang, F., Zhang, X., Wang, Y., Gao, H., Liu, J., Huang, T., & Wang, X. (2024). EMU: Generative pretraining in multimodality. In *International conference on learning representations*. https://openreview.net/forum?id=mL8Q9OOamV.

Sun, S., Ren, Y., Ma, C., & Zhang, X. (2023). Large language models as topological structure enhancers for text-attributed graphs. arXiv:2311.14324

Suzgun, M., Scales, N., Schärli, N., Gehrmann, S., Tay, Y., Chung, H. W., Chowdhery, A., Le, Q. V., Chi, E. H., Zhou, D. et al. (2023). Challenging big-bench tasks and whether chain-of-thought can solve them. In *Findings of the association for computational linguistics* (pp. 13003–13051). Association for Computational Linguistics. https://doi.org/10.18653/V1/2023.FINDINGS-ACL.824

Tan, Q., Ng, H. T., & Bing, L. (2023). Towards benchmarking and improving the temporal reasoning capability of large language models. arXiv:2306.08952

Tan, Y., Lv, H., Huang, X., Zhang, J., Wang, S., & Yang, C. (2024a). MuseGraph: Graph-oriented instruction tuning of large language models for generic graph mining. arXiv:2403.04780

Tan, Y., Zhou, Z., Lv, H., Liu, W., & Yang, C. (2024b). WalkLM: A uniform language model fine-tuning framework for attributed graph embedding. In *Advances in neural information processing systems*. https://openreview.net/forum?id=ZrG8kTbt70.

Tang, J., Li, J., Gao, Z., & Li, J. (2022). Rethinking graph neural networks for anomaly detection. In *International conference on machine learning* (pp. 21076–21089). PMLR. https://proceedings.mlr.press/v162/tang22b.html.

Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015). LINE: Large-scale information network embedding. In *Proceedings of the International conference on world wide web* (pp. 1067–1077). https://doi.org/10.1145/2736277.2741093

Tang, J., Yang, Y., Wei, W., Shi, L., Su, L., Cheng, S., Yin, D., & Huang, C. (2024a). GraphGPT: Graph instruction tuning for large language models. In *Proceedings of the 47th International ACM SIGIR conference on research and development in information retrieval*. https://doi.org/{10.1145/3626772.3657775}.

Tang, J., Yang, Y., Wei, W., Shi, L., Xia, L., Yin, D., & Huang, C. (2024b). HiGPT: Heterogeneous graph language model. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*. https://doi.org/{10.1145/3637528.3671987}.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). ArnetMiner: Extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD International conference on knowledge discovery and data mining, KDD '08* (p. 990-998). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1401890.1402008

Tang, L., & Liu, H. (2009). Relational learning via latent social dimensions. In *Proceedings of the 15th ACM SIGKDD International conference on knowledge discovery and data mining* (pp. 817–826). https://doi.org/10.1145/1557019.1557109

Thapaliya, B., Nguyen, A., Lu, Y., Xie, T., Grudetskyi, I., Lin, F., Valkanas, A., Liu, J., Chakraborty, D., & Fehri, B. (2024). ECGN: A cluster-aware approach to graph neural networks for imbalanced classification. arXiv:2410.11765

Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). FEVER: A large-scale dataset for fact extraction and verification. arXiv:1803.05355

Tian, Y., Song, H., Wang, Z., Wang, H., Hu, Z., Wang, F., Chawla, N. V., & Xu, P. (2024). Graph neural prompting with large language models. In *Proceedings of the AAAI conference on artificial intelligence*. https://doi.org/{10.1609/aaai.v38i17.29875}.

Tianchi (2018a). Ad display/click data on taobao.com. https://tianchi.aliyun.com/dataset/dataDetail?dataId=56.

Tianchi (2018b). IJCAI-16 brick-and-mortar store recommendation dataset. https://tianchi.aliyun.com/dataset/dataDetail?dataId=53.

Toutanova, K., & Chen, D. (2015). Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd workshop on continuous vector space models and their compositionality* (pp. 57–66). https://doi.org/10.18653/V1/W15-4007

Trivedi, R., Farajtabar, M., Biswal, P., & Zha, H. (c). DyRep: Learning representations over dynamic graphs. *International Conference on Learning Representation* https://par.nsf.gov/biblio/10190736.

Tsai, Y.-H. H., Yeh, Y.-R., & Wang, Y.-C. F. (2016). Learning cross-domain landmarks for heterogeneous domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. https://openaccess.thecvf.com/content_cvpr_2016/html/Tsai_Learning_Cross-Domain_Landmarks_CVPR_2016_paper.html.

Tsatsaronis, G., Balikas, G., Malakasiotis, P., Partalas, I., Zschunke, M., Alvers, M. R., Weissenborn, D., Krithara, A., Petridis, S., Polychronopoulos, D. et al. (2015). An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition. *BMC Bioinformatics* https://doi.org/{10.1186/s12859-015-0564-6}.

Tu, W., Zhou, S., Liu, X., Liu, Y., Cai, Z., Zhu, E., Changwang, Z., & Cheng, J. (2022). Initializing then refining: A simple graph attribute imputation network. In *Proceedings*

of the International joint conference on artificial intelligence (pp. 3494–3500). https://doi.org/10.24963/IJCAI.2022/485

Um, D., Park, J., Park, S., & Choi, J. Y. (2023). Confidence-based feature imputation for graphs with partially known features. In *The eleventh International conference on learning representations*. https://openreview.net/forum?id=YPKBIILy-Kt.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems, 30*. arXiv:1706.03762

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. In *International conference on learning representations*. https://openreview.net/forum?id=rJXMpikCZ.

Velickovic, P., Fedus, W., Hamilton, W. L., Liò, P., Bengio, Y., & Hjelm, R. D. (2019). Deep graph infomax. In *Proceedings of the International conference on learning representations* (p. 4). (vol. 2). https://openreview.net/forum?id=rklz9iAcKQ.

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: A free collaborative knowledgebase. *Communications of the ACM, 57*(10), 78–85. https://doi.org/10.1145/2629489

Wang, B., Jia, J., Cao, X., & Gong, N. Z. (2021a). Certified robustness of graph neural networks against adversarial structural perturbation. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining* (pp. 1645–1653). https://doi.org/10.1145/3447548.3467295

Wang, D., Zuo, Y., Li, F., & Wu, J. (2024a). LLMs as zero-shot graph learners: Alignment of GNN representations with LLM token embeddings. In *Advances in neural information processing systems*. https://openreview.net/forum?id=32g9BWTndc.

Wang, H., Feng, S., He, T., Tan, Z., Han, X., & Tsvetkov, Y. (2023a). Can language models solve graph problems in natural language? In *Advances in neural information processing systems*. https://openreview.net/forum?id=UDqHhbqYJV.

Wang, J., Hu, J., Qian, S., Fang, Q., & Xu, C. (2020). Multimodal graph convolutional networks for high quality content recognition. *Neurocomputing* https://doi.org/{10.1016/j.neucom.2020.04.145}.

Wang, J., Sun, K., Luo, L., Wei, W., Hu, Y., Liew, A. W.-C., Pan, S., & Yin, B. (2024b). Large language models-guided dynamic adaptation for temporal knowledge graph reasoning. arXiv:2405.14170

Wang, J., Wang, K., Zhang, Y., Zhang, W., Xu, X., & Lin, X. (2025). On LLM-enhanced mixed-type data imputation with high-order message passing. arXiv:2501.02191

Wang, K., Xu, Y., Wu, Z., & Luo, S. (2024c). LLM as prompter: Low-resource inductive reasoning on arbitrary knowledge graphs. In *Findings of the association for computational linguistics*. https://doi.org/10.18653/V1/2024.FINDINGS-ACL.224

Wang, L., Wang, Y., Ni, B., Zhao, Y., & Derr, T. (2024d). Large language model-based augmentation for imbalanced node classification on text-attributed graphs. arXiv:2410.16882

Wang, L., Zhou, W., Liu, L., Yang, Z., & Wen, J. (2023b). Deep adaptive collaborative graph neural network for social recommendation. *Expert Systems with Applications, 229*, 120410. https://doi.org/10.1016/J.ESWA.2023.120410

Wang, S., Huang, J., Chen, Z., Song, Y., Tang, W., Mao, H., Fan, W., Liu, H., Liu, X., Yin, D. et al. (2024e). Graph machine learning in the era of large language models (llms). *ACM Transactions on Intelligent Systems and Technology* https://doi.org/10.1145/3732786

Wang, X., Gao, T., Zhu, Z., Zhang, Z., Liu, Z., Li, J., & Tang, J. (2021b). KEPLER: A unified model for knowledge embedding and pre-trained language representation. *Transactions of the Association for Computational Linguistics, 9*, 176–194. https://doi.org/{10.1162/tacl_a_00360}.

Wang, X., Liu, N., Han, H., & Shi, C. (2021c). Self-supervised heterogeneous graph neural network with co-contrastive learning. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining* (pp. 1726–1736). https://doi.org/10.1145/3447548.3467415

Wang, X., Zhuang, B., & Wu, Q. (2024f). ModaVerse: Efficiently transforming modalities with LLMs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. https://openaccess.thecvf.com/content/CVPR2024/html/Wang_ModaVerse_Efficiently_Transforming_Modalities_with_LLMs_CVPR_2024_paper.html.

Wang, Y., Zhao, Y., Shah, N., & Derr, T. (2022). Imbalanced graph classification via graph-of-graph neural networks. In *Proceedings of the 31st ACM International conference on information & knowledge management* (pp. 2067–2076). https://doi.org/10.1145/3511808.3557356

Wei, W., Ren, X., Tang, J., Wang, Q., Su, L., Cheng, S., Wang, J., Yin, D., & Huang, C. (2024a). LLMRec: Large language models with graph augmentation for recommendation. In *Proceedings of the 17th ACM International conference on web search and data mining*. https://doi.org/10.1145/3616855.3635853}.

Wei, Y., Fu, S., Jiang, W., Zhang, Z., Zeng, Z., Wu, Q., Kwok, J., & Zhang, Y. (2024b). GITA: Graph to visual and textual integration for vision-language graph reasoning. In *Advances in neural information processing systems*. https://openreview.net/forum?id=SaodQ13jga.

Wei, Y., Huang, Q., Zhang, Y., & Kwok, J. (2023). KICGPT: Large language model with knowledge in context for knowledge graph completion. In *Findings of the association for computational linguistics: EMNLP 2023* (pp. 8667–8683). https://doi.org/10.18653/V1/2023.FINDINGS-EMNLP.580

Wei, Y., Wang, X., Nie, L., He, X., Hong, R., & Chua, T.-S. (2019). MMGCN: Multi-modal graph convolution network for personalized recommendation of micro-video. In *Proceedings of the 27th ACM International conference on multimedia*. https://doi.org/{10.1145/3343031.3351034}.

Wen, H., Ding, J., Jin, W., Wang, Y., Xie, Y., & Tang, J. (2022). Graph neural networks for multimodal single-cell data integration. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*. https://doi.org/{10.1145/3534678.3539213}.

Wen, Z., & Fang, Y. (2023). Augmenting low-resource text classification with graph-grounded pre-training and prompting. In *Proceedings of the 46th International ACM SIGIR conference on research and development in information retrieval* (pp. 506–516). https://doi.org/{10.1145/3539618.3591641}.

Wilcke, W. X., Bloem, P., de Boer, V., van t Veer, R. H., & van Harmelen, F. (2020). End-to-end entity classification on multimodal knowledge graphs. arXiv:2003.12383

Wu, F., Qiao, Y., Chen, J.-H., Wu, C., Qi, T., Lian, J., Liu, D., Xie, X., Gao, J., Wu, W. et al. (2020). MIND: A large-scale dataset for news recommendation. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 3597–3606). https://aclanthology.org/2020.acl-main.331/. https://doi.org/10.18653/v1/2020.acl-main.331

Wu, F., Souza, A., Zhang, T., Fifty, C., Yu, T., & Weinberger, K. (2019). Simplifying graph convolutional networks. In *Proceedings of the International conference on machine learning* (pp. 6861–6871). PMLR. http://proceedings.mlr.press/v97/wu19e.html.

Wu, H., Yin, J., Rajaratnam, B., & Guo, J. (2022). Hierarchical relational learning for few-shot knowledge graph completion. arXiv:2209.01205

Wu, M., Zheng, X., Zhang, Q., Shen, X., Luo, X., Zhu, X., & Pan, S. (2024a). Graph learning under distribution shifts: A comprehensive survey on domain adaptation, out-of-distribution, and continual learning. arXiv:2402.16374

Wu, Y., Wang, L., Han, X., & Ye, H.-J. (2024b). Graph contrastive learning with cohesive subgraph awareness. In *Proceedings of the ACM web conference* (pp. 629–640). https://doi.org/10.1145/3589334.3645470

Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., Leswing, K., & Pande, V. (2018). MoleculeNet: A benchmark for molecular machine learning. *Chemical Science* https://doi.org/{10.1039/C7SC02664A}.

Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S., & Liu, H. (2021). Graph learning: A survey. *IEEE Transactions on Artificial Intelligence*, 2(2), 109–127. https://doi.org/10.1109/TAI.2021.3076021

Xia, R., Zhang, C., Li, A., Liu, X., & Yang, B. (2024a). Attribute imputation autoencoders for attribute-missing graphs. *Knowledge-Based Systems*, 291, 111583. https://doi.org/10.1016/J.KNOSYS.2024.111583

Xia, Y., Wang, D., Liu, Q., Wang, L., Wu, S., & Zhang, X.-Y. (2024b). Chain-of-history reasoning for temporal knowledge graph forecasting. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), *Findings of the association for computational linguistics: ACL 2024* (pp. 16144–16159). Bangkok, Thailand: Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-acl.955

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE computer society conference on computer vision and pattern recognition*. https://doi.org/{10.1109/CVPR.2010.5539970}.

Xie, Y., Wang, X., Wang, P., & Bi, X. (2025). A pseudo-label supervised graph fusion attention network for drug–target interaction prediction. *Expert Systems with Applications*, 259, 125264. https://doi.org/10.1016/J.ESWA.2024.125264

Xiong, J., Zareie, A., & Sakellariou, R. (2025). A survey of link prediction in temporal networks. arXiv:2502.21185

Xiong, S., Payani, A., Kompella, R., & Fekri, F. (2024a). Large language models can learn temporal reasoning. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), *Proceedings of the 62nd meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 10452–10470). Bangkok, Thailand: Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.563

Xiong, S., Payani, A., Kompella, R., & Fekri, F. (2024b). Large language models can learn temporal reasoning. arXiv:2401.06853

Xiong, W., Hoang, T., & Wang, W. Y. (2017). DeepPath: A reinforcement learning method for knowledge graph reasoning. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 564–573). Association for Computational Linguistics. https://doi.org/10.18653/V1/D17-1060

Xiong, W., Yu, M., Chang, S., Guo, X., & Wang, W. Y. (2018). One-shot relational learning for knowledge graphs. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 1980–1990). Association for Computational Linguistics. https://doi.org/10.18653/V1/D18-1223

Xu, d., ruan, c., korpeoglu, e., kumar, s., & achan, k. (2020). Inductive representation learning on temporal graphs. In *International conference on learning representations (ICLR)*. https://openreview.net/forum?id=rJeW1yHYwH.

Xu, D., Zhang, Z., Lin, Z., Wu, X., Zhu, Z., Xu, T., Zhao, X., Zheng, Y., & Chen, E. (2024a). Multi-perspective improvement of knowledge graph completion with large language models. arXiv:2403.01972

Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How powerful are graph neural networks? In *Proceedings of the International conference on learning representations*. https://openreview.net/forum?id=ryGs6iA5Km.

Xu, W., Liu, B., Peng, M., Jia, X., & Peng, M. (2023). Pre-trained language model with prompts for temporal knowledge graph completion. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Findings of the association for computational linguistics: ACL 2023* (pp. 7790–7803). Toronto, Canada: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.493

Xu, Y., He, S., Chen, J., Wang, Z., Song, Y., Tong, H., Liu, G., Liu, K., & Zhao, J. (2024b). Generate-on-graph: Treat LLM as both agent and kg in incomplete knowledge graph question answering. arXiv:2404.14741

Xue, R., Shen, X., Yu, R., & Liu, X. (2023). Efficient large language models fine-tuning on graphs. arXiv:2312.04737

Yan, H., Li, C., Long, R., Yan, C., Zhao, J., Zhuang, W., Yin, J., Zhang, P., Han, W., Sun, H. et al. (2023). A comprehensive study on text-attributed graphs: Benchmarking and rethinking. *Advances in Neural Information Processing Systems*, 36, 17238–17264. https://proceedings.neurips.cc/paper_files/paper/2023/file/37d00f567a18b478065f1a91b95622a0-Paper-Datasets_and_Benchmarks.pdf.

Yan, H., Li, C., Yin, J., Yu, Z., Han, W., Li, M., Zeng, Z., Sun, H., & Wang, S. (2024). When graph meets multimodal: Benchmarking and mediating on multimodal attributed graphs learning. arXiv:2410.09132

Yang, C., Liu, Z., Zhao, D., Sun, M., & Chang, E. Y. (2015). Network representation learning with rich text information. In *International joint conference on artificial intelligence*. https://www.ijcai.org/Proceedings/15/Papers/299.pdf.

Yang, C., Yu, X., Wu, J., Zhang, B., & Yang, H. (2024a). Graph-aware multi-feature interacting network for explainable rumor detection on social network. *Expert Systems with Applications*, 249, 123687. https://doi.org/10.1016/J.ESWA.2024.123687

Yang, J., Liu, Z., Xiao, S., Li, C., Lian, D., Agrawal, S., Singh, A., Sun, G., & Xie, X. (2021). GraphFormers: GNN-nested transformers for representation learning on textual graph. In *Advances in neural information processing systems*. https://openreview.net/forum?id=yILzFBjR0Y.

Yang, J., Xu, H., Mirzoyan, S., Chen, T., Liu, Z., Liu, Z., Ju, W., Liu, L., Xiao, Z., Zhang, M. et al. (2024b). Poisoning medical knowledge using large language models. *Nature Machine Intelligence*, 6(10), 1156–1168. https://doi.org/10.1038/S42256-024-00899-3

Yang, R., Zhu, J., Man, J., Fang, L., & Zhou, Y. (2024c). Enhancing text-based knowledge graph completion with zero-shot large language models: A focus on semantic enhancement. *Knowledge-Based Systems*, 300, 112155. https://doi.org/10.1016/J.KNOSYS.2024.112155

Yang, R., Zhu, J., Man, J., Liu, H., Fang, L., & Zhou, Y. (2025). GS-KGC: A generative subgraph-based framework for knowledge graph completion with large language models. *Information Fusion*, 117, 102868. https://doi.org/10.1016/J.INFFUS.2024.102868

Yang, X., Liu, Y., Zhou, S., Wang, S., Tu, W., Zheng, Q., Liu, X., Fang, L., & Zhu, E. (2023). Cluster-guided contrastive graph clustering network. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 10834–10842). (vol. 37). arXiv:2301.01098

Yang, Z., Zhang, G., Wu, J., Yang, J., Sheng, Q. Z., Xue, S., Zhou, C., Aggarwal, C., Peng, H., Hu, W. et al. (2024d). State of the art and potentialities of graph-level learning. *ACM Computing Surveys*, 57(2), 1–40. https://doi.org/10.1145/3695863

Yao, H., Zhang, C., Wei, Y., Jiang, M., Wang, S., Huang, J., Chawla, N., & Li, Z. (2020). Graph few-shot learning via knowledge transfer. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 6656–6663). (vol. 34). https://doi.org/10.1609/AAAI.V34I04.6142

Yao, L., Mao, C., & Luo, Y. (2019). KG-BERT: Bert for knowledge graph completion. arXiv:1909.03193

Yao, L., Peng, J., Mao, C., & Luo, Y. (2023). Exploring large language models for knowledge graph completion. arXiv:2308.13916

Ye, R., Zhang, C., Wang, R., Xu, S., & Zhang, Y. (2024). Language is all a graph needs. In *Findings of the association for computational linguistics: EACL 2024*. https://aclanthology.org/2024.findings-eacl.132/.

Yi, S.-Y., Mao, Z., Ju, W., Zhou, Y.-D., Liu, L., Luo, X., & Zhang, M. (2023). Towards long-tailed recognition for graph classification via collaborative experts. *IEEE Transactions on Big Data*, 9(6), 1683–1696. https://doi.org/10.1109/TBDATA.2023.3313029

Yih, W.-t., Richardson, M., Meek, C., Chang, M.-W., & Suh, J. (2016). The value of semantic parse labeling for knowledge base question answering. In *Proceedings of the 54th annual meeting of the association for computational linguistics* (pp. 201–206). https://doi.org/10.18653/V1/P16-2033

Ying, Z., Bourgeois, D., You, J., Zitnik, M., & Leskovec, J. (2019). GNNExplainer: Generating explanations for graph neural networks. *Advances in Neural Information Processing Systems*, 32. https://proceedings.neurips.cc/paper/2019/hash/d80b7040b773199015de6d3b4293c8ff-Abstract.html.

Yoo, J., Jeon, H., Jung, J., & Kang, U. (2022). Accurate node feature estimation with structured variational graph autoencoder. In *Proceedings of the ACM SIGKDD conference on knowledge discovery and data mining* (pp. 2336–2346). https://doi.org/10.1145/3534678.3539337

Yoon, M., Koh, J. Y., Hooi, B., & Salakhutdinov, R. (2023). Multimodal graph learning for generative tasks. In *NeurIPS 2023 workshop: New frontiers in graph learning*. https://openreview.net/forum?id=YILik4gFBk.

You, J., Ying, R., Ren, X., Hamilton, W., & Leskovec, J. (2018). GraphRNN: Generating realistic graphs with deep auto-regressive models. In *International conference on machine learning* (pp. 5708–5717). PMLR. http://proceedings.mlr.press/v80/you18a.html.

You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., & Shen, Y. (2020). Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems*, 33, 5812–5823. https://proceedings.neurips.cc/paper/2020/hash/3fe230348e9a12c13120749e3f9fa4cd-Abstract.html.

You, Y., Chen, T., Wang, Z., & Shen, Y. (2023). Graph domain adaptation via theory-grounded spectral regularization. In *International conference on learning representations*. https://openreview.net/forum?id=OysfLgrk8mk.

Yu, J., Ren, Y., Gong, C., Tan, J., Li, X., & Zhang, X. (2025). Leveraging large language models for node generation in few-shot learning on text-attributed graphs. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 13087–13095). (vol. 39). https://doi.org/10.1609/aaai.v39i12.33428

Yuan, C., Xie, Q., Huang, J., & Ananiadou, S. (2024). Back to the future: Towards explainable temporal reasoning with large language models. In *Proceedings of the ACM web conference 2024, WWW '24* (p. 1963-1974). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3589334.3645376

Zeng, H., Zhou, H., Srivastava, A., Kannan, R., & Prasanna, V. (2020). GraphSAINT: Graph sampling based inductive learning method. In *International conference on learning representations*. https://openreview.net/forum?id=BJe8pkHFwS.

Zeng, L., Li, L., Gao, Z., Zhao, P., & Li, J. (2023). ImGCL: Revisiting graph contrastive learning on imbalanced node classification. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 11138–11146). (vol. 37).

Zhang, C., Song, D., Huang, C., Swami, A., & Chawla, N. V. (2019a). Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD International conference on knowledge discovery & data mining*. https://doi.org/{10.1145/3292500.3330961}.

Zhang, C., Yao, H., Huang, C., Jiang, M., Li, Z., & Chawla, N. V. (2020). Few-shot knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 3041–3048). (vol. 34). https://doi.org/10.1609/AAAI.V34I03.5698

Zhang, G., Yang, Z., Wu, J., Yang, J., Xue, S., Peng, H., Su, J., Zhou, C., Sheng, Q. Z., Akoglu, L. et al. (2022). Dual-discriminative graph neural network for imbalanced graph-level anomaly detection. *Advances in Neural Information Processing Systems*, 35, 24144–24157. http://papers.nips.cc/paper_files/paper/2022/hash/98a625423070cfc6ae3d82d4b59408a0-Abstract-Conference.html.

Zhang, H., Zheng, T., Gao, J., Miao, C., Su, L., Li, Y., & Ren, K. (2019b). Data poisoning attack against knowledge graph embedding. In *Proceedings of the International joint conference on artificial intelligence* (pp. 4853–4859). https://doi.org/10.24963/IJCAI.2019/674

Zhang, J., Mao, S., Yang, L., Ma, W., Li, S., & Gao, Z. (2024a). Physics-informed deep learning for traffic state estimation based on the traffic flow model and computational graph method. *Information Fusion*, *101*, 101971. https://doi.org/10.1016/J.INFFUS.2023.101971

Zhang, J., Zhang, S., Zhao, H., Yang, Y., & Liang, M. (2025a). Multi-frequency spatial-temporal graph neural network for short-term metro OD demand prediction during public health emergencies. *Transportation*, (pp. 1–23). https://doi.org/{10.1007/s11116-025-10582-0}.

Zhang, P., Cao, C., & Groth, P. (2024b). TIGER: Temporally improved graph entity linker. In *Proceedings of the 27th European conference on artificial intelligence* (pp. 3733–3740). IOS Press (*vol. 392*). https://doi.org/10.3233/FAIA240933

Zhang, P., Cao, C., Zaporojets, K., & Groth, P. (2024c). CYCLE: Cross-year contrastive learning in entity-linking. In *Proceedings of the 33rd ACM International conference on information and knowledge management* (p. 3197-3206). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3627673.3679702

Zhang, R., Xing, W., Li, M., Wang, Z., Chen, J., Ma, X., Liu, Z., & He, Z. (2025b). MM-STFlowNet: A transportation hub-oriented multi-mode passenger flow prediction method via spatial-temporal dynamic graph modeling arXiv:2504.06325

Zhang, T., Yang, R., Yan, M., Ye, X., Fan, D., & Lai, Y. (2024d). Cost-effective label-free node classification with LLMs. arXiv:2412.11983

Zhang, X., Huang, C., Xu, Y., Xia, L., Dai, P., Bo, L., Zhang, J., & Zheng, Y. (2021). Traffic flow forecasting with spatial-temporal graph diffusion network. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 15008–15015). https://doi.org/10.1609/AAAI.V35I17.17761

Zhang, X., Zhang, L., Zhou, D., & Xu, G. (2024e). Fine-grainedly synthesize streaming data based on large language models with graph structure understanding for data sparsity. arXiv:2403.06139

Zhang, X., & Zitnik, M. (2020). GNNGuard: Defending graph neural networks against adversarial attacks. *Advances in Neural Information Processing Systems*, *33*, 9263–9275. https://proceedings.neurips.cc/paper/2020/hash/690d83983a63aa1818423fd6edd3bfdb-Abstract.html.

Zhang, Y., Chen, Z., Guo, L., Xu, Y., Zhang, W., & Chen, H. (2024f). Making large language models perform better in knowledge graph completion. In *Proceedings of the 32nd ACM International conference on multimedia* (pp. 233–242). https://doi.org/10.1145/3664647.3681327

Zhang, Y., Dai, H., Kozareva, Z., Smola, A., & Song, L. (2018). Variational reasoning for question answering with knowledge graph. In *Proceedings of the AAAI conference on artificial intelligence*. https://doi.org/{10.1609/aaai.v32i1.12057}.

Zhang, Y., Song, G., Du, L., Yang, S., & Jin, Y. (2019c). DANE: Domain adaptive network embedding. In *International joint conference on artificial intelligence*. https://www.ijcai.org/Proceedings/2019/0606.pdf.

Zhang, Z., Chen, J., & Yang, D. (2024g). DARG: Dynamic evaluation of large language models via adaptive reasoning graph. arXiv:2406.17271

Zhang, Z., Wang, X., Zhang, Z., Li, H., Qin, Y., & Zhu, W. (2024h). LLM4DyG: Can large language models solve spatial-temporal problems on dynamic graphs? In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining, KDD '24* (p. 4350-4361). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3637528.3671709

Zhang, Z., Wang, X., Zhou, H., Yu, Y., Zhang, M., Yang, C., & Shi, C. (2024i). Can large language models improve the adversarial robustness of graph neural networks? arXiv:2408.08685

Zhao, J., Zhuo, L., Shen, Y., Qu, M., Liu, K., Bronstein, M., Zhu, Z., & Tang, J. (2023). GraphText: Graph reasoning in text space. arXiv:2310.01089

Zhao, R., Tang, J., Zeng, W., Chen, Z., & Zhao, X. (2024a). Zero-shot knowledge graph question generation via multi-agent LLMs and small models synthesis. In *Proceedings of the ACM International conference on information and knowledge management* (pp. 3341–3351). https://doi.org/10.1145/3627673.3679805

Zhao, T., Zhang, X., & Wang, S. (2021). GraphSMOTE: Imbalanced node classification on graphs with graph neural networks. In *Proceedings of the 14th ACM International conference on web search and data mining* (pp. 833–841). https://doi.org/10.1145/3437963.3441720

Zhao, Z., Lin, F., Zhu, X., Zheng, Z., Xu, T., Shen, S., Li, X., Yin, Z., & Chen, E. (2024b). DynLLM: When large language models meet dynamic graph recommendation. arXiv:2405.07580

Zheng, M., Hao, Y., Jiang, W., Lin, Z., Lyu, Y., She, Q., & Wang, W. (2023a). IM-TQA: A Chinese table question answering dataset with implicit and multi-type table structures. In *Proceedings of the 61st annual meeting of the association for computational linguistics* (pp. 5074–5094). https://doi.org/10.18653/V1/2023.ACL-LONG.278

Zheng, W., Huang, E. W., Rao, N., Katariya, S., Wang, Z., & Subbian, K. (2021). Cold brew: Distilling graph node representations with incomplete or missing neighborhoods. In *The tenth International conference on learning representations*. https://openreview.net/forum?id=1ugNpm7W6E.

Zheng, X., Zhang, M., Chen, C., Molaei, S., Zhou, C., & Pan, S. (2023b). GNNEvaluator: Evaluating GNN performance on unseen graphs without labels. In *Advances in neural information processing systems*.

Zhou, M., Huang, M., & Zhu, X. (2018). An interpretable reasoning network for multi-relation question answering. In *Proceedings of the 27th International conference on computational linguistics* (pp. 2010–2022).

Zhu, J., Song, X., Ioannidis, V., Koutra, D., & Faloutsos, C. (2024a). TouchUp-G: Improving feature representation through graph-centric finetuning. In *Proceedings of the 47th International ACM SIGIR conference on research and development in information retrieval*. https://doi.org/{10.1145/3626772.3657978}.

Zhu, J., Zhou, Y., Qian, S., He, Z., Zhao, T., Shah, N., & Koutra, D. (2024b). Multimodal graph benchmark. arXiv:2406.16321v1

Zhu, Q., Ponomareva, N., Han, J., & Perozzi, B. (2021). Shift-robust GNNs: Overcoming the limitations of localized graph training data. In *Advances in neural information processing systems*. https://openreview.net/forum?id=uY-XMIbyXec.

Zhu, Y., Shi, H., Wang, X., Liu, Y., Wang, Y., Peng, B., Hong, C., & Tang, S. (2024c). GraphCLIP: Enhancing transferability in graph foundation models for text-attributed graphs. arXiv:2410.10329

Zhuang, Z., Wang, J., Qi, Q., Sun, H., & Liao, J. (2019). Toward greater intelligence in route planning: A graph-aware deep learning approach. *IEEE Systems Journal* https://doi.org/10.1109/JSYST.2019.2922217

Zügner, D., Akbarnejad, A., & Günnemann, S. (2018). Adversarial attacks on neural networks for graph data. In *Proceedings of the 24th ACM SIGKDD International conference on knowledge discovery & data mining* (pp. 2847–2856). https://doi.org/10.1145/3219819.3220078