A Survey of Imbalanced Learning on Graphs: Problems, Techniques, and Future Directions

Zemin Liu, Yuan Li, Nan Chen, Qian Wang, Bryan Hooi, Bingsheng He

Abstract—Graphs represent interconnected structures prevalent in a myriad of real-world scenarios. Effective graph analytics, such as graph learning methods, enables users to gain profound insights from graph data, underpinning various tasks including node classification and link prediction. However, these methods often suffer from data imbalance, a common issue in graph data where certain segments possess abundant data while others are scarce, thereby leading to biased learning outcomes. This necessitates the emerging field of imbalanced learning on graphs, which aims to correct these data distribution skews for more accurate and representative learning outcomes. In this survey, we embark on a comprehensive review of the literature on imbalanced learning on graphs. We begin by providing a definitive understanding of the concept and related terminologies, establishing a strong foundational understanding for readers. Following this, we propose two comprehensive taxonomies: (1) the *problem taxonomy*, which describes the forms of imbalance we consider, the associated tasks, and potential solutions; (2) the *technique taxonomy*, which details key strategies for addressing these imbalances, and aids readers in their method selection process. Finally, we suggest prospective future directions for both problems and techniques within the sphere of imbalanced learning on graphs, fostering further innovation in this critical area.

Index Terms—Imbalanced learning on graphs, graph representation learning, class imbalance, structure imbalance.

✦

1 INTRODUCTION

GRAPHS, or networks, refer to interconnected structures that are commonly found in real-world scenarios, where entities often interact with each other. Graphs are ubiquitous in various domains, such as social networks on platforms like Facebook, citation networks on DBLP, e-commerce networks on Amazon, *etc.* The prevalence of graph structures has sparked significant interest in graph analytics, which aims to leverage the inherent information in graphs for downstream tasks such as node classification, link prediction, and graph classification.

Early studies in graph analytics often rely on conventional techniques like feature engineering [1]–[3], which can be computationally expensive and require significant effort. However, with the advent of graph representation learning [4], [5], new opportunities have emerged for graph analytics. Graph representation learning aims to embed the structures of graphs (*e.g.*, nodes, edges, or graphs) into a low-dimensional space while preserving their structural information. Prior graph embedding approaches [4] such as Deepwalk [6], LINE [7], and node2vec [8] rely on contextual connectivity between nodes to capture their proximity for node representation learning. Recently, more attention has shifted to graph neural networks (GNNs) [9]–[12], a family of graph representation learning approaches that capitalize on neighborhood aggregation. They conduct neighborhood aggregation to recursively pass and receive messages along the edges in an end-to-end manner, to effectively encode the graph structure. Consequently, GNNs have achieved state-of-the-art performance in many downstream tasks.

Imbalance phenomenon. While these graph representation learning approaches can be effective, like many machine learning models, they often require a large amount of labeled data for training. Real-world data, however, frequently display imbalanced distributions, where some segments have an abundance of data and others do not. For instance, in a classification task, the distribution of labeled data—such as images or documents—may skew towards certain classes, resulting in imbalanced label distribution. This data imbalance can notably impact the training process. Specifically, the model tends to be well-trained on *high-resource* groups that have ample data, while underperforming on *low-resource* groups that are data-limited, leading to unclear class boundaries [13]–[15]. As a result, the model performance is usually satisfactory within high-resource groups, but degrades within low-resource groups [16]–[18]. Thus, severe challenges of imbalanced learning have garnered significant attention in the literature and addressing those challenges is of paramount importance [17], [19]–[22].

Imbalanced learning on graphs. To tackle the imbalance issues, various approaches have been proposed in the fields of vision and language [23], [24]. However, graph data differs from them in the fact that samples are inherently non-*i.i.d.* and multifarious, with diverse structural aspects (*e.g.*, node degrees). Directly applying these approaches to address imbalance issues on graphs may not be feasible [25]–[27]. Thus, graph learning faces the challenge of new imbalance types beyond those found in these settings.

The considerable impact of imbalance issues on the performance of graph-based tasks has recently attracted significant research interest [28]–[30], as illustrated in Fig. 1. The annually increasing volume of literature mirrors the escalating importance and substantial impact of address-
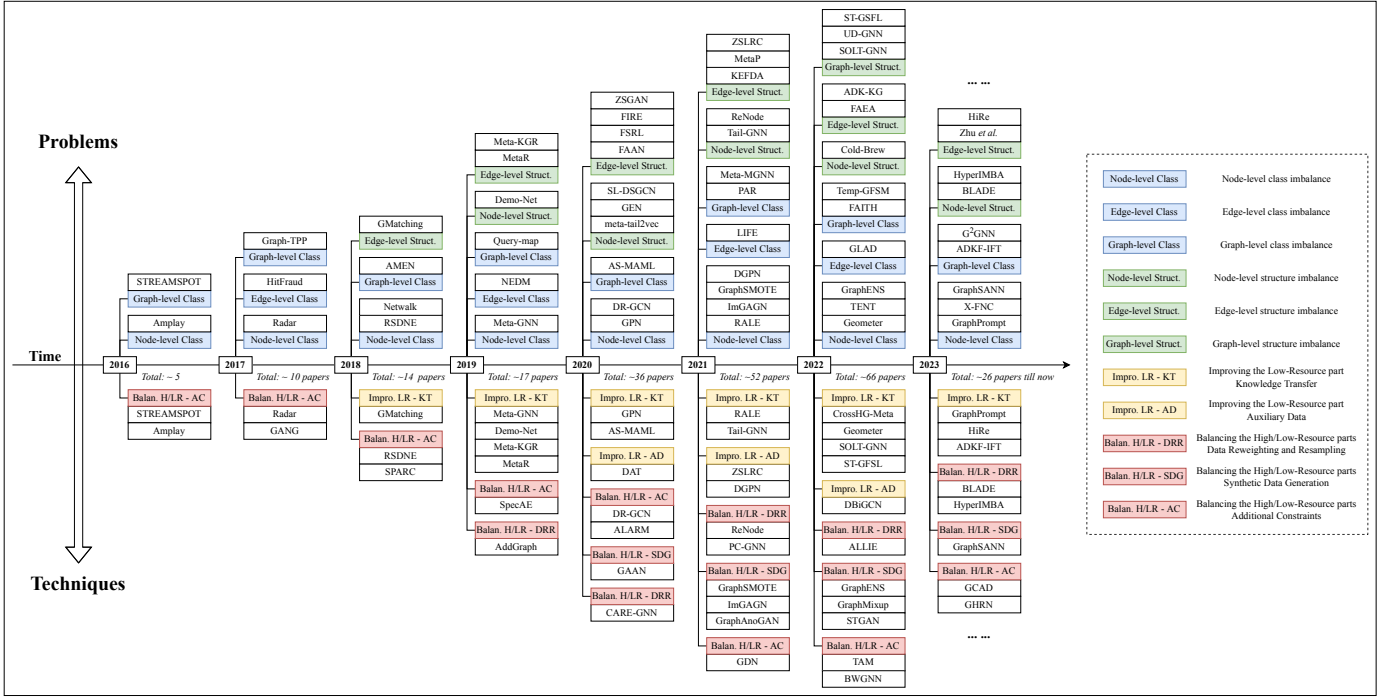
Fig. 1: This timeline showcases the representative research literature on ILoGs in relation to our taxonomies, *i.e.*, Problems and Techniques. The total number of research works for each year is also displayed, illustrating an upward trend.

ing challenges in imbalanced learning within the realm of graph learning tasks. These research efforts endeavor to address a variety of real-world applications [18], [26], [31]–[34], formalized into several distinct research *problems*. Each problem, characterized by its unique features, calls for the development of specialized *techniques* tailored to effectively address the imbalance issues specific to each scenario. However, the diversity of problems and techniques has resulted in a scattered landscape of imbalanced learning on graphs, lacking a comprehensive framework to identify their commonalities and disparities.

Therefore, in this survey, we focus on **I**mbalanced **L**earning **o**n **G**raph**s** (**ILoGs**) to bridge this gap by reviewing and summarizing the *problems* and *techniques* in the context of addressing imbalance issues on graphs. The essence of ILoGs lies in the observation that graph learning models with imbalanced input typically exhibit varying performance across groups with different levels of graph resource abundance [18], [32], [34]–[36]. More precisely, as shown in Fig. 2, given the input graph(s), graph data is often divided into multiple segments, creating an imbalanced distribution of graph resources. This phenomenon can be observed in various tasks, such as imbalanced node classification [25], [37] and node representation learning with different degrees [18], [27], as depicted in Fig. 2(b). This uneven distribution often leads to imbalanced outcomes: graph models typically perform well on high-resource parts while marginalizing the low-resource parts, ultimately resulting in disparities in performance across parts, as shown in Fig. 2(c).

However, the multifaceted nature of imbalanced learning on graphs presents considerable complexity in its investigation. On one hand, graph structures spawn a wide array of *problems* characterized by varying forms of imbalance, tasks, and solutions. This diversity presents a challenge in unifying these elements. Therefore, creating an organized taxonomy to categorize these imbalanced learning problems on graphs is a significant task. Additionally, a detailed categorization will benefit future research by identifying unexplored areas. On the other hand, the numerous imbalance problems on graphs yield diverse solutions, and some tasks demand specific *techniques* due to their unique attributes. This creates a complex landscape of solutions, making research in this area challenging. Therefore, classifying literature from a technical perspective is crucial. Moreover, this classification can also aid readers in selecting appropriate techniques to handle their specific graph imbalance issues.

To resolve the complexity, we categorize the literature from the perspectives of problems and techniques to provide a comprehensive overview. Primarily, for the taxonomy of *problems*, we classify the literature based on class imbalance and structure imbalance, both stemming from imbalanced input. We further distill this into more specific categories: node-, edge-, and graph-level imbalance, offering a comprehensive understanding of graph imbalance, as depicted in Fig. 3. Besides, to provide a clearer insight into these imbalance issues, we delve deeper by detailing and contrasting their types, settings, and information abundance in Table 1. Additionally, for the taxonomy of *techniques*, we classify literature based on imbalance types and the associated strategies to mitigate it, as summarized in Fig. 4. This is because, specific types of imbalance may require particular techniques to effectively address the associated issues, and the techniques employed may vary across different imbalance types.

Using these taxonomies allows us to capture commonalities and disparities across imbalanced graph learning liter-
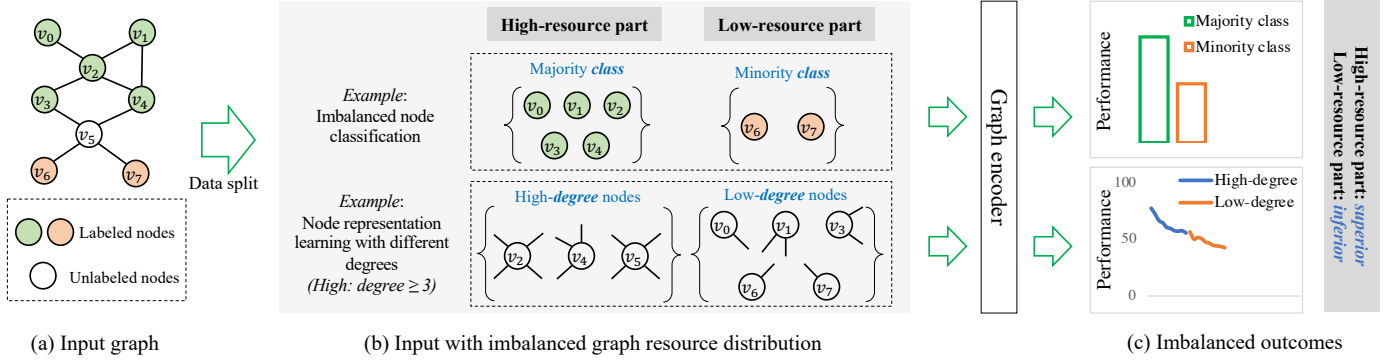
Fig. 2: Imbalanced learning on graphs: imbalanced graph resource distribution results in imbalanced outcomes.

ature. These hierarchical taxonomies group related studies based on shared features, such as imbalance type, imbalance level (*i.e.*, node, edge, and graph), and techniques. Simultaneously, the individual characteristics of each research task or technique are highlighted, underscoring the unique aspects that set them apart from each other. This approach provides a holistic view of the field, shedding light on both shared trends and unique advances. A more in-depth exploration of these taxonomies will be provided in Section 3.

Building on the taxonomies of problems and techniques, we further delineate promising directions for future research in the realm of imbalanced learning on graphs. Specifically, when considering future directions concerning *problems*, we dissect potential research challenges through the lenses of class imbalance and structure imbalance. As for future directions concerning *techniques*, we contemplate innovative solutions that could fuel the growth of this field.

Note that, given the sheer volume of research papers tackling this intricate issue, our survey does not delve into the nuances of each study. Instead, our primary objective is to present a holistic, structured overview in line with our established taxonomies. This approach ensures we offer a broad perspective on the domain and spotlight the emergent frontiers in imbalanced learning on graphs.

Relationship with existing surveys. Imbalanced learning has been the subject of several surveys in the past decade, covering various areas from general imbalanced classification [19], [20], [38] to more specific tasks like anomaly detection [39]–[41], few-shot learning [42], [43], long-tailed distribution [17], *etc.* However, these surveys have primarily focused on imbalanced learning in a general context or in specific tasks, and lack comprehensive coverage of imbalanced learning on graphs. In the domain of graph-related tasks, there are surveys that focus on specific tasks on graphs such as class-imbalanced learning [44], anomaly detection [28], [45]–[47], few-shot classification [29], [30], and fairness learning [48]. Despite their relevance, these surveys concentrate on individual tasks and lack a comprehensive overview of imbalanced learning on graphs.

Our survey fills this gap by providing a holistic view of imbalanced learning on graphs, covering diverse tasks with a focus on both class imbalance and structure imbalance. In contrast to previous surveys [28]–[30], [45]–[48], our work elucidates the shared traits and unique characteristics of these tasks, offering fresh insights into their commonalities

and differences within the sphere of imbalanced learning on graphs, based on the two taxonomies. Furthermore, we also incorporate fairness learning on graphs (see Section 7.1 for more details) as a specific type of imbalanced learning. Unlike other forms of imbalanced learning, fairness learning focuses on mitigating the potential bias and discrimination that may exist in the predictions made by the models, rather than solely on improving performance on imbalanced data.

Contributions. To summarize, our survey makes the following major contributions.

- We present the first comprehensive survey that broadly encompasses the field of imbalanced learning on graphs. Given the importance of this research area and the increasing number of papers exploring it, our survey serves as an invaluable resource for both researchers and practitioners.
- To provide a comprehensive and structured overview of the field, we introduce two novel taxonomies: problems and techniques. These taxonomies are designed to facilitate a thorough understanding of the existing literature, and provide a clear picture of the commonalities and distinctions through structured categorizations.
- We identify potential future research directions in the field of imbalanced learning on graphs, providing insights and guidance for those interested in advancing the state-of-the-art in this fast-paced field.

2 BACKGROUND

We present the main terminologies and definitions in this survey, for graphs and imbalanced learning, respectively.

2.1 Graphs and Graph Representation Learning

2.1.1 Graph Formalization

Graph. A **graph** can be represented as $G = \{\mathcal{V}, \mathcal{E}, \mathbf{X}_v, \mathbf{X}_e, \phi, \varphi, \mathcal{T}, \mathcal{R}\}$, where $\mathcal{V}$ is the set of nodes, $\mathcal{E}$ is the set of edges, $\mathbf{X}_v \in \mathbb{R}^{|\mathcal{V}| \times d_{\mathbf{X}_v}}$ and $\mathbf{X}_e \in \mathbb{R}^{|\mathcal{V}| \times d_{\mathbf{X}_e}}$ are the feature matrices of nodes and edges, respectively, and $\mathcal{T}$ and $\mathcal{R}$ are the sets of node types and edge types. For simplicity, we utilize $\mathbf{x}_v \in \mathbb{R}^{d_{\mathbf{X}_v}}$ and $\mathbf{x}_e \in \mathbb{R}^{d_{\mathbf{X}_e}}$ to denote the feature vectors of node v and edge e, respectively. The function $\phi : \mathcal{V} \to \mathcal{T}$ maps a node $v \in \mathcal{V}$ to its corresponding node type $\phi(v)$, while the function $\varphi : \mathcal{E} \to \mathcal{R}$ maps an edge $e_{\langle u,v \rangle} \in \mathcal{E}$ to its corresponding edge type $\varphi(e_{\langle u,v \rangle})$.

The variety of node and edge types leads to the differentiation of graphs into various forms. A **homogeneous graph** is a graph without considering the node and edge types, *i.e.*, $|\mathcal{T}| = |\mathcal{R}| = 1$, and has been the subject of most research in graph analytics due to its simplicity. A **heterogeneous graph** [49], also called a Heterogeneous Information Network (HIN), is a graph that has different node or edge types, *i.e.*, $|\mathcal{T}| + |\mathcal{R}| > 2$, and preserves diverse semantics due to the presence of node and edge types. In particular, heterogeneous graphs can be further classified into several categories, based on the diversity of heterogeneity. A **bipartite graph** requires that there only exist two types of nodes, with edges connecting nodes with different types, *i.e.*, $|\mathcal{T}| = 2$ and $|\mathcal{R}| = 1$, and $\phi(u) \neq \phi(v)$ $\forall e_{\langle u,v \rangle} \in \mathcal{E}$. Bipartite graphs are extensively investigated in the field of recommender systems [50] by formalizing the user-item interactions into graph format. Additionally, a **knowledge graph** (KG for short) [51] is a graph that consists of various types of edges between nodes, which are also known as *relations* and *entities*, respectively. KGs are prevalently employed in natural language processing to represent extracted knowledge from natural language.

For further information on these graph types, we refer readers to the relevant surveys mentioned above. As the main focus of this survey is on imbalanced learning on graphs, we do not extensively discuss or investigate the specific issues in these fields.

2.1.2 Graph Representation Learning

Graph representation learning [4], [5] is a powerful and efficient way for graph analytics. It offers a cost-effective alternative to traditional graph engineering techniques by mapping graph structures (*e.g.*, nodes) into a low-dimensional space while preserving the structural information.

Formally, let $\mathbf{h}_v \in \mathbb{R}^d$ denote the representation of node v. Node representation learning can be formalized as

$$\mathbf{h}_v = f(v, G; \theta_f), \qquad (1)$$

where θ_f represents the learnable parameters of the graph representation learning function $f(\cdot; \theta_f)$. Researchers can devise various objectives for the function $f(\cdot; \theta_f)$.

Recently, more attention has been shifted to GNNs [5] which are based on the key operation of neighborhood aggregation. GNNs aggregate messages from neighboring nodes recursively to generate node representations, enabling end-to-end learning. Formally, in the l-th GNN layer, the representation of node v can be calculated as

$$\mathbf{h}_v^l = \text{AGGR}(\mathbf{h}_v^{l-1}, \{\mathbf{h}_u^{l-1} : u \in \mathcal{N}_v\} \mid \theta_g^l), \qquad (2)$$

where $\mathcal{N}_v$ is the neighbors set of node v, $\text{AGGR}(\cdot)$ is the aggregation function, and θ_g^l is the parameters set in layer l. The aggregation function can take various forms, such as mean [9], sum [12], or attentive aggregation [11]. In the first layer, node representations can be initialized with node feature vectors, *i.e.*, $\mathbf{h}_v^0 = \mathbf{x}_v$. For simplicity, we directly use $\mathbf{h}_v$ to denote the output representation of node v.

Furthermore, graph-level representations [52] are often needed for graph-level tasks, such as molecular classification [53]. They typically leverage learned substructure representations (*e.g.*, nodes [12] or edges [54]) for further

aggregation to obtain graph representation. Given a graph G and considering node representations as the foundation, graph representation learning can be formalized as

$$\mathbf{h}_G = \text{READOUT}(\{\mathbf{h}_v : v \in \mathcal{V}\}; \theta_r), \qquad (3)$$

where $\text{READOUT}(\cdot)$ is the readout function that aggregates node representations into the graph representation, parameterized by θ_r. In general, the READOUT function can primarily be implemented in two ways, *i.e.*, global-pooling [12], [55] and hierarchical-pooling [52], [56].

2.2 Imbalanced Learning on Graphs

Imbalanced learning. The conventional approaches for imbalanced learning primarily focus on developing learning algorithms to handle imbalanced classes [19], [20], [38].

Definition 1 (Conventional Imbalanced Learning). *In the context of conventional imbalanced learning, consider a labeled data set $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$, which can be partitioned into K classes (groups) such that $\mathcal{D} = \bigcup_{1 \leq j \leq K} \mathcal{G}_j$ (given each group $\mathcal{G}_j = \{(x_i, y_i) : y_i = j\}$). There exists a notable imbalance in the number of labeled samples across these groups. Under this setting, the imbalanced distribution of samples across groups would give rise to biases in the performance of a learning algorithm. In particular, the low-resource groups, i.e., the classes with less labeled data, are usually marginalized by the learning model due to the domination of the high-resource groups, resulting in performance degradation for the former. The goal of imbalanced learning is to develop a balanced model that can improve the performance of low-resource groups, potentially reaching levels comparable to those of high-resource groups.*

The degree of imbalance in a given dataset can be evaluated using the *imbalance ratio*, a metric particularly pertinent in classification tasks [17], [25]. Formally, for a labeled dataset encompassing K classes (groups) $\mathcal{D} = \bigcup_{1 \leq i \leq K} \mathcal{G}_i$, and with classes ordered by their cardinality in a descending manner (*i.e.*, if $i < j$, then $|\mathcal{G}_i| \geq |\mathcal{G}_j|$, where $|\mathcal{G}_i|$ denotes the size of group $\mathcal{G}_i$), the imbalance ratio is defined as $|\mathcal{G}_1|/|\mathcal{G}_K|$. Essentially, the imbalance ratio quantifies the skewness of the distribution across classes in terms of their labeled data—a larger ratio signifies a more skewed distribution. However, due to the multifaceted nature of imbalance issues in graphs, the specific form of the imbalance ratio can vary. Next, we will introduce the concept of imbalanced learning on graphs as well as the imbalance ratio for graphs.

Imbalanced learning on graphs. Imbalanced learning on graphs is complex due to its distinct graph structures. In addition to class imbalance, the graph structure itself is a significant source of imbalance, which can cause varying performance across groups w.r.t. their structures.

Definition 2 (Imbalanced Learning on Graphs). *In addition to the number of labeled instances, imbalance in graph data can stem from disparities in structural abundance across groups, leading to a more complex imbalance pattern. For a given graph dataset comprising a set of elements (i.e., nodes, edges, or (sub)graphs) represented as $\mathcal{G} = \{x_i\}_{i=1}^N$, these elements can be further grouped into K subsets, i.e., $\mathcal{G} = \bigcup_{1 \leq i \leq K} \mathcal{G}_i$, according to specific criteria based on classes or structures, where $1 < K \leq N$. It is important to note that these groups differ*

Imbalance Types	Imbalance Tasks	Settings	Information Abundance s	Explanations
Node-Level Class Imbalance	Imbalanced node classification / Node-level anomaly detection	A set of (or two) node classes $\mathcal{D} = \bigcup_{1 \le i \le K} \mathcal{C}_i$	$|\mathcal{C}_i|$ (# labeled nodes in each class $\mathcal{C}_i$)	Labeled nodes are unevenly distributed across classes.
	Few-shot node classification / Zero-shot node classification	A set of base node classes $\mathcal{D}_b = \bigcup_{1 \le i \le K_1} \mathcal{C}_i$, and novel node classes $\mathcal{D}_n = \bigcup_{K_1 < i \le K_2} \mathcal{C}_i$	$|\mathcal{C}_i|$ (# labeled nodes in each class $\mathcal{C}_i$)	Base classes have abundant labeled nodes, while novel classes have few/no labeled nodes.
Edge-Level Class Imbalance	Few-shot link prediction	A set of base graphs $\mathcal{D}_b = \{G_i\}_{i=1}^{K_1}$ and novel graphs $\mathcal{D}_n = \{G_i\}_{i=K_1+1}^{K_2}$, where $G_i = \{\mathcal{V}_i, \mathcal{E}_i\}$	$|\mathcal{E}_i|$ (# edges in each graph G_i)	Base graphs have abundant edges, while novel graphs have limited edges.
	Edge-level anomaly detection	Two edge classes $\mathcal{D} = \mathcal{C}_1 \cup \mathcal{C}_2$	$|\mathcal{C}_i|$ (# labeled edges in each class $\mathcal{C}_i$)	Labeled edges are unevenly distributed across classes.
Graph-Level Class Imbalance	Imbalanced graph classification / Graph-level anomaly detection	A set of (or two) graph classes $\mathcal{D} = \bigcup_{1 \le i \le K} \mathcal{C}_i$	$|\mathcal{C}_i|$ (# labeled graphs in each class $\mathcal{C}_i$)	Labeled graphs are unevenly distributed across classes.
	Few-shot graph classification	A set of base graph classes $\mathcal{D}_b = \bigcup_{1 \le i \le K_1} \mathcal{C}_i$, and novel node classes $\mathcal{D}_n = \bigcup_{K_1 < i \le K_2} \mathcal{C}_i$	$|\mathcal{C}_i|$ (# labeled graphs in each class $\mathcal{C}_i$)	Base classes have abundant labeled graphs, while novel classes have few labeled graphs.
Node-Level Structure Imbalance	Imbalanced node degrees	A set of node groups $\mathcal{D} = \bigcup_{1 \le i \le K} \mathcal{G}_i$, where $\mathcal{G}_i = \{v_j : d_j = i\}$ (d_j is the degree of node v_j)	d_j (the degree of each node v_j)	Head nodes have high degrees, while tail/cold-start nodes have few/no degrees.
	Node topology imbalance	A set of node classes $\mathcal{D} = \bigcup_{1 \le i \le K} \mathcal{C}_i$	The consistency between true class boundaries and influence boundaries of labeled nodes	Classes with more consistent boundaries tend to propagate label information more effectively.
	Long-tail entity embedding	A set of entity groups $\mathcal{D} = \bigcup_{1 \le i \le K} \mathcal{G}_i$, where $\mathcal{G}_i = \{e_j : d_j = i\}$ (d_j is the # triplets of entity e_j)	d_j (# triplets of each entity e_j)	Head entities have more triplets, while tail/cold-start entities have few/no triplets.
Edge-Level Structure Imbalance	Few-shot relation classification / Zero-shot relation classification / Few-shot reasoning on KGs	A set of base relations $\mathcal{D}_b = \bigcup_{1 \le i \le K_1} \mathcal{R}_i$, and novel relations $\mathcal{D}_n = \bigcup_{K_1 < i \le K_2} \mathcal{R}_i$	$|\mathcal{R}_i|$ (# labeled triplets of each relation $\mathcal{R}_i$)	Base relations have abundant labeled triplets, while novel relations have few/no labeled triplets.
Graph-Level Structure Imbalance	Imbalanced graph sizes	A set of graph groups $\mathcal{D} = \bigcup_{1 \le i \le K} \mathcal{G}_i$, where $\mathcal{G}_i = \{G_j : |\mathcal{V}_j| = i\}$ ($|\mathcal{V}_j|$ is the size of graph G_j)	$|\mathcal{V}_j|$ (the size of each graph G_j)	Head graph have large sizes, while tail graphs have small sizes.
	Imbalanced topology groups	A set of topology motifs $\mathcal{D} = \bigcup_{1 \le i \le K} \mathcal{M}_i$	$|\mathcal{M}_i|$ (# instances of each motif $\mathcal{M}_i$ in one class)	Motifs with more instances have stronger associations with the class than the less frequent motifs.

TABLE 1: Category of existing graph imbalance issues.

in terms of information abundance, which results in unequal performance among them when used as input for a learning model.

Deviating from the initially defined imbalance ratio for class imbalance tasks, the multifaceted nature of imbalance issues inherent in graphs calls for a broader definition of the imbalance ratio. Formally, given K subsets of a set of elements $\mathcal{G} = \{x_i\}_{i=1}^N$, *i.e.*, $\mathcal{G} = \bigcup_{1 \le i \le K} \mathcal{G}_i$, let $s_{\mathcal{G}_i}$ symbolize the abundance of information for group $\mathcal{G}_i$ based on particular criteria related to classes or structures (as depicted in Table 1). In this context, groups are sorted by descending order of information abundance (*i.e.*, if $i < j$, then $s_{\mathcal{G}_i} \ge s_{\mathcal{G}_j}$). Thus, the *imbalance ratio on graphs* can be defined as $s_{\mathcal{G}_1}/s_{\mathcal{G}_K}$. In essence, this imbalance ratio on graphs quantifies the skewness in distribution across groups concerning their information abundance—a larger ratio denotes a more skewed distribution. Crucially, this generalized definition of the imbalance ratio provides a more flexible tool for evaluating the wide array of imbalance issues within graphs.

To better grasp graph imbalance issues, we have compiled them in Table 1 to demonstrate their commonalities and disparities. For each type of imbalance, we enumerate the associated imbalance tasks, all of which are covered in this survey. Based on these tasks, we then outline the main settings pertinent to each task. Furthermore, we introduce the concept of information abundance s, as discussed earlier, to pinpoint the imbalance resources that contribute to the imbalance issues. Lastly, for clarity, we furnish detailed explanations for each task.

3 OVERVIEW OF TAXONOMIES

We categorize the literature from the perspectives of problems and techniques to provide a comprehensive overview.

Taxonomy of Problems. In terms of the taxonomy of *problems*, our summarization begins by investigating the research tasks of ILoGs. Imbalance issues on graphs are multifaceted, stemming from the multi-faceted nature of graph data. Specifically, graph imbalance can arise from two key sources: *classes* and *structures*, which usually serve as the key inputs to learning models. On one hand, for classes in graphs, labeled data may be distributed unevenly across classes, resulting in imbalanced outcomes in tasks such as node classification [25], [35]. On the other hand, the unique graph structures can introduce another source of imbalance in graph learning models. For instance, the neighborhood abundance of nodes [27], [57], can result in discrepancies in the exposed information, which may further skew the output of the learning model accordingly.

Therefore, we first categorize graph imbalance issues based on their types, *i.e.*, *class* and *structure imbalance*, both arising from imbalanced input. Secondly, considering the specific graph structures, we further categorize the imbalance issues into finer groups based on the graph-related elements, *i.e.*, *node-*, *edge-*, and *graph-level imbalance*, to gain a more detailed perspective on these issues. Finally, ILoGs includes a variety of research tasks, which we use as the finest categories for illustration, such as imbalanced node classification [35]–[37], few-shot node classification [26], [58], [59], *etc.* Fig. 3 illustrates the taxonomy of problems.

Taxonomy of Techniques. For the taxonomy of *techniques*, we categorize the literature based on imbalance forms (*i.e.*, what kind of imbalance) and the corresponding strategies to mitigate the imbalance (*i.e.*, how this imbalance is addressed), as summarized in Fig. 4. This is because specific forms of imbalance may require particular techniques to effectively address the associated issues, and the techniques employed may vary across different imbalance forms. Initially, we separate data into high- and low-resource segments (as shown in Fig. 2(b)) based on their information abundance, which results in two primary task branches contingent on the prediction target: *improving the low-resource part*, and *balancing the high/low-resource parts*. The former seeks to improve the performance of the low-resource part without considering the other part; while the latter strives for a balance in performance across both parts, given that
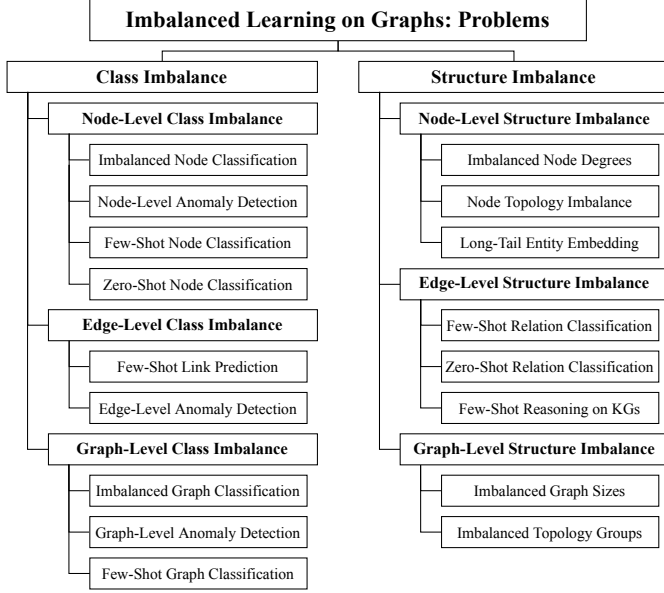
Fig. 3: Taxonomy of Problems.

Techniques		Literature
Algo-level	Constraints Knowledge distillation	DRGCN [35], DPGNN [61], TAM [62] LTE4G [63]
Data-level	GAN SMOTE Mixup Resampling Reweighting	DRGCN [35], ImGAGN [37] GraphSMOTE [25] GraphMixup [65], GraphSANN [67], GraphENS [66] LTE4G [63], ALLIE [69] TAM [62]

TABLE 2: Summary of imbalanced node classification.

predictions will be drawn from both. Subsequently, we summarize the major techniques employed to address the imbalance issues in each branch, and further provide detailed illustrations of techniques under each main category.

Organization. In the following three sections, we introduce the two taxonomies of ILoGs. Specifically, Sections 4 and 5 address the problems of ILoGs from the standpoints of class imbalance and structure imbalance, respectively. In Section 6, we classify the relevant literature based on the techniques used to address imbalance issues and offer guidance to help readers choose appropriate solutions. Subsequently, we discuss other studies related to imbalanced learning on graphs in Section 7. In Section 8, we explore future research directions from the perspective of both problems and techniques, and Section 9 concludes the survey.

4 ILoGs PROBLEMS: CLASS IMBALANCE

In the vast landscape of ILoGs, this section offers a succinct overview of class imbalance issues at the node (Section 4.1), edge (Section 4.2), and graph levels (Section 4.3), as illustrated in the left segment of Fig. 3. Moreover, structure imbalance will be discussed in Section 5.

4.1 Node-Level Class Imbalance

Node-level class imbalance refers to the disproportionate distribution of labeled nodes across classes. The model often learns more from classes with numerous labeled instances, at the risk of overlooking those with fewer. This challenge surfaces as four principal node-level tasks, which we will detail in the following sections.

4.1.1 Imbalanced Node Classification

Imbalanced node classification is a common challenge in real-world applications where labeled nodes are unevenly distributed across classes (as depicted in Table 1), such as in fake account detection or abusive review detection [19]. This

imbalance usually leads to a performance bias, as majority (high-resource) classes tend to outperform minority (low-resource) classes due to their larger number of labeled nodes [25], [35]. This imbalance has incited a growing number of studies [25], [35], [37], trying to enhance the performance of minority groups. For clarity, we categorize these studies into two perspectives: algorithm-level and data-level solutions.

For *algorithm-level* strategies, early work like DRGCN [35] implements a GNN-based approach, featuring a generative adversarial network (GAN) [60] for synthetic node generation to level the class distribution, paired with a KL-divergence constraint to synchronize the representation distribution of unlabeled nodes with labeled ones. Differently, DPGNN [61] employs a class prototype-driven training approach to balance training loss across classes, with the help of distance metric learning to accurately capture the relative positions of nodes concerning class prototypes. TAM [62] addresses the class imbalance issue by accounting for the reduced homogeneity among minority nodes. Specifically, it innovatively introduces connectivity- and distribution-aware margins to guide the model, emphasizing class-wise connectivity and neighbor-label distribution. LTE4G [63] considers the imbalance in both node classes and degrees. It partitions nodes into balanced subsets assigned to expert GNNs, then employs knowledge distillation to train class-specific students, to enhance classification performance.

The *data-level* approaches are typically designed to balance data distribution through tail class oversampling [16], head class undersampling [64], or instance-based class reweighting [14]. On graphs, GraphSMOTE [25] and ImGAGN [37] both revolve around the generation of synthetic nodes to balance classes. GraphSMOTE achieves this through a technique inspired by SMOTE [16], generating new node representations by averaging two sampled minority class nodes. ImGAGN, in contrast, applies GAN to generate synthetic nodes. GraphMixup [65], GraphENS [66], and GraphSANN [67] offer a unique approach to synthetic node generation, drawing inspiration from Mixup [68]. They create new nodes by merging semantic features with contextual edges, promoting balanced model training. Notably, GraphENS places greater emphasis on minority node neighbors, acknowledging their informational bias and employing neighbor sampling along with saliency-based node mixing to mitigate it. ALLIE [69] deviates from synthetic node generation, employing active learning instead. It involves sampling instances from both majority and minority classes, with an imbalance-aware reward function implemented via reinforcement learning [70] aiding the sampling process.

Summary. The core challenge in imbalanced node classification lies in *achieving balanced information distribution across classes for uniform model training*. Algorithm-level methods

Graph objects	Graph types	Base models	Literature
Node-level	Homogeneous	GA GE GNN	[81]–[91] [92]–[97] [77], [98]–[124]
	Heterogeneous	GA GE GNN	[125]–[131] [132], [133] [36], [75], [76], [123], [134]–[149]
	Dynamic	GA GE GNN	[80], [150]–[157] [158]–[162] [74], [163]–[167]
Edge-level	Homogeneous	GNN	[168], [169]
	Heterogeneous	GNN	[170]
	Dynamic	GA GE GNN	[78], [171] [79] [172], [173]
Graph-level	Homogeneous	GA GNN	[174]–[179] [72], [73], [180], [181]
	Heterogeneous	GA	[182]
	Dynamic	GA GNN	[183], [184] [185]

TABLE 3: Summary of anomaly detection on graphs.

addressing imbalance often leverage strategies such as additional constraints [35], prototype-driven training [61], ensuring homogeneity among minority nodes [62], or employing knowledge distillation [63]. As summarized in Table 2, a significant portion of research in this area tackles this issue from the data-level. This commonly involves creating synthetic nodes for underrepresented classes, or opting for reweighting or resampling nodes from both the majority and minority classes. For a more in-depth understanding of these techniques, please refer to Section 6.2. Note that, there is no significant gap between algorithm- and data-level approaches, as some methods, like DRGCN and TAM, can incorporate elements from both. Despite extensive research, this area still calls for further exploration. Innovative techniques, such as using diffusion models [71] to generate synthetic nodes or employing transfer learning to transfer knowledge from majority to minority classes for their enhancement, could be promising avenues of investigation.

A particular type of imbalanced node classification, node-level anomaly detection, has sparked increasing interest recently. We will separately provide a comprehensive review of the associated literature in Section 4.1.2.

4.1.2 Node-Level Anomaly Detection

Node-level anomaly detection, as a specific manifestation of imbalanced node classification, generally centers on binary classification between normal and anomalous nodes. This task plays a vital role in numerous applications, such as outlier detection [72], [73], fraudulent user detection [74]–[77], and social spammer detection [78]–[80]. Considering the diversity of graphs (*e.g.,* homogeneous, heterogeneous, or dynamic graphs), a range of techniques have been developed to address this issue, such as traditional graph algorithms (GA), graph embedding approaches (GE), and graph neural network approaches (GNN). They are categorized in Table 3, which also includes edge- and graph-level anomaly detection approaches. While we provide an overview of recent literature in this domain, for an in-depth analysis, we refer readers to the comprehensive survey [28].

Homogeneous graphs. Node-level anomaly detection on homogeneous graphs is a focal point of interest within this task. Traditional graph algorithms (GA) address anomaly detection by capturing structural anomalies through various handcrafted metrics such as applying residual analysis on attributes [82], [85], or utilizing human feedback in an interactive manner [87]. Following this, graph embedding (GE) based methods were developed, aiming to separately embed anomalous and normal nodes to capture their specific patterns [93], [95], or to generate representations that can form clear boundaries between classes [96], among others.

Recently, GNNs are widely used for anomaly detection on graphs. These methods are usually coupled with different learning objectives to obtain distinguishable node embeddings. We outline some popular techniques below. Contrastive learning methods typically leverage the similarity between target nodes and their context during model pre-training. For graph data containing suspicious nodes—nodes that often attempt to camouflage themselves—identifying their dissimilarity to the context (*i.e.,* heterophily) proves beneficial for their detection [107], [111]. Generative models like GANs are also exploited to create synthetic nodes during training, so the model can learn the distinct distributions of normal nodes as well as the out-of-distribution anomalous nodes [102], [110]. In addition, autoencoders are also employed to reconstruct graph structures or node features, where hard-to-reconstruct samples are usually identified as anomalies [102], [106].

Heterogeneous graphs. Anomaly detection is extensively studied in the context of heterogeneous graphs as well. In early stages, traditional graph algorithms (GA) usually define anomalous structural patterns on heterogeneous graphs to identify anomalies, such as detecting dense subgraphs [127], [128], or finding suspicious node attributes in clusters [126]. Later in the literature, more GE-based methods are proposed for heterogeneous graphs, such as characterizing different types of malware using meta-path based random walks [132], or identifying fraudsters based on their propensity to form unidirectional links with normal users [133].

In addition, many recent studies leverage heterogeneous GNNs to concurrently address graph heterogeneity and imbalance challenges, employing a variety of effective techniques. For instance, some studies [134], [144] resort to reinforcement learning to select the most informative neighbors, and jointly consider multiple relations to learn distinguishable node embeddings. Similarly, some studies [36], [136] utilize neighborhood sampling techniques to tackle class imbalance and aggregate multiple relations to differentiate normal and anomalous neighboring nodes. Additionally, spectral-based methods [141], [147] segregate aggregations of normal and anomalous nodes using distinct pass filters and manage multiple relations individually.

Dynamic graphs. Node-level anomaly detection is also widely explored within the realm of dynamic graphs. Early GA approaches address anomaly detection on dynamic graphs typically by identifying changes or unusual patterns within the graph structure, such as detecting temporal bursts [151] or tracking sudden changes on anomaly scores [155]. Additionally, GE-based methods are proposed to detect graph anomalies by learning distinguishable node embeddings, such as applying support vector data description (SVDD) to differentiate time-sensitive anomalous patterns

[158], using autoencoder to learn dynamic node embeddings and detect anomalous nodes by their distance to cluster centers [159], or combining autoencoders with hypersphere learning to isolate the spatiotemporal anomalies [160].

Several GNN-based methods are also proposed to address anomaly detection on dynamic graphs [74], [163]–[167]. For example, STGAN [164] involves a spatiotemporal sequence generator and a discriminator that determines whether an input sequence is real or not. During inference, the discriminator reports samples that have low scores as anomalies due to their out-of-distribution representations. Additionally, AddGraph [172] leverages an extended temporal graph convolutional network to discern both long-term and short-term patterns in dynamic graphs, using a margin loss to learn distinctive representations for anomalous and normal nodes.

Summary. Node-level anomaly detection on graphs is inherently challenging due to the intricate nature of graph structures and the infrequency of anomalies. Prevalent techniques in this arena largely focus on establishing distinct boundaries between normal and anomalous nodes. This can be achieved through feature engineering (GA), discriminative node representations (GE), or neighborhood aggregation for structure encoding (GNN). For a more comprehensive insight into state-of-the-art comparisons, readers may refer to benchmarks such as [186] or visit leaderboards[1]. To further propel progress in this domain, embracing innovative methodologies is essential. For example, sophisticated generative learning models, like diffusion [71], can be harnessed to bolster synthetic instance generation, thereby enhancing the discernibility of anomalous nodes. Furthermore, innovative foundational models [187] hold promise for enriching the comprehension of graph structures, and thus can potentially facilitate more effective anomaly detection.

4.1.3 Few-Shot Node Classification

In various real-world settings, *novel classes* are often encountered, marked by the availability of only a few labeled instances or *few-shot* instances. Such scarcity imposes significant hurdles for model training. To counter this, a set of *base classes* abundant in labeled data is typically harnessed to assist the learning process [42], as summarized in Table 1. Thus, the primary objective of *Few-Shot Node Classification* (FSNC) on graphs is to yield robust and superior performance on these few-shot novel classes by efficiently exploiting the information derived from the base classes, and this task has recently garnered considerable attention [58], [188]–[190]. This task finds applications in several contexts, such as novel intrusion detection on traffic networks, predicting new types of goods in e-commerce, and anticipating newly discovered chemical properties in protein networks, *etc.* In this section, we will delve into FSNC across various settings.

Generic FSNC. Meta-learning [205], with a primary focus on MAML [206] and prototypical networks [207], is a common approach in many studies addressing FSNC.

MAML-based approaches. Meta-GNN [58] set a precedent in the field by utilizing meta-learning to transfer meta-knowledge from data-rich base classes to novel classes using

a MAML-based adaptation [206]. Building on this, AMM-GNN [188] introduces attribute-level adaptation using FiLM [208] to address potential attribute disparities across meta-tasks, complementing the MAML adaptation. TRGM [190] adds another dimension by capturing inter-task relations via contrastive learning in a constructed task graph, thereby enhancing few-shot learning. In a distinct approach, RALE [189], while rooted in MAML, emphasizes both relative and absolute node positions for node placement, aiming to bolster representations of nodes in novel classes.

Prototypical network-based approaches. GFL [26] utilizes prototypical network to enable cross-graph FSNC, and further introduces a hierarchical graph representation gate to regulate global transferable knowledge. Moreover, GPN [59] enriches this paradigm by incorporating a node valuator to re-weight labeled nodes according to their informativeness in prototype calculation. In addition, MuL-GRN [194], akin to RALE [189], calculates relations at node, global subgraph, and local subgraph levels between query and support nodes to facilitate the classification of novel class nodes.

Some approaches merge the benefits of prototypical networks with adaptation-based methods like MAML or FiLM. ST-GFSL [209] and G-Meta [191], building on the prototypical network, employ MAML for optimization to tackle FSNC, with the former addressing traffic flow prediction on spatio-temporal graphs and the latter emphasizing local subgraph utilization for target node representation. TENT [192] tackles variance across meta-tasks via node-, class-, and task-level adaptations. Moreover, Meta-GPS [193] refines aggregation and combination functions in GNNs to consider both graph homophily and heterophily, employing FiLM [208] to modulate model parameters per meta-task.

Other techniques. While most existing studies employ meta-learning for FSNC, alternative techniques have also been proposed. For instance, IA-FSNC [195] initializes novel classes with the first layer of a pre-trained GNN model, and employs a semi-supervised method for synthetic label generation to realize information augmentation for the support sets. In addition, Tan *et al.* [196], [197] employ contrastive learning, by constructing a subgraph around a target node based on connectivity and generating three types of contrast pairs (node-node, node-subgraph, and subgraph-subgraph) for model training. Another work GraphPrompt [198] employs GNN prompting techniques, unifying pre-training and downstream tasks for effective knowledge transfer, achieving impressive results even with limited supervision.

Generalized FSNC. In a broader FSNC setting, both base and novel class nodes appear in the test set rather than only novel classes. This complicates matters by introducing challenges like asymmetric classification and inconsistent preference between base and novel classes [201]. Stager [201] addresses this by using prediction uncertainty to measure class shot levels and decompose prediction probability, tackling asymmetric classification. It also employs a meta-learner to weigh various receptive fields, addressing the inconsistent preference challenge.

Other works [199], [200] focus on class-incremental FSNC, where novel classes continually emerge. Both HAG-Meta [199] and Geometer [200] address this challenge by establishing an attention-based framework on the founda-

Tasks	Meta-learning techniques			Other techniques		
	MAML	Prototypical network	Others	Label generation	Contrastive Learning	Prompting
Generic FSNC	[58], [188]–[190]	[26], [59], [191]–[194]	-	[195]	[196], [197]	[198]
Generalized FSNC	-	[199], [200]	[201]	-	-	-
Multi-label FSNC	-	[202]	-	-	-	-
FSNC with extremely weak supervision	-	-	-	[203]	-	-
FSNC on HINs	[30], [204]	-	-	-	-	-

TABLE 4: Summary of few-shot node classification.

tion of the prototypical network, allowing for the weighted calculation of prototypes from labeled instances. HAG-Meta extends its approach beyond GNN pre-training on base classes, using an attention mechanism to adjust the significance of meta-tasks. On the other hand, Geometer applies geometric loss functions to maintain intra-class proximity, inter-class uniformity, and inter-class separability, effectively tackling the challenge of class incrementality.

Multi-label FSNC. While the majority of prior research primarily focuses on multi-class node classification, the task of few-shot multi-label node classification, where each node may be associated with multiple labels instead of just one, continues to pose a considerable challenge. To address this issue, MetaTNE [202] treats it as a label-wise binary classification problem. It identifies discrepancies in node embeddings across different meta-tasks and implemented an embedding transformation. This transformation, contingent on the query node within each task, allows it to obtain task-specific representations beneficial for classification.

FSNC with extremely weak supervision. In the challenging scenario of FSNC with extremely weak supervision, where only a limited number of labeled nodes are available during meta-training, models typically struggle to extract sufficient prior knowledge for effective knowledge transfer. To overcome this limitation, X-FNC [203] introduces a novel approach to generate pseudo-labels as additional references for the base classes, in order to learn effectively from extremely weak supervision.

FSNC on HINs. Recent research has delved into FSNC on heterogeneous graphs, necessitating the management of heterogeneity. To enhance classification in the target domain, CrossHG-Meta [210], building on heterogeneous graph neural networks (HGNN) [211], [212], extracts transferable meta-knowledge from a source domain. It employs dual adaptations at the domain- and task-levels for model optimization, along with a cross-domain contrastive regularization. Similarly, HG-Meta [204], another study founded on HGNN, tackles differences across meta-tasks through task-level modulation. It applies task feature scaling to adjust node representations and incorporates an attention mechanism to re-evaluate the importance of each meta-task.

Summary. The primary challenge in few-shot node classification is to *extract transferable knowledge from base classes to benefit novel classes*. Recent studies in this area, as summarized in Table 4, have focused on various aspects of FSNC, including generic, generalized (class incremental), multi-label, extremely weak supervision settings, and on HINs. Methods frequently employed include meta-learning tech-

niques like MAML and prototypical networks, sometimes used simultaneously. Yet, other studies have also ventured into utilizing innovative techniques such as data augmentation, contrastive learning, and prompting methods. For a more comprehensive understanding of these techniques, we direct readers to Section 6.1.1. Despite considerable research, this complex task continues to necessitate further exploration. As demonstrated in Table 4, specific settings like generalized, multi-label, extremely weak supervision, or those concerning HINs, remain largely underexplored. Notably, the adoption of innovative techniques like prompt tuning [198], generative models [71], and others could potentially enhance the representation of novel classes, indicating promising future directions. For a broader perspective, we suggest consulting the survey of few-shot learning on graphs [30] for additional insights.

4.1.4 Zero-Shot Node Classification

Zero-Shot Node Classification (ZSNC) necessitates the absence of labeled data for novel classes during model training. Wang *et al.* [213] build on DeepWalk [6], modulating the objective constraints on base classes to learn effective representations for unlabeled novel classes. They relax the intra-class similarity constraint and remove inter-label edges to enhance dissimilarity. Its further extensions [214], [215] utilize GCN as the backbone, adding label-related semantic descriptions. Differently, DGPN [216] provides a formalization for ZSNC, introducing class semantic descriptions and corresponding evaluation metrics (based on the proximity of text-based class embeddings and proximity in real data). They refine GNN models by constraining the similarity between hidden embeddings (in each layer) and the class embeddings, to enhance the model training. Similarly, another work DBiGCN [217] also follows the paradigm of utilizing class descriptions for ZSNC.

Summary. ZSNC remains an underexplored field due to the absence of descriptions for elements like nodes, edges, or graphs, which are crucial in typical zero-shot settings. Therefore, this challenging task warrants further research attention. Currently, the existing strategies involve constraining model training on base classes, while others utilize auxiliary information, such as descriptions, to supplement model training. The metrics established by previous work, such as [216], lay a promising foundation for further exploration in this challenging domain.

4.2 Edge-Level Class Imbalance

In addition to node-level class imbalance, addressing edge-level class imbalance is also crucial. Two representative

tasks of interest are few-shot link prediction and edge-level anomaly detection. These tasks underscore the importance of tackling edge-level class imbalance in various real-world scenarios, such as inductive recommendation [31], anomalous transaction detection on financial networks, *etc.*

4.2.1 Few-Shot Link Prediction

Recent studies have focused on few-shot link prediction across graphs, particularly in event-based social networks [31]. This presents a challenge as the novel graphs in this context have limited edges and pose difficulties for link prediction. To address this, EA-GAT [31], an event-aware graph attention network, attempts to effectively encode fine-grained events from existing graphs and few-shot target events. It further utilizes gradient-based episode learning to acquire transferable knowledge and adapt to event-based social networks with sparse connections.

Additionally, other settings such as few-shot link prediction across different sections of a single graph, which might demonstrate varied patterns [218], could also be explored as a potential avenue for research.

4.2.2 Edge-Level Anomaly Detection

Edge-level anomaly detection is designed to identify anomalous edges within a network. This endeavor is crucial in areas like transactions and social networks to detect potentially dubious transactions or fake news. Prevailing research primarily focuses on characterizing edge distributions and pinpointing anomalies that exhibit significant deviations from these established distributions.

Existing works on homogeneous graphs tend to use graph auto-encoders to model the distribution of normal edges and use reconstruction errors as indicators of anomalies [168], or enrich node and edge features using auxiliary descriptions [169] for better discrimination. On heterogeneous graphs, research often tries to capture both heterogeneity and anomalous patterns, such as using type-aware edge representations and tailored classifiers to identify fraudulent transactions on live-streaming platforms [170].

Some works consider detecting abnormal edges in dynamic graphs by keeping track of edge dynamics [78], [79], [171]–[173]. For instance, Miz *et al.* [79] use the Hopfield model of memory to combine graph and temporal information and detect temporal spikes in activities as anomalies. Another work StrGNN [173] extracts subgraph snapshots around an edge and applies a GNN model with gated recurrent units [219] to capture subgraph structural changes as anomaly indicators.

Summary. The main challenge in edge-level anomaly detection lies in the highly imbalanced distribution of normal and abnormal edges. To counter this challenge, previous works usually model edge distributions based on the abundant information on normal edges, using deviations from the normal distributions as anomaly indicators. Despite its significance, this area remains relatively underexplored. There is ample room for further investigation, especially in tasks involving heterogeneous graphs where edge relations are intricate and significant.

4.3 Graph-Level Class Imbalance

Addressing graph-level class imbalance forms a vital aspect of ILoGs, surpassing the scope of node- and edge-level imbalance. Three crucial areas warrant attention within this domain: imbalanced graph classification, few-shot graph classification, and graph-level anomaly detection. They highlight the importance of addressing graph-level class imbalance across a variety of real-world scenarios, such as novel drug detection, and protein function prediction.

4.3.1 Imbalanced Graph Classification

Imbalanced graph classification parallels the challenges of imbalanced node classification. It commonly arises in real-world scenarios (*e.g.,* imbalanced chemical compound classification) where class distributions of labeled graphs are skewed (as illustrated in Table 1), often favoring the majority class with more labeled graphs [32].

To counter this, G^2GNN [32] utilizes additional supervision both globally, through neighboring graphs, and locally, via stochastic augmentations. It builds a Graph of Graphs (GoG) by leveraging kernel similarity and implements GoG propagation for information aggregation. Additionally, topological augmentation paired with self-consistency regularization is employed at the local level. These strategies collectively enhance model generalizability, thereby elevating classification performance.

Despite the advance, this crucial task continues to call for more in-depth study due to the limited amount of current research, indicating a potentially promising direction.

4.3.2 Graph-Level Anomaly Detection

Graph-level anomaly detection aims to detect anomalous graphs or subgraphs according to a predefined anomaly measure, which can be seen as a variant of imbalanced graph classification with binary labels. This task is commonly seen in transaction networks where fraudsters conspire to do money laundering.

Much of the emphasis has been placed on *homogeneous graphs*. A segment of research employs traditional graph algorithms (GA) to identify anomalous subgraphs [174]–[179]. For instance, AMEN [175] defines a normality measure and discovers communities with an aim to maximize their normality scores, where communities that cannot achieve high normality scores are considered to be fraudulent. Chen *et al.* [179] employ Benford's law to assign anomaly scores to edges and use the densest sub-graph discovery algorithm to find anomalous sub-graphs.

Another line of works uses GE-based methods to analyze the anomaly degree of graphs [72], [73], [180], [181]. For instance, OCGTL [72] concentrates graph embeddings in the training set within a hypersphere where the distance to the hypersphere is used as the anomaly degree. AS-GA [181] identifies suspicious sub-graphs based on the reconstruction error of a graph auto-encoder and rates anomaly scores of extracted areas with a graph supermodular neural network.

Some studies consider graph-level anomaly detection on *heterogeneous graphs*. For instance, ACGPMiner [182] introduces conditional graph patterns to model abnormal patterns in property graphs and follows a generation-and-validation paradigm to mine the defined abnormal patterns and their matches.

Tasks	Meta-learning techniques		Other techniques
	MAML	Prototypical network	
Generic FSGC	[220]	[221], [222]	Adaptive step controller [220], super-class graph [223], task correlations [221]
Cross-domain FSGC	-	-	Data augmentation [224]
Few-shot temporal graph classification	[225]	[225]	-
Few-shot molecular property prediction	[226]–[229]	-	Meta-task reweighting [227], implicit function theorem [230]

TABLE 5: Summary of few-shot graph classification.

In *dynamic graphs*, existing works usually try to detect the deviation of graph structures among different time steps [183]–[185]. For instance, Zambon *et al.* [184] propose to embed each graph in the graph stream into a vector and perform change detection on the converted vector stream to flag unusual graph patterns. MTAD-GAT [185] employs self-attention to capture temporal relations across different time steps for nodes and aggregates node-level anomaly scores to generate graph-level anomaly scores.

Summary. Graph-level anomaly detection is a logical progression from node- and edge-level tasks. Many current approaches determine node- or edge-level anomaly scores and subsequently aggregate these to gauge graph-level anomalies. An alternative strategy involves embedding entire graphs into vectors, effectively transforming graph-level anomaly detection into the well-established realm of vector-based anomaly detection. Presently, the primary focus lies on homogeneous and dynamic graphs. Nevertheless, detecting anomalies at the graph-level on heterogeneous graphs remains a crucial area meriting further exploration.

4.3.3 Few-Shot Graph Classification

Few-Shot Graph Classification (FSGC), a crucial facet of imbalanced learning on graphs, aims to categorize target graphs into respective novel classes, commonly applied in real-world scenarios such as novel molecular or protein prediction. Like few-shot node classification, the objective of FSGC is to distill transferable knowledge from base classes and apply it to novel classes, thereby enhancing the classification performance for the latter.

Generic FSGC. To address the challenge of FSGC, several noteworthy studies offer distinct solutions. AS-MAML [220] employs MAML [206] to extract meta-knowledge and uses a reinforcement learning-based adaptive controller to adjust the step size [231]. In contrast, Chauhan *et al.* [223] utilize distance metrics to cluster closely related graph classes into super-classes, forming a super-graph. This structure facilitates hierarchical classification at both super-class and class levels, thereby highlighting the importance of inter-class relationships. Different from the class correlations, FAITH [221] focuses on the role of task correlations in knowledge transfer. They build a three-layer hierarchical graph capturing sample-, prototype-, and task-level data to facilitate classification, emphasizing the role of task correlations in model performance. In addition, Crisostomi *et al.* [222] combine a simple metric learning model with advanced graph embedding techniques, demonstrating excellent results. Their strategy of using a task-adaptive mechanism for layer-wise adaptation and Mixup data augmentation showcases

an effective blend of traditional and novel techniques for superior performance.

Cross-domain FSGC. For cross-domain FSGC, Hassani *et al.* [224] devise three graph augmentations, including contextual and two topological augmentations, to enhance representation learning, particularly task-specific information for fast adaptation and task-agnostic information for knowledge transfer.

Few-shot temporal graph classification. Fu *et al.* [225] explore a novel setting of few-shot temporal graph classification, where each class may have only a few labeled temporal graphs and novel classes might emerge in the future. They present Temp-GFSM [225], a temporal graph metric learning framework. Utilizing attention mechanisms like node-level lifelong attention and both intra- and inter-snapshot attention, they derive temporal graph representations. They are subsequently used in few-shot metric learning, backed by prototypical networks and MAML.

Few-shot molecular property prediction. In the realm of few-shot molecular property prediction, several studies exploit meta-learning (*e.g.*, MAML) to bridge different tasks. PAR [226] applies MAML, utilizing detailed molecular representations to enhance prediction performance. Meta-MGNN [227] also employs MAML, but enhances its approach by applying self-supervised constraints for graph knowledge extraction and an attention mechanism to weight meta-task contributions. Similarly, Borde *et al.* [228] highlight the applicability of Reptile [232], a MAML-like model-agnostic algorithm, with GNN models in molecular property regression tasks.

Additionally, various other techniques are also employed. ADKF-IFT [230] integrates meta-learning with deep kernel Gaussian processes [233], proposing adaptive deep kernel fitting for versatile feature representation learning across tasks. In contrast, to combat overfitting and improve generalizability, MTA [229] generates new labeled samples using motifs from a pre-defined vocabulary, enabling connections between tasks through these augmented motifs.

Summary. Few-shot graph classification presents the significant challenge of *effectively transferring knowledge from base graph classes to novel graph classes to enhance the performance of the latter*. This challenge has inspired a multitude of research, showcasing diverse strategies to tackle this issue, as summarized in Table 5. These studies not only leverage meta-learning strategies, but also employ additional techniques to bolster performance, such as data augmentation, super-class graph construction, *etc.* Nonetheless, certain specific settings within this critical task still demand further exploration. For instance, in a cross-domain scenario, the transfer of knowledge from base to novel classes may pose a greater

challenge; on temporal graphs, the complexity of few-shot classification is further amplified by the temporal patterns. Furthermore, few-shot graph-level classification on HINs could also present an intriguing research direction.

5 ILoGs Problems: Structure Imbalance

Unlike other data types such as images or text, graph data inherently contains topological structures that may exhibit imbalance. The investigation of structure imbalance in graphs has increasingly drawn the attention of researchers due to its prevalence in real-world scenarios. In this section, we extend our discussion for the taxonomy of Problems, examining structure imbalance in graphs from the perspectives of node (Section 5.1), edge (Section 5.2), and graph (Section 5.3) levels, as depicted in the right part of Fig. 3.

5.1 Node-Level Structure Imbalance

Node-level structure imbalance arises when the contextual structures surrounding each node exhibit an unequal distribution. A prominent indicator of node-level structure is node degree, representing the count of neighboring nodes and reflecting the vicinity richness of a node. In this section, we explore node-level structure imbalance from two perspectives: imbalanced node degrees and topologies, as summarized in Table 6. These aspects have significant implications for various real-world applications, such as cold-start recommendations, and knowledge graph enrichment.

5.1.1 Imbalanced Node Degrees

In graphs, node degrees often follow a long-tail distribution, with *head nodes*—those with high degrees—benefiting from richer structural information and thus achieving superior performance in downstream tasks like node classification [18], [27]. Conversely, *tail nodes* with low degrees have limited topological information, hindering their performance. Research typically targets two main tasks: improving the performance of tail nodes, and addressing the more challenging *cold-start nodes* that are isolated in the graph. These tasks find applications in various domains, such as predicting new papers in a citation network or new items in an e-commerce network.

Tail node embedding. Tail node embedding seeks to boost performance for tail nodes, which often underperform compared to head nodes. To tackle this issue, Demo-Net [234] and SL-DSGCN [235] utilize degree-specific GNNs to capture unique structural patterns across nodes of different degrees. Contrastingly, meta-tail2vec [18] and Tail-GNN [27] implement knowledge transfer mechanisms. The former leverages meta-learning to transfer knowledge from head to tail nodes, redefining tail node embedding as a few-shot embedding regression task. Meanwhile, the latter proposes a neighborhood translation technique to bridge the head and tail nodes for knowledge transfer.

Degree imbalance has also been explored in various other tasks. For instance, Residual2Vec [239] delves into the task of fair path sampling, where imbalance may stem from the differing degrees of nodes. To achieve fairness, it reweighs the loss function based on both node labels and degrees. In contrast, CenGCN [240] addresses the issue of

dominance by hub vertices in scale-free networks, where these vertices can disproportionately propagate influential information due to vertex imbalance. Notably, CenGCN quantifies the similarity between hub vertices and their neighbors, and applies graph transformations through edge weight adjustments and self-connections, effectively mitigating this issue. Meanwhile, BLADE [241] confronts the degree discrepancy problem in link prediction by using biased neighborhood sampling. By creating neighborhoods of varying sizes based on node connectivity, it optimizes performance across different node degrees.

In *e-commerce networks*, tail node embedding aims to enhance the performance of less-frequent users and shops. One typical work DHGAT [238] achieves this by leveraging first- and second-order proximities of user histories, utilizing attentive mechanisms to further exploit the neighbors of target entities. Hao *et al.* [236], drawing from the concept of meta-tail2vec, apply regression to tail nodes to facilitate knowledge transfer, enabling cold-start recommendations.

In *dynamic networks*, newly arriving nodes typically confront structural limitations. To address this, MetaDyGNN [237] targets few-shot link prediction. It harnesses a meta-learning approach to create a hierarchical architecture that applies interval- and node-wise adaptations, thereby facilitating the knowledge transfer from existing to new nodes. The approach also includes a dynamic GNN model to efficiently exploit local node structures.

Cold-start node embedding. In certain extreme cases, graphs may contain numerous isolated nodes, known as cold-start nodes, which have no neighbors and thus pose a challenge for conventional graph representation learning approaches. To tackle this issue, techniques typically employ the knowledge distillation paradigm [252]. These approaches utilize a teacher network for knowledge extraction on information-sufficient nodes and a student network for knowledge application on cold-start nodes, ultimately improving the performance of the latter.

Cold Brew [57] is a notable approach designed to address cold-start node representation learning by leveraging the concept of knowledge distillation. Specifically, the method employs a GNN network as the teacher network to extract knowledge from information-sufficient nodes and a multilayer perceptron (MLP) as the student network to distill the knowledge from the teacher. Consequently, the student network can predict suitable representations for cold-start nodes without relying on neighborhood information.

Summary. The central challenge in managing imbalanced node degrees lies in the *efficient knowledge transfer from head nodes to tail or cold-start nodes*. Various strategies have emerged to address this, as summarized in Table 6. Notably, degree-aware model modulation and meta-learning are frequently utilized for tail node embedding, while knowledge distillation is favored for cold-start nodes. For further details of these techniques, we direct readers to Section 6.1.1. Certain areas still warrant further exploration. For instance, there is a notable dearth of research on the pivotal task of cold-start node embedding. Moreover, while knowledge distillation is typically utilized for cold-start node embedding, its potential for handling other similar tasks, such as tail node embedding, could also be investigated. Further-

Tasks	Degree-aware modulation	Meta-learning	Knowledge distillation	Other techniques
Tail node embedding	[234], [235]	[18], [236], [237]	-	Neighborhood translation [27] Hybrid-order proximities [238] Reweighting [239]–[241]
Cold-start node embedding	-	-	[57]	-
Node topology imbalance	-	-	-	Reweighting [242], [243] Graph geometric embedding [244]
Long-tail entity embedding on KGs	[245]	[246]–[249]	-	Open knowledge enrichment [250] Synthetic data generation [251]

TABLE 6: Summary of node-level structure imbalance.

more, we will discuss long-tail entity embedding on KGs, a related task, in Section 5.1.3.

5.1.2 Node Topology Imbalance

In contrast to the frequently studied class imbalance rooted in the quantity-centered disparity in the number of labeled samples across classes (*i.e.*, imbalanced node classification as discussed in Section 4.1.1), *topology imbalance* on graphs shifts the focus to the positional distribution of labeled samples. This positional discrepancy critically influences the label propagation process on the graph. Ideally, in label propagation, the influence boundaries of labeled nodes should coincide with the true class boundaries. However, the specific positions of labeled nodes can trigger a displacement of these influence boundaries away from the true class boundaries, resulting in skewed model decision boundaries. Classes with labeled nodes having influence boundaries more closely aligned with the true class boundaries tend to propagate label information more effectively than those without such alignment, leading to an imbalance, as illustrated in Table 1. This uneven distribution could potentially induce biases in learning representations, thereby impacting the performance of nodes across different classes [242].

To address this, ReNode [242] reweights labeled nodes based on their proximity to class boundaries, improving performance particularly for boundary-near and remote nodes. They also develop a metric to quantify this imbalance using influence conflict detection. Another work PASTEL [243] combats topology imbalance by optimizing information propagation paths. It aims to alleviate under-reaching and over-squashing effects by enhancing intra-class connectivity and employing a position encoding mechanism. PASTEL also uses a class-wise conflict measure for edge weights to facilitate node class separation.

In graph data with hierarchical structures, *hierarchy imbalance* represents an unequal distribution of labeled nodes across hierarchical levels. This issue may impede the effective learning and classification of nodes in underrepresented hierarchy levels, also affecting the decision boundary of the classifier [244]. To tackle this challenge, HyperIMBA [244] utilizes hyperbolic geometric embedding to gauge the hierarchy of labeled nodes. Subsequently, it adjusts label information propagation and alters the objective margin based on the hierarchy of the node, thereby addressing the issues caused by hierarchy imbalance.

Summary. In addition to the extensively studied imbalance issues (*e.g.*, class or degree imbalance) that affect the performance of node classification, the topology of nodes, which depicts the interconnectedness of nodes from the label information, remains underexplored. Specifically, given the importance of label propagation on graphs, the connecting topologies could play a pivotal role in the node classification task, indicating a need for further investigation in this area.

5.1.3 Long-Tail Entity Embedding on KGs

Knowledge graphs often present a long-tail distribution of triplets across entities, potentially causing performance bias [253] in applications such as KG enrichment. To address this issue, a series of approaches have been proposed.

GEN [246] is a transductive meta-learning framework predicting links for new entities. It uses aggregation on neighboring entities and relations, and employs MAML [206] for knowledge transfer from seen (head) to unseen (tail) entities. The works of MaKEr [247], MTKGE [249], and MorsE [248] follow a similar paradigm, employing both meta-learning and GNN encoding to tackle this issue. Specifically, MaKEr operates within a federated setting, basing its process on representations calculated on their constructed relation position graph. MTKGE, on the other hand, concentrates on the relative position and temporal sequence patterns between relations to extrapolate missing facts in emerging temporal knowledge graphs. MorsE, meanwhile, focuses on inductive knowledge graph embedding, by incorporating two key modules—an entity initializer and a GNN modulator—for enhanced entity embedding. In addition, KG-Mixup [251] generates synthetic triples with Mixup for data augmentation, aiming to enhance KG completion performance, particularly on tail entities.

Some approaches also leverage additional knowledge to address this issue. For instance, Zeng *et al.* [245] propose a degree-aware co-attention network to guide information fusion from different sources for tail entity alignment, forming an iterative paradigm to improve KG completion. Another work OKELE [250] aims to increase tail entity triplets by extracting open knowledge from the Web. It constructs an entity-property graph on the given KG for attentive GNN-based representation learning.

Summary. Tail entities on KGs share similar characteristics with tail nodes on graphs, as both have a scarcity of linked neighbors. Notably, some similar techniques, such as meta-learning, are employed to address tail entity embedding. Moreover, supplementary knowledge, such as auxiliary KGs, serves as an additional source of information to aid in resolving this issue. Consequently, the solutions proposed for tail entity or node embedding can draw inspiration from each other to enhance their respective performance.

5.2 Edge-Level Structure Imbalance

Edge-level structure imbalance in complex graphs, like KGs, manifests as varying edge (relation) frequencies, often following a long-tail distribution [33], [254]. This imbalance can bias the model, marginalizing long-tail relations and leading to suboptimal performance—a challenge that is garnered significant research attention due to its implications in real-world scenarios. We delve into edge-level topology imbalance through three lenses: few-shot relation classification, zero-shot relation classification, and reasoning with few-shot relations, and summarize them in Table 7. Note that the frequency of edges reflects their structural distribution in KGs, hence we classify this aspect as structure imbalance. Alternatively, it can also be viewed as class imbalance, considering edges as labeled instances of the edge class.

5.2.1 Few-Shot Relation Classification

In scenarios such as KG enrichment, there are often *novel relations* that have only a few available triplets. In contrast, numerous *base relations* are provided to support classification tasks involving these novel relations. In this context, the objective of Few-Shot Relation Classification (FSRC) is to classify triplets into novel relations, with only a limited number of triplets serving as supervision.

Generic FSRC. Within the realm of FSRC on KGs using *metric-based meta-learning* methods [207], [282], various studies exhibit similarities in approach, albeit with distinct contributions. Certain investigations implement unique techniques for metric computation and encoding. For instance, Gmatching [33] and FSRL [260] employ matching networks, differing in their methods of entity representation. While Gmatching utilizes a GNN encoder, FSRL infers entity embeddings by encoding heterogeneous neighbors. Furthermore, MetaP [266] exploits co-occurrence relation patterns in meta-tasks to enhance the effectiveness of proximity calculations within matching networks.

A common approach revolves around the usage of attention mechanisms and prototypical networks. For instance, Gao *et al.* [254] and KEFDA [262] utilize attention-based prototypical networks. The former [254] is distinguished by its instance- and feature-level hybrid attention mechanism, enhancing prototype calculation accuracy. KEFDA [262], on the other hand, stands out by incorporating general and domain knowledge into the model. The theme of attention mechanisms continues in IAN [261] and HMNet [263], with each focusing on different aspects. IAN [261] computes inter- and intra-instance correlations to guide fusion operations, while HMNet [263] conducts dual scoring from both entity and relation perspectives for effective link prediction.

A number of studies present other contributions based on prototypical networks. Both FAAN [264] and work [269] utilize adaptive learning models to tackle FSRC. Specifically, FAAN introduces an adaptive attentional network to capture dynamic properties, while the latter proposes an adaptive mixture mechanism that incorporates label words into class prototypes. On the other hand, MULTIFORM [267] enriches entity representations by incorporating multi-modal contexts. And another research [268] adopts a granularity-aware method, depicting each relation as an area to account for variations in granularities.

Among studies employing *optimization-based meta-learning* methods for FSRC on KGs, MAML [206] serves as a key foundation. Both MetaR [255] and MTransH [256] employ MAML for link prediction and few-shot training respectively, with MetaR using an MLP for relation representation and MTransH integrating a gated and attentive neighbor aggregator. Li *et al.* [257] deviate by jointly training a prototype and instance encoder, enhancing the relation classification accuracy of the model. In addition, Mick [283], HiRe [258] and Meta-iKG [259] explore different facets: Mick explores FSRC from the perspective of limiting available data during training, HiRe focuses on hierarchical relational learning across different levels, while Meta-iKG utilizes local subgraphs for efficient pattern learning and generalization across few-shot and large-shot relations.

Some studies highlight *other techniques* or *aspects of KGs* for FSRC. For instance, Gao *et al.* [271] deploy an iterative neural snowball process with fake labels for novel relations, relying on a relational Siamese network [284] and a relation classifier. REFORM [272] uses an error mitigation mechanism alongside attentive neighborhood encoding and cross-relation aggregation for error-aware completion. In contrast, P-INT [275] encodes relations via expressive paths for accurate matching. Similar to IAN [261], CIAN [270] leverages attention mechanisms to capture both intra- and inter-entity interactions for better entity pair representations. Meanwhile, NP-FKGC [276] merges normalizing flows and neural processes to manage complex relations and estimate uncertainties, aided by an attentive relation path-based GNN to better capture KG path information.

FSRC with uncommon entities. Wang *et al.* [273] consider an extreme setting with uncommon entities and long-tailed relations. They use textual descriptions (*i.e.,* description encoder) to extract crucial information for entity and relation embeddings. To address data limitations, they propose a triplet generator to synthesize fake triplets for augmentation, augmenting both limited entities and relations. Reptile [285] is employed for model optimization.

FSRC on uncertain KGs. In uncertain KGs, where each triplet has a confidence score indicating its certainty, few-shot relation classification remains a challenge. Zhang *et al.* address this issue with Gaussian Metric Learning [274], which completes missing facts and confidence scores with few available examples. They propose a Gaussian neighbor encoder to represent relation facts as multi-dimensional Gaussian distributions, simultaneously learning semantic features and internal uncertainty. Additionally, they propose a Gaussian matching function considering fact qualities to discover new facts and predict confidence scores.

Few-shot inverse relation classification. For few-shot inverse relation classification, FAEA [265] employs a hybrid attention model to attend class-related words based on meta-learning. It leverages function-words enhanced attention to effectively compute representations for both support and query instances, facilitating message passing and similarity calculation between queries and prototypes.

Summary. The principal challenge of few-shot relation classification is *the effective transfer of knowledge from data-rich relations to long-tail relations*, aiming to bolster the performance of the latter. As depicted in Table 7, most existing

Tasks	Meta-learning techniques		Other techniques
	Optimization-based	Metric-based	
Few-shot relation classification	[255]–[259]	[33], [254], [260]–[270]	Neural snowball [271], error mitigation [272], data augmentation [273], gaussian metric learning [274], path-based interactions [275], entity interactions [270], neural process [276]
Zero-shot relation classification	-	[277]	GAN [278]
Few-shot multi-hop reasoning on KGs	[260], [279]–[281]	-	-

TABLE 7: Summary of edge-level structure imbalance.

approaches rely heavily on meta-learning techniques, such as prototypical networks and MAML, to counter the imbalance issue. Some methods also employ different techniques like the neural snowball [271], error mitigation [272], among others. Moreover, several extreme scenarios have also been explored, including few-shot relation classification with uncommon entities, within uncertain knowledge graphs, and few-shot inverse relation classification. Delving deeper into these extreme scenarios could further advance progress in FSRC. Additionally, benchmarks and leaderboards, such as FewRel[2] [286], [287], have been introduced for comparison.

5.2.2 Zero-Shot Relation Classification

Zero-shot relation classification on KGs, classifying triplets without given supervision, typically leverages auxiliary text descriptions for zero-shot relation representation. For instance, ZSGAN [278] employs GAN [60] to transfer knowledge from labeled to zero-shot relations. The generator aims to produce zero-shot relation embeddings using text descriptions, while the discriminator classifies the generated embeddings into respective classes. Putting it differently, ZSLRC [277] utilizes side information through hypernym and keyword extraction for new relation type detection. It then implements a side-information-enhanced prototypical network for few-shot relation classification by computing a weighted side information embedding for each relation.

5.2.3 Few-Shot Multi-Hop Reasoning on KGs

Multi-hop reasoning on KGs offers an effective method for inferencing target entities given query entities and relations, a technique crucial for tasks like query answering. However, it is often challenged by a limited number of labeled triplets, leading to a few-shot setting.

Efforts to tackle this issue, such as Meta-KGR [279] and FIRE [288], usually utilize reinforcement learning [231] for multi-hop reasoning and MAML to extract meta-knowledge to address the few-shot problem. To improve the generalization abilities, THML [280] further introduces a hardness-aware meta-reinforcement learning method. This method trains hardness-aware batches using a two-level hardness-aware sampling approach, addressing low reasoning accuracies over challenging relations. In a similar vein, ADK-KG [281] enhances the standard reinforcement learning and MAML modules with text-enhanced heterogeneous GNN for better node embeddings. It also incorporates a knowledge distillation module using unlabeled data to generate synthetic labels, further improving model training.

2. https://github.com/thunlp/FewRel.

5.3 Graph-Level Structure Imbalance

The intricate interconnections within graphs can lead to structural imbalance across graphs. This type of imbalance often manifests as differences in graph sizes [34], or topology groups [289], as illustrated in Table 1. Typically, graphs with advantageous structures, such as larger sizes, are more expressive and subsequently yield superior performance compared to their counterparts, which may introduce bias in applications such as molecular or protein prediction.

5.3.1 Imbalanced Graph Sizes

Graphs with larger sizes (*e.g.*, number of nodes), which offer more intricate structures, are generally more expressive than smaller ones. This disparity in size often results in performance biases in graph-level tasks such as graph classification [34]. To address this challenge, SOLT-GNN [34] is designed to enhance the performance of smaller, or "tail", graphs. SOLT-GNN first identifies co-occurrence patterns in the structures of larger, or "head", graphs, to generate transferable knowledge. This information is then utilized to augment tail graphs by predicting co-occurrence patterns that supplement their existing structures. Thus, the performance of tail graphs can be improved accordingly.

5.3.2 Imbalanced Topology Groups

Contrasting with the node topology imbalance discussed in Section 5.1.2, *imbalanced topology groups* [289] refer to the unequal distribution of topology groups, such as topological motifs, within individual classes in graph data. Take, for example, the Mutag dataset [289], where molecular graphs of the Mutagenic class contain two distinct topology groups: one associated with the motif NO_2 and the other with NH_2. In certain scenarios, the NO_2 group may substantially outnumber the NH_2 group, which results in the NO_2 motif having stronger associations with the class than the less frequent NH_2 motif. This imbalance, anchored in motif distribution, can lead to a paucity of training instances for minority groups. Consequently, classifiers may overfit to the majority topology (motif) groups, hampering effective learning and classification of instances in the underrepresented minority topology groups. To address this issue, TopoImb [289] dynamically updates the discovery of topology groups and assigns importance weights to under-represented instances during training, by incorporating a topology extractor and a training modulator. This approach improves the effectiveness of learning on minority topology groups and addresses the issue of over-fitting to majority groups.
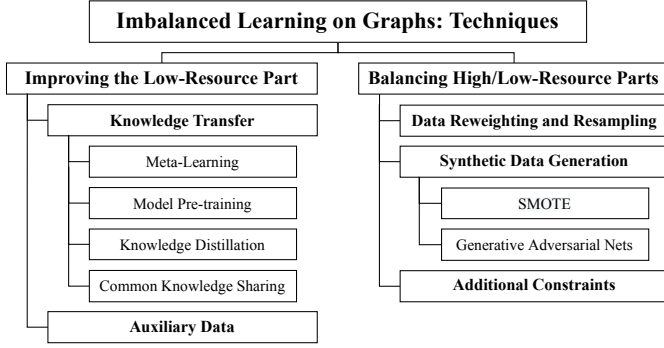
Fig. 4: Taxonomy of Techniques.

Summary In conclusion, the problem of graph-level structure imbalance is still a largely underexplored domain that calls for more in-depth study. The core challenge in this area is *the enhancement of the expressiveness of tail (sub)graphs or groups*. Predominantly, current methods tend to utilize knowledge transfer or reweighting techniques to address this issue. By drawing inspiration from the tasks previously discussed, a variety of strategies could potentially be adapted and applied to grapple with this intricate yet understudied task. Additionally, other types of graph-level imbalance issues that could potentially impact the performance of graph learning models may also require investigation, further emphasizing the need for continued research.

6 TECHNIQUES OF ILoGs

As detailed in Section 1, graph imbalance issues typically involve two primary components: the high- and low-resource parts. This leads to two main task categories: *improving the low-resource part* and *balancing both high and low-resource parts*. In this case, Section 6.1 will elaborate on tasks aimed at improving the performance of the low-resource part, such as few-shot node classification [58], [59], [189] and tail/cold-start node representation learning [27], [57]. In addition, Section 6.2 will explore tasks aiming to equalize performance across both parts, essential in scenarios like imbalanced classification tasks [25], [35], [37].

Our taxonomy, as summarized in Fig. 4, offers a well-organized and extensive overview of the existing literature across multiple tasks. This classification approach not only effectively differentiates various scenarios, but it also aids in identifying strategies bespoke to these specific settings, thereby facilitating effective management of graph imbalance issues, which we will discuss further in Section 6.3.

We also summarize the literature w.r.t. this taxonomy in Table 8. It is important to note that this categorization is not inflexible: some studies may incorporate techniques from another branch to mitigate imbalance issues in their own domains. For example, some studies [251] apply Mixup for synthetic data generation, enhancing the performance of the low-resource part. Others, such as [32], [63], use techniques like common knowledge sharing or knowledge distillation for knowledge transfer, seeking balanced performance across both parts. Note that, despite these cross-branch applications, our discussion primarily centers on major strategies for tackling imbalance.

6.1 Improving the Low-Resource Part

The primary goal of tasks designed to improve the low-resource part is to boost their performance, despite having less data compared to the high-resource part. They commonly exploit the high-resource part, which often holds a wealth of knowledge, as a resource to enrich the low-resource part. This category includes various graph tasks such as few-/zero-shot node/edge/graph classification, few-shot link prediction, or few-shot reasoning on KGs, and cold-start/tail node/entity/graph embedding, among others. To handle the imbalance issues, two main techniques are typically used: *knowledge transfer* and the incorporation of *auxiliary data*, in order to leverage existing resources to supplement low-resource parts.

6.1.1 Knowledge Transfer

In typical settings, there often exists a data part that is characterized by a wealth of source knowledge, such as the base classes within the framework of few-shot node classification, or high-degree nodes in tail node embedding situations. Knowledge transfer intends to harness this reservoir of knowledge residing within the data-rich source, and transfer it to the knowledge-deficient segment, thus bolstering its performance. To enable this, a suite of techniques has been engineered to facilitate efficient knowledge transfer between the source and target data, thereby enhancing the efficacy of imbalanced learning on graphs.

Meta-learning. Meta-learning, distinct from traditional machine learning methods, extracts versatile meta-knowledge from a series of base meta-tasks with a shared distribution. This knowledge is then applied to novel meta-tasks to enhance predictive performance. Many studies on ILoGs have utilized this approach to extract transferable meta-knowledge from high-resource parts (base meta-tasks) and apply it to low-resource parts (novel meta-tasks). Techniques employed often include optimization-based approaches (*e.g.*, MAML [206]) and metric-based approaches (*e.g.*, prototypical networks [207]), among others.

MAML, compatible with any optimization-based models, learns a base task prior that allows quick adaptation to the support set of a meta-task, facilitating rapid predictions on the query set. Its effectiveness supports various approaches addressing imbalanced learning on graphs [30], [58], [188], [189], [204], [220].

Conversely, within a meta-task, prototypical networks aim to identify the center of each class based on the support set, using distance-based metrics to classify novel instances in the query set. This simple but effective method is widely used in imbalanced learning on graphs to leverage transferable knowledge [26], [59], [191]–[193], [202]. Additionally, techniques like Matching Networks [282] and Siamese Networks [284] are also commonly used in this context.

Model pre-training. The paradigm of "pre-training, fine-tuning" is a widely accepted two-step model training strategy in machine learning. Initially, models are pre-trained on a large dataset using self-supervised learning, which uncovers the underlying knowledge embedded within the data. Following this, the pre-trained models are fine-tuned on a smaller, task-specific supervised dataset. This strategy

Techniques			Literature
Improving the low-resource part	Knowledge transfer	Meta-learning	Optimization-based
			[18], [31], [58], [76], [108], [188]–[190], [193], [203], [204], [209], [210], [220], [225]–[229], [237], [246], [255], [256], [258], [259], [273], [279]–[281], [283], [288]
			Metric-based
			[26], [33], [59], [191], [192], [194], [199], [200], [202], [221], [222], [224], [254], [257], [260]–[270], [274], [277]
		Model pre-training	GNN parameters transfer [195], contrastive learning [196], [197], prompting [198]
		Knowledge distillation	GNNs to MLPs [57], KG models to MLPs [278], Random knowledge distillation [73]
		Common knowledge sharing	Data sharing
			Super-classes [223]
			Model sharing
			[27], [34], [234], [235]
	Auxiliary data		Alignment data [245], auxiliary descriptions [216], [217], [277]
Balancing high/low-resource parts	Reweighting and resampling		Reweighting [61], [75], [120], [125], [147], [179], [240], [242], [244], [289], resampling [36], [69], [83], [131], [134], [136], [143]–[145], [145], [148], [165], [172], [241], [243]
	Synthetic data generation	SMOTE	SMOTE [25], Mixup [65]–[67]
		GAN	[37], [102], [110], [164], [180]
		Other methods	Predictive data generation [74], [135], label generation [77], [167], [170]
	Additional constraints		Condition relax constraints [139], [213]–[215], imbalance constraints [62], [112], class separation constraints [35], [72], [78]–[101], [103]–[107], [109], [111], [113]–[119], [121]–[124], [126]–[130], [132], [133], [137], [138], [140]–[144], [146]–[163], [166], [168], [169], [171], [173]–[179], [181]–[185], [290]

TABLE 8: Literature categorization of imbalanced learning on graphs w.r.t. the taxonomy of techniques.

benefits the fine-tuning stage by providing useful initializations for the models, particularly useful when dealing with limited labeled data. This strategy has demonstrated success across various tasks, setting benchmark results and establishing itself as a robust paradigm [291]–[293].

Within the sphere of ILoGs, a typical strategy involves pre-training models on the high-resource parts to distill the embedded knowledge, such as pre-training GNN models in alignment with the label information of the high-resource parts [195], [196]. Subsequently, these pre-trained models undergo fine-tuning with limited supervision from the low-resource parts, thereby facilitating the transfer of knowledge to enhance the performance of these low-resource parts.

Knowledge distillation. Knowledge distillation [252] is a technique wherein a smaller "student" model learns from a larger "teacher" model by mimicking its behavior. The teacher model offers soft targets to guide the student model, thereby transferring its knowledge to the student, aiming to boost efficiency while preserving performance [294], [295].

Regarding ILoGs, one approach involves constructing a comprehensive teacher model for the high-resource part to extract knowledge. To tackle data scarcity in the low-resource part, a student model is developed, guided by the teacher, facilitating training on limited data. This technique has shown noteworthy usage in situations such as cold-start node embedding [57], as detailed in Section 5.1.1. In summary, knowledge distillation is particularly beneficial in scenarios where distinct models are needed to decode patterns within high- and low-resource parts of the graph.

Common knowledge sharing. Common knowledge sharing forms the crux of knowledge transfer, establishing a conduit between high- and low-resource parts through shared elements. Some studies propose shared data [223] or shared sub-models [27], [34], [234], [235] as bridges for knowledge transfer. These shared components serve as critical resources for models associated with both high- and low-resource parts, facilitating further model development while maintaining a conversation via this shared bridge.

These methods are particularly beneficial in ILoGs. Strategies employing shared data construct overarching relationships for both high- and low-resource parts, like super-classes based on original classes [223]. Establishing this hierarchical structure enables knowledge transfer from high-resource (*e.g.*, majority classes) to low-resource parts (*e.g.*, minority classes) via shared relationships. In contrast, studies addressing structure imbalance commonly design a shared model for high- and low-resource parts, further incorporating model modulation to adapt the globally shared model to the unique characteristics of each part [27], [34], [234], [235]. The shared component, or globally shared model, retains common knowledge for both parts and serves as a bridge for knowledge transfer from high-resource (*e.g.*, high-degree nodes [27], [234], [235] or large-size graphs [34]) to low-resource parts (*e.g.*, low-degree nodes or small-size graphs). This approach is generally effective in scenarios where certain common attributes are shared by both parts, such as high-level features from the data-sharing perspective (*e.g.*, super-classes) or common knowledge which can be preserved by a shared model.

6.1.2 Auxiliary Data

Auxiliary data in machine learning refers to supplemental information that enhances the primary dataset, consequently improving model performance. For ILoGs, auxiliary data often underpins the learning process of low-resource parts by contributing additional knowledge [296], [297].

Auxiliary data in the realm of ILoGs takes various forms, enabling it to support multiple tasks. For zero-shot learning, text descriptions can offer class information for each class, facilitating classification in the absence of label support [216], [217], [277]. Some research centered on knowledge graph completion exploits auxiliary KGs to align with the target KG, providing extra information for KG completion concerning long-tailed entities [245] or zero-shot relations [277], [278]. As such, the use of auxiliary data can be contemplated in most scenarios of ILoGs where it is available.

6.2 Balancing the High- and Low-Resource Parts

In contrast to the objective of solely improving the low-resource part, balancing the high- and low-resource parts aims to enhance the performance of the low-resource part while minimizing potential performance degradation in the high-resource part. Typical tasks often involve imbalanced node/graph classification [25], [32], anomaly detection [36], [147], *etc.* To accomplish this, three primary techniques have been proposed: *data reweighting and resampling*, *synthetic data generation*, and the incorporation of *additional constraints*.

6.2.1 Data Reweighting and Resampling

Data resampling and reweighting techniques are commonly employed in imbalanced learning to address class imbalance issues within training data. Class imbalance occurs when there is an unequal distribution of classes, leading to a bias towards the majority class and reduced performance on minority classes. Both data resampling and reweighting aim to rebalance the class distribution, allowing the model to learn from minority class examples more effectively. Within the scope of ILoGs, data reweighting and resampling methods are prevalently employed to address class imbalance challenges [69], [242]. These techniques are applicable in numerous scenarios plagued by imbalance issues.

6.2.2 Synthetic Data Generation

Synthetic data generation techniques are widely employed in ILoGs to address the challenge of class imbalance by generating synthetic data for data rebalance. These techniques offer valuable solutions to achieve balanced performance in scenarios where there is a significant disparity between the high- and low-resource parts of the data. Specifically, two commonly used techniques for synthetic data generation are *SMOTE* (Synthetic Minority Over-sampling Technique) [16] and *GAN* (Generative Adversarial Networks) [60].

SMOTE. SMOTE [16] is a well-known approach utilized in imbalanced learning to combat the class imbalance problem within training data. It involves creating synthetic samples by interpolating between existing minority class instances. For each minority class instance, SMOTE selects one or more k-nearest neighbors from the same class and generates synthetic examples by linearly interpolating between the selected instance and its neighbors. This process effectively expands the minority class and introduces additional data points to improve its representation during training.

Furthermore, Mixup [68], a data augmentation technique initially proposed for training deep neural networks, can also be leveraged in imbalanced learning scenarios. Mixup extends the idea of interpolation by generating virtual training examples through linear combinations of instance pairs. As a special case of SMOTE, Mixup can also promot model generalization and reducing overfitting [68].

In the context of ILoGs, researchers have explored the adaptation of SMOTE and Mixup to address the class imbalance issue [25], [65], [66]. These methods are particularly beneficial when there is a need to bolster the low-resource segment with additional instances for data rebalance, for tasks like imbalanced node/graph classification.

GAN. Generative Adversarial Networks (GANs) [60] are a class of machine learning techniques composed of two
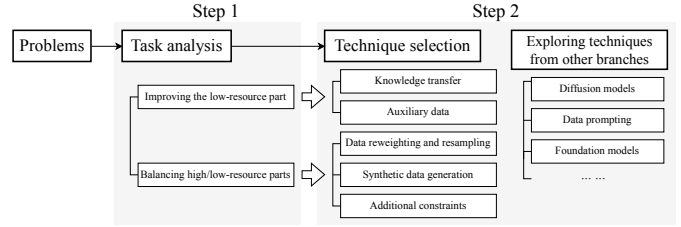

Fig. 5: Procedure of techniques selection.

modules, a generator and a discriminator, which are trained together in an adversarial manner. GANs are known for generating realistic and high-quality data, such as images, that closely resemble the original data distribution.

In the context of ILoGs, some research (*e.g.*, IMGAGN [37]) has explored the use of GANs to generate synthetic samples for minority classes, functioning similarly to SMOTE. These GAN-based techniques can help balance the data distribution and improve the model performance on minority classes in imbalanced graph learning scenarios.

6.2.3 Additional Constraints

Additional constraints can be effectively incorporated in ILoGs to bridge the gap between high- and low-resource parts, to allow knowledge transfer between them while adhering to predefined constraints, thereby enhancing performance. For instance, some studies [213]–[215] have attempted to relax the intra-class similarity requirement by allowing nodes with the same labels to reside on the same manifold in the embedding space. In particular, the selection and design of these constraints depend on the specific characteristics of the dataset and the learning tasks at hand.

6.3 Appropriate Techniques Selection

Selecting the proper technique for ILoGs is pivotal, following a specific paradigm, as shown in Fig. 5. The process comprises task analysis, technique selection, and further exploration of methods from related machine learning fields.

Firstly, *task analysis* involves understanding the data imbalance and its challenges. This step requires determining the data split into high- and low-resource parts and identifying the primary goal: improving the performance of the low-resource part, or balancing both parts. This analysis is the foundation for subsequent technique selection.

Secondly, *technique selection* involves considering various methods to address the identified imbalance issue. Each technique offers a unique perspective and capability to address the problem, making it a potential selection candidate.

For tasks focused on enhancing low-resource part performance, knowledge transfer and the use of auxiliary data are key considerations. Knowledge transfer is particularly beneficial when ample base data is available. Meta-learning is effective for meta-task-oriented tasks, and model pre-training is recommended when valuable information can be extracted from unlabeled base data. Knowledge distillation works well when different models operate on different data parts. Common knowledge sharing technique aids in capturing shared patterns (data or model) across both parts, thus promoting knowledge transfer.

For tasks centered on balancing the performance of both high- and low-resource parts, data reweighting and resampling, synthetic data generation, and additional constraints are essential considerations. Data reweighting and resampling can rectify imbalance issues and enhance data distribution. Synthetic data generation techniques, such as SMOTE or GAN, can generate synthetic instances for the low-resource part, thus also rebalancing the data distribution. Additionally, when specific conditions exist for model training based on data characteristics, additional constraints can guide the model training to address imbalance issues.

Finally, *exploring techniques from other branches* is a worthwhile endeavor. On one hand, although most techniques are categorized under the two branches—improving the low-resource part or balancing the high- and low-resource parts—the boundaries are not strict. Namely, techniques intended for one branch may be applicable to tasks in the other. For instance, in the context of improving low-resource part performance, synthetic data generation and additional task-specific constraints, originally meant for balancing both parts, can potentially bolster the data and performance. Conversely, techniques like knowledge transfer, can potentially foster balancing both parts by transferring knowledge from high-resource to low-resource parts. On the other hand, exploring novel techniques that are not commonly used in ILoGs can also be beneficial, *e.g.*, diffusion models [71], data prompting [291], foundation models [187], *etc.* These techniques may bring new perspectives and potential solutions for addressing imbalance issues on graphs.

7   OTHER RELATED LITERATURE

In addition to the primary scenarios of ILoGs, there are other contexts that fuse graph learning techniques and imbalance issues, such as graph fairness learning [298], cold-start recommendation [299], and imbalanced road networks [300]. In this section, we present a concise overview of the tasks closely related to our discussion, including fairness learning on graphs and imbalanced learning in recommendations.

7.1   Fairness Learning on Graphs

Fairness learning addresses the imbalance in input attributes linked to certain predictions due to historical prejudices in data collection. This imbalance, particularly crucial when dealing with sensitive attributes like gender, race, or education, can produce discrimination in model outcomes, fostering unfairness. Unlike other imbalanced learning forms, fairness learning aims to minimize prediction discrepancies between different subgroups, instead of typically focusing on performance on imbalanced data.

Methods to achieve fairness learning on graphs can be mainly categorized into two classes: adding constraints to model training and pre-processing input data. The first line of works adds a fairness constraint to the objective of the task [298], [301]–[313], promoting independence between output and sensitive attributes. The second stream of works resorts to modify input features to create bias-free input, either by removing the influence of sensitive attributes [314]–[326], or by rendering the model invariant to these attributes [327]–[330]. Apart for these two main categories, some other

techniques are also deployed in this setting, such as post-processing output embeddings [331], partitioning training set [332], [333], and considering post-hoc explanation of bias [334]. We refer interesting readers to a comprehensive survey [48] for more details.

7.2   Imbalanced learning in Recommendations

Imbalanced learning in recommendations primarily addresses cold-start recommendations within user-item rating graphs, a bipartite graph [299], [335]. Distinct from ILoGs, cold-start recommendation usually grapples with insufficient user-item interactions. It manifests when making recommendations for new users (user cold-start) or new items (item cold-start) that lack adequate interaction data to guide traditional collaborative filtering methods. To mitigate the cold-start recommendation problem, numerous techniques have been employed, including meta-learning [336], [337], pre-training [338], and knowledge distillation [339], *etc.* For an in-depth understanding of this challenge, we encourage readers to consult the related survey [50].

8   FUTURE DIRECTIONS

In this section, we explore prospective avenues for advancing ILoGs, building on our established taxonomies of problems (Section 8.1) and techniques (Section 8.2). Our goal is to stimulate research that tackles the challenges inherent to imbalanced graph data and ultimately improves the performance in various scenarios.

8.1   Future Directions of Problems

The characteristics of graph data lead to various imbalance challenges in graph learning. Despite substantial research in this area, some facets remain underexplored and deserve further attention. Two pivotal dimensions—class and structure imbalance—provide a framework for future research when examined within our problem taxonomy. Furthermore, the varied nature of graph imbalance necessitates diverse imbalance ratios to quantify the degree of imbalance, offering a potential cornerstone for ensuing investigations.

Class imbalance. As indicated in Section 4, much of the current research in ILoGs centers on node-level imbalance, leaving edge- and graph-level tasks less explored.

Firstly, edge-level imbalance tasks, particularly for generic graphs, need more focus. Edge behaviors can vary across or even within a single graph, creating unique challenges [31]. Investigating imbalanced edge tasks like few-shot link prediction could significantly enhance the practicality of graph learning in real-world applications.

Secondly, graph-level imbalance tasks have received limited investigation and warrant further exploration. Beyond anomaly detection, there is a lack of research on more generic imbalanced graph classification [32], which is crucial in real-world applications. Moreover, considering the laborious nature of labeling in many molecular-related tasks, graph classification without supervision (*e.g.*, zero-shot graph classification) becomes increasingly important, which might be potentially assisted by text information.

Structure imbalance. In the realm of structure imbalance, the majority of recent work has primarily focused on node-level imbalance. However, the intricate connectivity patterns present in graphs can offer various perspectives on structure imbalance across nodes.

For instance, beyond the well-researched degree imbalance among nodes, another type of imbalance relating to generalized node degree [340], which captures the abundance of contextual structures of a node, could also trigger performance disparities among nodes. Switching to graph-level tasks, imbalance may also stem from graph size variations. Differences in graph sizes may introduce bias and impact outcomes [34], a factor that remains under-exploited and calls for further investigation.

8.2 Future Directions of Techniques

Two major directions can be considered in order to investigate future directions of techniques for ILoGs.

Cross-branch technique exploration. As discussed in Section 6.3, the techniques of one branch can address the imbalance issues of another, emphasizing the potential for cross-branch technique exploration. Thus, researchers are urged to consider this method to address imbalance issues.

Novel technique exploration. As real-world applications spur artificial intelligence innovations, exploring new techniques becomes essential. For example, in the context of synthetic data generation for balancing high/low-resource parts, diffusion models [71] offer an alternative to GANs, given their superior generation capabilities in some contexts [71], [341], [342]. Additionally, foundation models [187] like GPT models [343], known for their potential in various domains, could potentially offer additional benefits in ILoGs where the text information is available.

8.3 Other Directions

ILoGs is an area where current studies often feel fragmented and disconnected. One of the major challenges is the absence of unified benchmark datasets and leaderboards for each research task. Such tools would not only facilitate a more direct evaluation of individual studies but also encourage healthy competition among researchers. Furthermore, there is a valuable opportunity to consolidate and review existing work on each specific task within this domain. This would produce comprehensive survey reports, offering a holistic view and serving as invaluable resources for both newcomers and seasoned researchers in the field.

9 CONCLUSIONS

In this survey, we offer a comprehensive review of the literature on imbalanced learning on graphs. To provide a solid foundation, we begin by presenting a formal definition of the concept, followed by an introduction to basic terms in the field. Significantly, we develop two comprehensive taxonomies of ILoGs, categorizing existing studies based on their problems and techniques. Our problem taxonomy identifies class and structure imbalance, further detailing research tasks within each branch concerning graph-related elements, namely nodes, edges, and graphs. Meanwhile, the

technique taxonomy classifies the literature based on the type of imbalance and corresponding strategies deployed to rectify these imbalance on graphs. Finally, we outline promising future directions for both problems and techniques in ILoGs, in order to spark further innovation and exploration in this increasingly critical domain.

REFERENCES

[1] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford infolab, Tech. Rep., 1999.

[2] G. Jeh and J. Widom, "Simrank: a measure of structural-context similarity," in *KDD*, 2002, pp. 538–543.

[3] L. Backstrom and J. Leskovec, "Supervised random walks: predicting and recommending links in social networks," in *WSDM*, 2011, pp. 635–644.

[4] H. Cai, V. W. Zheng, and K. C.-C. Chang, "A comprehensive survey of graph embedding: Problems, techniques, and applications," *TKDE*, vol. 30, no. 9, pp. 1616–1637, 2018.

[5] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *TNNLS*, vol. 32, no. 1, pp. 4–24, 2020.

[6] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *KDD*, 2014, pp. 701–710.

[7] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding," in *WWW*, 2015, pp. 1067–1077.

[8] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *KDD*, 2016, pp. 855–864.

[9] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *ICLR*, 2017.

[10] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *NeurIPS*, 2017, pp. 1024–1034.

[11] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *ICLR*, 2018.

[12] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" in *ICLR*, 2019.

[13] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural networks*, vol. 106, pp. 249–259, 2018.

[14] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *CVPR*, 2019, pp. 9268–9277.

[15] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis, "Decoupling representation and classifier for long-tailed recognition," in *ICLR*, 2020.

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *JAIR*, vol. 16, pp. 321–357, 2002.

[17] Y. Zhang, B. Kang, B. Hooi, S. Yan, and J. Feng, "Deep long-tailed learning: A survey," *TPAMI*, 2023.

[18] Z. Liu, W. Zhang, Y. Fang, X. Zhang, and S. C. Hoi, "Towards locality-aware meta-learning of tail node embeddings on networks," in *CIKM*, 2020, pp. 975–984.

[19] H. He and E. A. Garcia, "Learning from imbalanced data," *TKDE*, vol. 21, no. 9, pp. 1263–1284, 2009.

[20] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221–232, 2016.

[21] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert systems with applications*, vol. 73, pp. 220–239, 2017.

[22] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, pp. 1–54, 2019.

[23] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *CVPR*, 2016, pp. 5375–5384.

[24] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, and J. Li, "Dice loss for data-imbalanced nlp tasks," in *ACL*, 2020, pp. 465–476.

[25] T. Zhao, X. Zhang, and S. Wang, "Graphsmote: Imbalanced node classification on graphs with graph neural networks," in *WSDM*, 2021, pp. 833–841.

[26] H. Yao, C. Zhang, Y. Wei, M. Jiang, S. Wang, J. Huang, N. Chawla, and Z. Li, "Graph few-shot learning via knowledge transfer," in *AAAI*, vol. 34, no. 04, 2020, pp. 6656–6663.

[27] Z. Liu, T.-K. Nguyen, and Y. Fang, "Tail-gnn: Tail-node graph neural networks," in *KDD*, 2021, pp. 1109–1119.

[28] X. Ma, J. Wu, S. Xue, J. Yang, C. Zhou, Q. Z. Sheng, H. Xiong, and L. Akoglu, "A comprehensive survey on graph anomaly detection with deep learning," *TKDE*, 2021.

[29] D. Mandal, S. Medya, B. Uzzi, and C. Aggarwal, "Metalearning with graph neural networks: Methods and applications," *ACM SIGKDD Explorations Newsletter*, vol. 23, no. 2, pp. 13–22, 2022.

[30] C. Zhang, K. Ding, J. Li, X. Zhang, Y. Ye, N. V. Chawla, and H. Liu, "Few-shot learning on graphs," in *IJCAI*, 2022, pp. 5662–5669.

[31] X. Zhu, P. Luo, Z. Zhao, T. Xu, A. Lizhiyu, Y. Yu, X. Li, and E. Chen, "Few-shot link prediction for event-based social networks via meta-learning," in *DASFAA*, 2023, pp. 31–41.

[32] Y. Wang, Y. Zhao, N. Shah, and T. Derr, "Imbalanced graph classification via graph-of-graph neural networks," in *CIKM*, 2022, pp. 2067–2076.

[33] W. Xiong, M. Yu, S. Chang, X. Guo, and W. Y. Wang, "One-shot relational learning for knowledge graphs," in *EMNLP*, 2018, pp. 1980–1990.

[34] Z. Liu, Q. Mao, C. Liu, Y. Fang, and J. Sun, "On size-oriented long-tailed graph classification of graph neural networks," in *WWW*, 2022, pp. 1506–1516.

[35] M. Shi, Y. Tang, X. Zhu, D. Wilson, and J. Liu, "Multi-class imbalanced graph convolutional network learning," in *IJCAI*, 2020, pp. 2879–2885.

[36] Y. Liu, X. Ao, Z. Qin, J. Chi, J. Feng, H. Yang, and Q. He, "Pick and choose: a gnn-based imbalanced learning approach for fraud detection," in *WWW*, 2021, pp. 3168–3177.

[37] L. Qu, H. Zhu, R. Zheng, Y. Shi, and H. Yin, "Imgagn: Imbalanced network embedding via generative adversarial graph networks," in *KDD*, 2021, pp. 1390–1398.

[38] A. Fernández, S. Garcia, F. Herrera, and N. V. Chawla, "Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary," *JAIR*, vol. 61, pp. 863–905, 2018.

[39] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *CSUR*, vol. 41, no. 3, pp. 1–58, 2009.

[40] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," *arXiv preprint arXiv:1901.03407*, 2019.

[41] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *CSUR*, vol. 54, no. 2, pp. 1–38, 2021.

[42] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *CSUR*, vol. 53, no. 3, pp. 1–34, 2020.

[43] Y. Song, T. Wang, S. K. Mondal, and J. P. Sahoo, "A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities," *CSUR*, 2023.

[44] Y. Ma, Y. Tian, N. Moniz, and N. V. Chawla, "Class-imbalanced learning on graphs: A survey," *arXiv preprint arXiv:2304.04300*, 2023.

[45] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: methods, systems and tools," *COMST*, vol. 16, no. 1, pp. 303–336, 2013.

[46] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," *DMKD*, vol. 29, no. 3, pp. 626–688, 2015.

[47] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *JNCA*, vol. 60, pp. 19–31, 2016.

[48] Y. Dong, J. Ma, S. Wang, C. Chen, and J. Li, "Fairness in graph mining: A survey," *TKDE*, no. 01, pp. 1–22, 2023.

[49] C. Yang, Y. Xiao, Y. Zhang, Y. Sun, and J. Han, "Heterogeneous network representation learning: A unified framework with survey and benchmark," *TKDE*, vol. 34, no. 10, pp. 4854–4873, 2020.

[50] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *CSUR*, vol. 52, no. 1, pp. 1–38, 2019.

[51] S. Ji, S. Pan, E. Cambria, P. Marttinen, and S. Y. Philip, "A survey on knowledge graphs: Representation, acquisition, and applications," *TNNLS*, vol. 33, no. 2, pp. 494–514, 2021.

[52] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," *NeurIPS*, vol. 31, pp. 4805–4815, 2018.

[53] Z. Guo, B. Nan, Y. Tian, O. Wiest, C. Zhang, and N. V. Chawla, "Graph-based molecular representation learning," *arXiv preprint arXiv:2207.04869*, 2022.

[54] X. Yu, Z. Liu, Y. Fang, and X. Zhang, "Learning to count isomorphisms with graph neural networks," in *AAAI*, 2023.

[55] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, "Convolutional networks on graphs for learning molecular fingerprints," *NeurIPS*, vol. 28, pp. 2224–2232, 2015.

[56] H. Gao and S. Ji, "Graph u-nets," in *ICML*, 2019, pp. 2083–2092.

[57] W. Zheng, E. W. Huang, N. Rao, S. Katariya, Z. Wang, and K. Subbian, "Cold brew: Distilling graph node representations with incomplete or missing neighborhoods," in *ICLR*, 2022.

[58] F. Zhou, C. Cao, K. Zhang, G. Trajcevski, T. Zhong, and J. Geng, "Meta-gnn: On few-shot node classification in graph meta-learning," in *CIKM*, 2019, pp. 2357–2360.

[59] K. Ding, J. Wang, J. Li, K. Shu, C. Liu, and H. Liu, "Graph prototypical networks for few-shot learning on attributed networks," in *CIKM*, 2020, pp. 295–304.

[60] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS*, vol. 27, 2014.

[61] Y. Wang, C. Aggarwal, and T. Derr, "Distance-wise prototypical graph neural network in node imbalance classification," *arXiv preprint arXiv:2110.12035*, 2021.

[62] J. Song, J. Park, and E. Yang, "Tam: Topology-aware margin loss for class-imbalanced node classification," in *ICML*, 2022, pp. 20369–20383.

[63] S. Yun, K. Kim, K. Yoon, and C. Park, "Lte4g: Long-tail experts for graph neural networks," in *CIKM*, 2022, pp. 2434–2443.

[64] C. Drummond, R. C. Holte *et al.*, "C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling," in *Workshop on learning from imbalanced datasets II*, vol. 11, 2003, pp. 1–8.

[65] L. Wu, H. Lin, Z. Gao, C. Tan, S. Li *et al.*, "Graphmixup: Improving class-imbalanced node classification by reinforcement mixup and self-supervised context prediction," *ECML-PKDD*, pp. 519–535, 2022.

[66] J. Park, J. Song, and E. Yang, "Graphens: Neighbor-aware ego network synthesis for class-imbalanced node classification," in *ICLR*, 2022.

[67] J. Liu, M. He, G. Wang, N. Q. V. Hung, X. Shang, and H. Yin, "Imbalanced node classification beyond homophilic assumption," in *IJCAI*, 2023.

[68] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *ICLR*, 2018.

[69] L. Cui, X. Tang, S. Katariya, N. Rao, P. Agrawal, K. Subbian, and D. Lee, "Allie: Active learning on large-scale imbalanced graphs," in *WWW*, 2022, pp. 690–698.

[70] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *JAIR*, vol. 4, pp. 237–285, 1996.

[71] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *CVPR*, 2022, pp. 10684–10695.

[72] C. Qiu, M. Kloft, S. Mandt, and M. Rudolph, "Raising the bar in graph-level anomaly detection," *IJCAI*, pp. 2196–2203, 2022.

[73] R. Ma, G. Pang, L. Chen, and A. van den Hengel, "Deep graph-level anomaly detection by glocal knowledge distillation," in *WSDM*, 2022, pp. 704–714.

[74] T. Zhao, B. Ni, W. Yu, Z. Guo, N. Shah, and M. Jiang, "Action sequence augmentation for early graph-based anomaly detection," in *CIKM*, 2021, pp. 2668–2678.

[75] G. Zhang, J. Wu, J. Yang, A. Beheshti, S. Xue, C. Zhou, and Q. Z. Sheng, "Fraudre: Fraud detection dual-resistant to graph inconsistency and imbalance," in *ICDM*, 2021, pp. 867–876.

[76] Y. Qian, Y. Zhang, Y. Ye, C. Zhang *et al.*, "Distilling meta knowledge on heterogeneous graph for illicit drug trafficker detection on social media," *NeurIPS*, vol. 34, pp. 26911–26923, 2021.

[77] J. Mathew, M. Negi, R. Vijjali, and J. Sathyanarayana, "Defraudnet: An end-to-end weak supervision framework to detect fraud in online food delivery," in *ECML-PKDD*, 2021, pp. 85–99.

[78] H. Wang and C. Qiao, "A nodes' evolution diversity inspired method to detect anomalies in dynamic social networks," *TKDE*, vol. 32, no. 10, pp. 1868–1880, 2019.

[79] V. Miz, B. Ricaud, K. Benzi, and P. Vandergheynst, "Anomaly detection in the dynamics of web and social networks using associative memory," in *WWW*, 2019, pp. 1290–1299.

[80] H. Lin, G. Liu, J. Wu, Y. Zuo, X. Wan, and H. Li, "Fraud detection in dynamic interaction network," *TKDE*, vol. 32, no. 10, pp. 1936–1950, 2019.

[81] R. Hu, C. C. Aggarwal, S. Ma, and J. Huai, "An embedding approach to anomaly detection," in *ICDE*, 2016, pp. 385–396.

[82] J. Li, H. Dani, X. Hu, and H. Liu, "Radar: Residual analysis for anomaly detection in attributed networks." in *IJCAI*, vol. 17, 2017, pp. 2152–2158.

[83] B. Wang, N. Z. Gong, and H. Fu, "Gang: Detecting fraudulent users in online social networks via guilt-by-association on directed graphs," in *ICDM*, 2017, pp. 465–474.

[84] A. Bojchevski and S. Günnemann, "Bayesian robust attributed graph clustering: Joint learning of partial anomalies and group structure," in *AAAI*, vol. 32, no. 1, 2018.

[85] Z. Peng, M. Luo, J. Li, H. Liu, and Q. Zheng, "Anomalous: A joint modeling approach for anomaly detection on attributed networks." in *IJCAI*, 2018, pp. 3513–3519.

[86] Y. Yang, Y. Xu, Y. Sun, Y. Dong, F. Wu, and Y. Zhuang, "Mining fraudsters and fraudulent strategies in large-scale mobile social networks," *TKDE*, vol. 33, no. 1, pp. 169–179, 2019.

[87] K. Ding, J. Li, and H. Liu, "Interactive anomaly detection on attributed networks," in *WSDM*, 2019, pp. 357–365.

[88] L. Gutiérrez-Gómez, A. Bovet, and J.-C. Delvenne, "Multi-scale anomaly detection on attributed networks," in *AAAI*, vol. 34, no. 01, 2020, pp. 678–685.

[89] M. Zhu and H. Zhu, "Mixedad: A scalable algorithm for detecting mixed anomalies in attributed graphs," in *AAAI*, vol. 34, no. 01, 2020, pp. 1274–1281.

[90] C. C. Aggarwal, Y. Li, and P. S. Yu, "Signature-based anomaly detection in networks," in *SDM*, 2021, pp. 109–117.

[91] H. T. Nguyen, P. J. Liang, and L. Akoglu, "Detecting anomalous graphs in labeled multi-graph databases," *TKDD*, vol. 17, no. 2, pp. 1–25, 2023.

[92] S. Yuan, X. Wu, J. Li, and A. Lu, "Spectrum-based deep neural networks for fraud detection," in *CIKM*, 2017, pp. 2419–2422.

[93] J. Liang, P. Jacobs, J. Sun, and S. Parthasarathy, "Semi-supervised embedding in attributed networks with outliers," in *SDM*, 2018, pp. 153–161.

[94] D. Zhou, J. He, H. Yang, and W. Fan, "Sparc: Self-paced network representation for few-shot rare category characterization," in *KDD*, 2018, pp. 2807–2816.

[95] S. Bandyopadhyay, N. Lokesh, and M. N. Murty, "Outlier aware network embedding for attributed networks," in *AAAI*, vol. 33, no. 01, 2019, pp. 12–19.

[96] S. Bandyopadhyay, S. V. Vivek, and M. N. Murty, "Integrating network embedding and community outlier detection via multiclass graph description," *ECAI*, pp. 976–983, 2020.

[97] I. Ahmed, T. Galoppo, X. Hu, and Y. Ding, "Graph regularized autoencoder and its application in unsupervised anomaly detection," *TPAMI*, vol. 44, no. 8, pp. 4110–4124, 2021.

[98] Y. Li, X. Huang, J. Li, M. Du, and N. Zou, "Specae: Spectral autoencoder for anomaly detection in attributed networks," in *CIKM*, 2019, pp. 2233–2236.

[99] K. Ding, J. Li, R. Bhanushali, and H. Liu, "Deep anomaly detection on attributed networks," in *SDM*, 2019, pp. 594–602.

[100] T. Zhao, C. Deng, K. Yu, T. Jiang, D. Wang, and M. Jiang, "Error-bounded graph anomaly loss for gnns," in *CIKM*, 2020, pp. 1873–1882.

[101] Z. Peng, M. Luo, J. Li, L. Xue, and Q. Zheng, "A deep multi-view framework for anomaly detection on attributed networks," *TKDE*, vol. 34, no. 6, pp. 2539–2552, 2020.

[102] Z. Chen, B. Liu, M. Wang, P. Dai, J. Lv, and L. Bo, "Generative adversarial attributed network anomaly detection," in *CIKM*, 2020, pp. 1989–1992.

[103] D. Cheng, X. Wang, Y. Zhang, and L. Zhang, "Graph neural network for fraud detection via spatial-temporal attention," *TKDE*, vol. 34, no. 8, pp. 3800–3813, 2020.

[104] K. Ding, K. Shu, X. Shan, J. Li, and H. Liu, "Cross-domain graph anomaly detection," *TNNLS*, vol. 33, no. 6, pp. 2406–2415, 2021.

[105] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *AAAI*, vol. 35, no. 5, 2021, pp. 4027–4035.

[106] U. Desai, S. Bandyopadhyay, and S. Tamilselvam, "Graph neural network to dilute outliers for refactoring monolith application," in *AAAI*, vol. 35, no. 1, 2021, pp. 72–80.

[107] M. Jin, Y. Liu, Y. Zheng, L. Chi, Y.-F. Li, and S. Pan, "Anemone: Graph anomaly detection with multi-scale contrastive learning," in *CIKM*, 2021, pp. 3122–3126.

[108] K. Ding, Q. Zhou, H. Tong, and H. Liu, "Few-shot network anomaly detection via cross-network meta-learning," in *WWW*, 2021, pp. 2448–2456.

[109] S. Zhou, Q. Tan, Z. Xu, X. Huang, and F.-l. Chung, "Subtractive aggregation for attributed network anomaly detection," in *CIKM*, 2021, pp. 3672–3676.

[110] K. Ding, J. Li, N. Agarwal, and H. Liu, "Inductive anomaly detection on attributed networks," in *IJCAI*, 2021, pp. 1288–1294.

[111] Y. Liu, Z. Li, S. Pan, C. Gong, C. Zhou, and G. Karypis, "Anomaly detection on attributed networks via contrastive self-supervised learning," *TNNLS*, vol. 33, no. 6, pp. 2378–2392, 2021.

[112] T. Zhao, T. Jiang, N. Shah, and M. Jiang, "A synergistic approach for graph anomaly detection with pattern mining and feature learning," *TNNLS*, vol. 33, no. 6, pp. 2393–2405, 2021.

[113] W. Chen, L. Tian, B. Chen, L. Dai, Z. Duan, and M. Zhou, "Deep variational graph convolutional recurrent network for multivariate time series anomaly detection," in *ICML*, 2022, pp. 3621–3633.

[114] P. Qi, D. Li, and S.-K. Ng, "Mad-sgcn: Multivariate anomaly detection with self-learning graph convolutional networks," in *ICDE*, 2022, pp. 1232–1244.

[115] X. Chen, Q. Qiu, C. Li, and K. Xie, "Graphad: A graph neural network for entity-wise multivariate time-series anomaly detection," in *SIGIR*, 2022, pp. 2297–2302.

[116] J. Zhang, S. Wang, and S. Chen, "Reconstruction enhanced multiview contrastive learning for anomaly detection on attributed networks," *IJCAI*, pp. 2376–2382, 2022.

[117] X. Luo, J. Wu, A. Beheshti, J. Yang, X. Zhang, Y. Wang, and S. Xue, "Comga: Community-aware attributed graph anomaly detection," in *WSDM*, 2022, pp. 657–665.

[118] A. Goodge, B. Hooi, S.-K. Ng, and W. S. Ng, "Lunar: Unifying local outlier detection methods via graph neural networks," in *AAAI*, vol. 36, no. 6, 2022, pp. 6737–6745.

[119] S. Zhou, X. Huang, N. Liu, Q. Tan, and F.-L. Chung, "Unseen anomaly detection on networks via multi-hypersphere learning," in *SDM*, 2022, pp. 262–270.

[120] F. Liu, X. Ma, J. Wu, J. Yang, S. Xue, A. Beheshti, C. Zhou, H. Peng, Q. Z. Sheng, and C. C. Aggarwal, "Dagad: Data augmentation for graph anomaly detection," *ICDM*, pp. 259–268, 2022.

[121] T. Huang, Y. Pei, V. Menkovski, and M. Pechenizkiy, "Hop-count based self-supervised anomaly detection on attributed networks," in *ECML-PKDD*, 2022, pp. 225–241.

[122] Z. Zhuang, K. M. Ting, G. Pang, and S. Song, "Subgraph centralization: A necessary step for graph anomaly detection," *SDM*, pp. 703–711, 2023.

[123] Y. Gao, X. Wang, X. He, Z. Liu, H. Feng, and Y. Zhang, "Addressing heterophily in graph anomaly detection: A perspective of graph spectrum," in *WWW*, 2023, pp. 1528–1538.

[124] Y. Huang, L. Wang, F. Zhang, and X. Lin, "Unsupervised graph outlier detection: Problem revisit, new insight, and superior method," *ICDE*, pp. 2556–2569, 2023.

[125] J. J.-C. Ying, J. Zhang, C.-W. Huang, K.-T. Chen, and V. S. Tseng, "Pfraudetector: a parallelized graph mining approach for efficient fraudulent phone call detection," in *ICPADS*, 2016, pp. 1059–1066.

[126] N. Liu, X. Huang, and X. Hu, "Accelerated local anomaly detection via resolving attributed networks." in *IJCAI*, 2017, pp. 2337–2343.

[127] B. Hooi, K. Shin, H. A. Song, A. Beutel, N. Shah, and C. Faloutsos, "Graph-based fraud detection in the face of camouflage," *TKDD*, vol. 11, no. 4, pp. 1–26, 2017.

[128] S. Zhang, D. Zhou, M. Y. Yildirim, S. Alcorn, J. He, H. Davulcu, and H. Tong, "Hidden: hierarchical dense subgraph detection with application to financial fraud detection," in *SDM*, 2017, pp. 570–578.

[129] B. Cao, M. Mao, S. Viidu, and S. Y. Philip, "Hitfraud: a broad learning approach for collective fraud detection in heterogeneous information networks," in *ICDM*, 2017, pp. 769–774.

[130] V. Ranjbar, M. Salehi, P. Jandaghi, and M. Jalili, "Qanet: Tensor decomposition approach for query-based anomaly detection in heterogeneous information networks," *TKDE*, vol. 31, no. 11, pp. 2178–2189, 2018.

[131] Y. Ren, H. Zhu, J. Zhang, P. Dai, and L. Bo, "Ensemfdet: An ensemble approach to fraud detection based on bipartite graph," in *ICDE*, 2021, pp. 2039–2044.

[132] Y. Fan, S. Hou, Y. Zhang, Y. Ye, and M. Abdulhayoglu, "Gotchasly malware! scorpion a metagraph2vec based malware detection system," in *KDD*, 2018, pp. 253–262.

[133] M. Yoon, "Graph fraud detection based on accessibility score distributions," in *ECML-PKDD*, 2021, pp. 483–498.

[134] Y. Dou, Z. Liu, L. Sun, Y. Deng, H. Peng, and P. S. Yu, "Enhancing graph neural network-based fraud detectors against camouflaged fraudsters," in *CIKM*, 2020, pp. 315–324.

[135] Y.-N. Zhu, X. Luo, Y.-F. Li, B. Bu, K. Zhou, W. Zhang, and M. Lu, "Heterogeneous mini-graph neural network and its application to fraud invitation detection," in *ICDM*, 2020, pp. 891–899.

[136] Z. Liu, Y. Dou, P. S. Yu, Y. Deng, and H. Peng, "Alleviating the inconsistency problem of applying graph neural network to fraud detection," in *SIGIR*, 2020, pp. 1569–1572.

[137] S. Zhang, H. Yin, T. Chen, Q. V. N. Hung, Z. Huang, and L. Cui, "Gcn-based user representation learning for unifying robust recommendation and fraudster detection," in *SIGIR*, 2020, pp. 689–698.

[138] S. Hu, X. Zhang, J. Zhou, S. Ji, J. Yuan, Z. Li, Z. Wang, Q. Chen, Q. He, and L. Fang, "Turbo: Fraud detection in deposit-free leasing service via real-time behavior network mining," in *ICDE*, 2021, pp. 2583–2594.

[139] B. Xu, H. Shen, B. Sun, R. An, Q. Cao, and X. Cheng, "Towards consumer loan fraud detection: Graph neural networks with role-constrained conditional random field," in *AAAI*, vol. 35, no. 5, 2021, pp. 4537–4545.

[140] L. Wang, P. Li, K. Xiong, J. Zhao, and R. Lin, "Modeling heterogeneous graph network on fraud detection: a community-based framework with attention mechanism," in *CIKM*, 2021, pp. 1959–1968.

[141] Y. Wang, J. Zhang, S. Guo, H. Yin, C. Li, and H. Chen, "Decoupling representation learning and classification for gnn-based anomaly detection," in *SIGIR*, 2021, pp. 1239–1248.

[142] C. Liu, L. Sun, X. Ao, J. Feng, Q. He, and H. Yang, "Intention-aware heterogeneous graph attention networks for fraud transactions detection," in *KDD*, 2021, pp. 3280–3288.

[143] L. Dong, Y. Liu, X. Ao, J. Chi, J. Feng, H. Yang, and Q. He, "Bi-level selection via meta gradient for graph-based fraud detection," in *DASFAA*, 2022, pp. 387–394.

[144] M. Huang, Y. Liu, X. Ao, K. Li, J. Chi, J. Feng, H. Yang, and Q. He, "Auc-oriented graph neural network for fraud detection," in *WWW*, 2022, pp. 1311–1321.

[145] Z. Li, D. Chen, Q. Liu, and S. Wu, "The devil is in the conflict: Disentangled information graph neural networks for fraud detection," *ICDM*, pp. 1059–1064, 2022.

[146] Z. Qin, Y. Liu, Q. He, and X. Ao, "Explainable graph-based fraud detection via neural meta-graph search," in *CIKM*, 2022, pp. 4414–4418.

[147] J. Tang, J. Li, Z. Gao, and J. Li, "Rethinking graph neural networks for anomaly detection," in *ICML*, 2022, pp. 21 076–21 089.

[148] F. Shi, Y. Cao, Y. Shang, Y. Zhou, C. Zhou, and J. Wu, "H2-fdetector: a gnn-based fraud detector with homophilic and heterophilic connections," in *WWW*, 2022, pp. 1486–1494.

[149] Y. Gao, X. Wang, X. He, Z. Liu, H. Feng, and Y. Zhang, "Alleviating structural distribution shift in graph anomaly detection," in *WSDM*, 2023, pp. 357–365.

[150] L. Rashidi, A. Kan, J. Bailey, J. Chan, C. Leckie, W. Liu, S. Rajasegarar, and K. Ramamohanarao, "Node re-ordering as a means of anomaly detection in time-evolving graphs," in *ECML-PKDD*, 2016, pp. 162–178.

[151] S. Liu, B. Hooi, and C. Faloutsos, "Holoscope: Topology-and-spike aware fraud detection," in *CIKM*, 2017, pp. 1539–1548.

[152] T. L. Fond, J. Neville, and B. Gallagher, "Designing size consistent statistics for accurate anomaly detection in dynamic networks," *TKDD*, vol. 12, no. 4, pp. 1–49, 2018.

[153] S. Liu, B. Hooi, and C. Faloutsos, "A contrast metric for fraud detection in rich graphs," *TKDE*, vol. 31, no. 12, pp. 2235–2248, 2018.

[154] V. Sharma, R. Kumar, W.-H. Cheng, M. Atiquzzaman, K. Srinivasan, and A. Y. Zomaya, "Nhad: Neuro-fuzzy based horizontal anomaly detection in online social networks," *TKDE*, vol. 30, no. 11, pp. 2171–2184, 2018.

[155] M. Yoon, B. Hooi, K. Shin, and C. Faloutsos, "Fast and accurate anomaly detection in dynamic graphs with a two-pronged approach," in *KDD*, 2019, pp. 647–657.

[156] S. Bhatia, B. Hooi, M. Yoon, K. Shin, and C. Faloutsos, "Midas: Microcluster-based detector of anomalies in edge streams," in *AAAI*, vol. 34, no. 04, 2020, pp. 3242–3249.

[157] X. Guo, B. Zhou, and S. Skiena, "Subset node anomaly tracking over large dynamic graphs," in *KDD*, 2022, pp. 475–485.

[158] X. Teng, Y.-R. Lin, and X. Wen, "Anomaly detection in dynamic networks using multi-view time-series hypersphere learning," in *CIKM*, 2017, pp. 827–836.

[159] W. Yu, W. Cheng, C. C. Aggarwal, K. Zhang, H. Chen, and W. Wang, "Netwalk: A flexible deep embedding approach for anomaly detection in dynamic networks," in *KDD*, 2018, pp. 2672–2681.

[160] X. Teng, M. Yan, A. M. Ertugrul, and Y.-R. Lin, "Deep into hypersphere: Robust and unsupervised anomaly discovery in dynamic networks," in *IJCAI*, 2018, pp. 2724–2730.

[161] Z. Chen and A. Sun, "Anomaly detection on dynamic bipartite graph with burstiness," in *ICDM*, 2020, pp. 966–971.

[162] D. Ofori-Boateng, I. S. Dominguez, C. Akcora, M. Kantarcioglu, and Y. R. Gel, "Topological anomaly detection in dynamic multi-layer blockchain networks," in *ECML-PKDD*, 2021, pp. 788–804.

[163] A. Z. Wang, R. Ying, P. Li, N. Rao, K. Subbian, and J. Leskovec, "Bipartite dynamic representations for abuse detection," in *KDD*, 2021, pp. 3638–3648.

[164] L. Deng, D. Lian, Z. Huang, and E. Chen, "Graph convolutional adversarial networks for spatiotemporal anomaly detection," *TNNLS*, vol. 33, no. 6, pp. 2416–2428, 2022.

[165] M. Lu, Z. Han, S. X. Rao, Z. Zhang, Y. Zhao, Y. Shan, R. Raghunathan, C. Zhang, and J. Jiang, "Bright-graph neural networks in real-time fraud detection," in *CIKM*, 2022, pp. 3342–3351.

[166] J. Cui, K. Kim, S. H. Na, and S. Shin, "Meta-path-based fake news detection leveraging multi-level social context information," in *CIKM*, 2022, pp. 325–334.

[167] S. Tian, J. Dong, J. Li, W. Zhao, X. Xu, B. Song, C. Meng, T. Zhang, L. Chen *et al.*, "Sad: Semi-supervised anomaly detection on dynamic graphs," in *IJCAI*, 2023.

[168] D. Duan, L. Tong, Y. Li, J. Lu, L. Shi, and C. Zhang, "Aane: Anomaly aware network embedding for anomalous link detection," in *ICDM*, 2020, pp. 1002–1007.

[169] J. Liu, F. Xia, X. Feng, J. Ren, and H. Liu, "Deep graph learning for anomalous citation detection," *TNNLS*, vol. 33, no. 6, pp. 2543–2557, 2022.

[170] Z. Li, H. Wang, P. Zhang, P. Hui, J. Huang, J. Liao, J. Zhang, and J. Bu, "Live-streaming fraud detection: a heterogeneous graph neural network approach," in *KDD*, 2021, pp. 3670–3678.

[171] Y.-Y. Chang, P. Li, R. Sosic, M. Afifi, M. Schweighauser, and J. Leskovec, "F-fade: Frequency factorization for anomaly detection in edge streams," in *WSDM*, 2021, pp. 589–597.

[172] L. Zheng, Z. Li, J. Li, Z. Li, and J. Gao, "Addgraph: Anomaly detection in dynamic graph using attention-based temporal gcn." in *IJCAI*, vol. 3, 2019, p. 7.

[173] L. Cai, Z. Chen, C. Luo, J. Gui, J. Ni, D. Li, and H. Chen, "Structural temporal graph neural networks for anomaly detection in dynamic graphs," in *CIKM*, 2021, pp. 3747–3756.

[174] N. Wu, F. Chen, J. Li, J. Huai, and B. Li, "Query-driven discovery of anomalous subgraphs in attributed graphs," *IJCAI*, vol. 6, pp. 3105–3111, 2017.

[175] B. Perozzi and L. Akoglu, "Discovering communities and anomalies in attributed graphs: Interactive visual exploration and summarization," *TKDD*, vol. 12, no. 2, pp. 1–40, 2018.

[176] N. Wu, W. Wang, F. Chen, J. Li, B. Li, and J. Huai, "Uncovering specific-shape graph anomalies in attributed graphs," in *AAAI*, vol. 33, no. 01, 2019, pp. 5433–5440.

[177] Y. Sun, W. Wang, N. Wu, W. Yu, and X. Chen, "Anomaly subgraph detection with feature transfer," in *CIKM*, 2020, pp. 1415–1424.

[178] C. Wang, D. B. Neill, and F. Chen, "Calibrated nonparametric scan statistics for anomalous pattern detection in graphs," in *AAAI*, vol. 36, no. 4, 2022, pp. 4201–4209.

[179] T. Chen and C. Tsourakakis, "Antibenford subgraphs: Unsupervised anomaly detection in financial networks," in *KDD*, 2022, pp. 2762–2770.

[180] S. Bhatia, Y. Wang, B. Hooi, and T. Chakraborty, "Graphanogan: Detecting anomalous snapshots from attributed graphs," in *ECML-PKDD*, 2021, pp. 36–51.

[181] Z. Zhang and L. Zhao, "Unsupervised deep subgraph anomaly detection," in *ICDM*, 2022, pp. 753–762.

[182] J. Hou, Y. Lei, Z. Peng, W. Lu, F. Zhang, and X. Du, "Efficient anomaly detection in property graphs," in *DASFFA*, 2023, pp. 120–136.

[183] E. Manzoor, S. M. Milajerdi, and L. Akoglu, "Fast memory-efficient anomaly detection in streaming heterogeneous graphs," in *KDD*, 2016, pp. 1035–1044.

[184] D. Zambon, C. Alippi, and L. Livi, "Concept drift and anomaly detection in graph streams," *TNNLS*, vol. 29, no. 11, pp. 5592–5605, 2018.

[185] H. Zhao, Y. Wang, J. Duan, C. Huang, D. Cao, Y. Tong, B. Xu, J. Bai, J. Tong, and Q. Zhang, "Multivariate time-series anomaly detection via graph attention network," in *ICDM*, 2020, pp. 841–850.

[186] K. Liu, Y. Dou, Y. Zhao, X. Ding, X. Hu, R. Zhang, K. Ding, C. Chen, H. Peng, K. Shu *et al.*, "Bond: Benchmarking unsupervised outlier node detection on static attributed graphs," *NeurIPS*, vol. 35, pp. 27 021–27 035, 2022.

[187] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

[188] N. Wang, M. Luo, K. Ding, L. Zhang, J. Li, and Q. Zheng, "Graph few-shot learning with attribute matching," in *CIKM*, 2020, pp. 1545–1554.

[189] Z. Liu, Y. Fang, C. Liu, and S. C. Hoi, "Relative and absolute location embedding for few-shot node classification on graph," in *AAAI*, vol. 35, no. 5, 2021, pp. 4267–4275.

[190] Y. Zhou, Y. Cao, Y. Shang, C. Zhou, C. Song, F. Shi, and Q. Li, "Task-level relations modelling for graph meta-learning," in *ICDM*, 2022, pp. 813–822.

[191] K. Huang and M. Zitnik, "Graph meta learning via local subgraphs," *NeurIPS*, vol. 33, pp. 5862–5874, 2020.

[192] S. Wang, K. Ding, C. Zhang, C. Chen, and J. Li, "Task-adaptive few-shot node classification," in *KDD*, 2022, pp. 1910–1919.

[193] Y. Liu, M. Li, X. Li, F. Giunchiglia, X. Feng, and R. Guan, "Few-shot node classification on attributed networks with graph meta-learning," in *SIGIR*, 2022, pp. 471–481.

[194] L. Zhang, S. Wang, J. Liu, Q. Lin, X. Chang, Y. Wu, and Q. Zheng, "Mul-grn: Multi-level graph relation network for few-shot node classification," *TKDE*, vol. 35, no. 6, pp. 6085–6098, 2022.

[195] Z. Wu, P. Zhou, G. Wen, Y. Wan, J. Ma, D. Cheng, and X. Zhu, "Information augmentation for few-shot node classification," in *IJCAI*, 2022, pp. 3601–3607.

[196] Z. Tan, K. Ding, R. Guo, and H. Liu, "Supervised graph contrastive learning for few-shot node classification," in *ECML-PKDD*, 2022, pp. 394–411.

[197] Z. Tan, S. Wang, K. Ding, J. Li, and huan liu, "Transductive linear probing: A novel framework for few-shot node classification," in *LoG*, 2022.

[198] Z. Liu, X. Yu, Y. Fang, and X. Zhang, "Graphprompt: Unifying pre-training and downstream tasks for graph neural networks," in *WWW*, 2023, pp. 417–428.

[199] Z. Tan, K. Ding, R. Guo, and H. Liu, "Graph few-shot class-incremental learning," in *WSDM*, 2022, pp. 987–996.

[200] B. Lu, X. Gan, L. Yang, W. Zhang, L. Fu, and X. Wang, "Geometer: Graph few-shot class-incremental learning via prototype representation," *KDD*, pp. 1152–1161, 2022.

[201] Z. Xu, K. Ding, Y.-X. Wang, H. Liu, and H. Tong, "Generalized few-shot node classification," in *ICDM*, 2022, pp. 608–617.

[202] L. Lan, P. Wang, X. Du, K. Song, J. Tao, and X. Guan, "Node classification on graphs with few-shot novel labels via meta transformed network embedding," *NeurIPS*, vol. 33, pp. 16 520–16 531, 2020.

[203] S. Wang, Y. Dong, K. Ding, C. Chen, and J. Li, "Few-shot node classification with extremely weak supervision," in *WSDM*, 2023, pp. 276–284.

[204] Q. Zhang, X. Wu, Q. Yang, C. Zhang, and X. Zhang, "Hg-meta: Graph meta-learning over heterogeneous graphs," in *SDM*, 2022, pp. 397–405.

[205] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-learning in neural networks: A survey," *TPAMI*, vol. 44, no. 9, pp. 5149–5169, 2021.

[206] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *ICML*, 2017, pp. 1126–1135.

[207] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *NIPS*, 2017, pp. 4080–4090.

[208] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "Film: Visual reasoning with a general conditioning layer," in *AAAI*, vol. 32, no. 1, 2018, pp. 3942–3951.

[209] B. Lu, X. Gan, W. Zhang, H. Yao, L. Fu, and X. Wang, "Spatio-temporal graph few-shot learning with cross-city knowledge transfer," in *KDD*, 2022, pp. 1162–1172.

[210] Q. Zhang, X. Wu, Q. Yang, C. Zhang, and X. Zhang, "Few-shot heterogeneous graph learning via cross-domain knowledge transfer," in *KDD*, 2022, pp. 2450–2460.

[211] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous graph attention network," in *WWW*, 2019, pp. 2022–2032.

[212] Q. Lv, M. Ding, Q. Liu, Y. Chen, W. Feng, S. He, C. Zhou, J. Jiang, Y. Dong, and J. Tang, "Are we really making much progress? revisiting, benchmarking and refining heterogeneous graph neural networks," in *KDD*, 2021, pp. 1150–1160.

[213] Z. Wang, X. Ye, C. Wang, Y. Wu, C. Wang, and K. Liang, "Rsdne: Exploring relaxed similarity and dissimilarity from completely-imbalanced labels for network embedding," in *AAAI*, vol. 32, no. 1, 2018, pp. 475–482.

[214] Z. Wang, X. Ye, C. Wang, J. Cui, and S. Y. Philip, "Network embedding with completely-imbalanced labels," *TKDE*, vol. 33, no. 11, pp. 3634–3647, 2020.

[215] Z. Wang, R. Shao, C. Wang, C. Hu, C. Wang, and Z. Gong, "Expanding semantic knowledge for zero-shot graph embedding," in *DASFAA*, 2021, pp. 394–402.

[216] Z. Wang, J. Wang, Y. Guo, and Z. Gong, "Zero-shot node classification with decomposed graph prototype network," in *KDD*, 2021, pp. 1769–1779.

[217] Q. Yue, J. Liang, J. Cui, and L. Bai, "Dual bidirectional graph convolutional networks for zero-shot node classification," in *KDD*, 2022, pp. 2408–2417.

[218] Z. Liu, Y. Fang, C. Liu, and S. C. Hoi, "Node-wise localization of graph neural networks," in *IJCAI*, 2021, pp. 1520–1526.

[219] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[220] N. Ma, J. Bu, J. Yang, Z. Zhang, C. Yao, Z. Yu, S. Zhou, and X. Yan, "Adaptive-step graph meta-learner for few-shot graph classification," in *CIKM*, 2020, pp. 1055–1064.

[221] S. Wang, Y. Dong, X. Huang, C. Chen, and J. Li, "Faith: Few-shot graph classification with hierarchical task graphs," in *IJCAI*, 2022, pp. 1–7.

[222] D. Crisostomi, S. Antonelli, V. Maiorca, L. Moschella, R. Marin, and E. Rodolà, "Metric based few-shot graph classification," in *LoG*, 2022.

[223] J. Chauhan, D. Nathani, and M. Kaul, "Few-shot learning on graphs via super-classes based on graph spectral measures," in *ICLR*, 2020.

[224] K. Hassani, "Cross-domain few-shot graph classification," in *AAAI*, vol. 36, no. 6, 2022, pp. 6856–6864.

[225] D. Fu, L. Fang, R. Maciejewski, V. I. Torvik, and J. He, "Meta-learned metrics over multi-evolution temporal graphs," in *KDD*, 2022, pp. 367–377.

[226] Y. Wang, A. Abuduweili, Q. Yao, and D. Dou, "Property-aware relation networks for few-shot molecular property prediction," *NeurIPS*, vol. 34, pp. 17 441–17 454, 2021.

[227] Z. Guo, C. Zhang, W. Yu, J. Herr, O. Wiest, M. Jiang, and N. V. Chawla, "Few-shot graph learning for molecular property prediction," in *WWW*, 2021, pp. 2559–2567.

[228] H. S. de Ocáriz Borde and F. Barbero, "Graph neural network expressivity and meta-learning for molecular property regression," in *LoG*, 2022.

[229] Z. Meng, Y. Li, P. Zhao, Y. Yu, and I. King, "Meta-learning with motif-based task augmentation for few-shot molecular property prediction," in *SDM*, 2023, pp. 811–819.

[230] W. Chen, A. Tripp, and J. M. Hernández-Lobato, "Meta-learning adaptive deep kernel gaussian processes for molecular property prediction," in *ICLR*, 2023.

[231] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Reinforcement learning*, pp. 5–32, 1992.

[232] A. Nichol and J. Schulman, "Reptile: a scalable metalearning algorithm," *arXiv preprint arXiv:1803.02999*, vol. 2, no. 3, p. 4, 2018.

[233] M. Seeger, "Gaussian processes for machine learning," *IJNS*, vol. 14, no. 02, pp. 69–106, 2004.

[234] J. Wu, J. He, and J. Xu, "Demonet: Degree-specific graph neural networks for node and graph classification," in *KDD*, 2019, pp. 406–415.

[235] X. Tang, H. Yao, Y. Sun, Y. Wang, J. Tang, C. Aggarwal, P. Mitra, and S. Wang, "Investigating and mitigating degree-related biases in graph convoltuional networks," in *CIKM*, 2020, pp. 1435–1444.

[236] B. Hao, J. Zhang, H. Yin, C. Li, and H. Chen, "Pre-training graph neural networks for cold-start users and items representation," in *WSDM*, 2021, pp. 265–273.

[237] C. Yang, C. Wang, Y. Lu, X. Gong, C. Shi, W. Wang, and X. Zhang, "Few-shot link prediction in dynamic networks," in *WSDM*, 2022, pp. 1245–1255.

[238] X. Niu, B. Li, C. Li, R. Xiao, H. Sun, H. Deng, and Z. Chen, "A dual heterogeneous graph attention network to improve long-tail performance for shop search in e-commerce," in *KDD*, 2020, pp. 3405–3415.

[239] S. Kojaku, J. Yoon, I. Constantino, and Y.-Y. Ahn, "Residual2vec: Debiasing graph embedding with random graphs," *NeurIPS*, vol. 34, pp. 24 150–24 163, 2021.

[240] F. Xia, L. Wang, T. Tang, X. Chen, X. Kong, G. Oatley, and I. King, "Cengcn: centralized convolutional networks with vertex imbalance for scale-free graphs," *TKDE*, pp. 4555–4569, 2022.

[241] S. Virinchi and A. Saladi, "Blade: Biased neighborhood sampling based graph neural network for directed graphs," in *WSDM*, 2023, pp. 42–50.

[242] D. Chen, Y. Lin, G. Zhao, X. Ren, P. Li, J. Zhou, and X. Sun, "Topology-imbalance learning for semi-supervised node classification," *NeurIPS*, vol. 34, pp. 29 885–29 897, 2021.

[243] Q. Sun, J. Li, H. Yuan, X. Fu, H. Peng, C. Ji, Q. Li, and P. S. Yu, "Position-aware structure learning for graph topology-imbalance by relieving under-reaching and over-squashing," in *CIKM*, 2022, pp. 1848–1857.

[244] X. Fu, Y. Wei, Q. Sun, H. Yuan, J. Wu, H. Peng, and J. Li, "Hyperbolic geometric graph representation learning for hierarchy-imbalance node classification," in *WWW*, 2023, pp. 460–468.

[245] W. Zeng, X. Zhao, W. Wang, J. Tang, and Z. Tan, "Degree-aware alignment for entities in tail," in *SIGIR*, 2020, pp. 811–820.

[246] J. Baek, D. B. Lee, and S. J. Hwang, "Learning to extrapolate knowledge: Transductive few-shot out-of-graph link prediction," *NeurIPS*, vol. 33, pp. 546–560, 2020.

[247] M. Chen, W. Zhang, Z. Yao, X. Chen, M. Ding, F. Huang, and H. Chen, "Meta-learning based knowledge extrapolation for knowledge graphs in the federated setting," in *IJCAI*, 2022, pp. 1966–1972.

[248] M. Chen, W. Zhang, Y. Zhu, H. Zhou, Z. Yuan, C. Xu, and H. Chen, "Meta-knowledge transfer for inductive knowledge graph embedding," in *SIGIR*, 2022, pp. 927–937.

[249] Z. Chen, C. Xu, F. Su, Z. Huang, and Y. Dou, "Meta-learning based knowledge extrapolation for temporal knowledge graph," in *WWW*, 2023, pp. 2433–2443.

[250] E. Cao, D. Wang, J. Huang, and W. Hu, "Open knowledge enrichment for long-tail entities," in *WWW*, 2020, pp. 384–394.

[251] H. Shomer, W. Jin, W. Wang, and J. Tang, "Toward degree bias in embedding-based knowledge graph completion," in *WWW*, 2023, pp. 705–715.

[252] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *IJCV*, vol. 129, pp. 1789–1819, 2021.

[253] M. Farid, I. F. Ilyas, S. E. Whang, and C. Yu, "Lonlies: estimating property values for long tail entities," in *SIGIR*, 2016, pp. 1125–1128.

[254] T. Gao, X. Han, Z. Liu, and M. Sun, "Hybrid attention-based prototypical networks for noisy few-shot relation classification," in *AAAI*, vol. 33, no. 01, 2019, pp. 6407–6414.

[255] M. Chen, W. Zhang, W. Zhang, Q. Chen, and H. Chen, "Meta relational learning for few-shot link prediction in knowledge graphs," in *EMNLP-IJCNLP*, 2019, pp. 4217–4226.

[256] G. Niu, Y. Li, C. Tang, R. Geng, J. Dai, Q. Liu, H. Wang, J. Sun, F. Huang, and L. Si, "Relational learning with gated and attentive neighbor aggregator for few-shot knowledge graph completion," in *SIGIR*, 2021, pp. 213–222.

[257] L. Zhenzhen, Y. Zhang, J.-Y. Nie, and D. Li, "Improving few-shot relation classification by prototypical representation learning with definition text," in *NAACL*, 2022, pp. 454–464.

[258] H. Wu, J. Yin, B. Rajaratnam, and J. Guo, "Hierarchical relational learning for few-shot knowledge graph completion," in *ICLR*, 2023.

[259] S. Zheng, S. Mai, Y. Sun, H. Hu, and Y. Yang, "Subgraph-aware few-shot inductive link prediction via meta-learning," *TKDE*, vol. 35, no. 6, pp. 6512–6517, 2022.

[260] C. Zhang, H. Yao, C. Huang, M. Jiang, Z. Li, and N. V. Chawla, "Few-shot knowledge graph completion," in *AAAI*, vol. 34, no. 03, 2020, pp. 3041–3048.

[261] Y. Han, L. Qiao, J. Zheng, Z. Kan, L. Feng, Y. Gao, Y. Tang, Q. Zhai, D. Li, and X. Liao, "Multi-view interaction learning for few-shot relation classification," in *CIKM*, 2021, pp. 649–658.

[262] J. Zhang, J. Zhu, Y. Yang, W. Shi, C. Zhang, and H. Wang, "Knowledge-enhanced domain adaptation in few-shot relation classification," in *KDD*, 2021, pp. 2183–2191.

[263] S. Xiao, L. Duan, G. Xie, R. Li, Z. Chen, G. Deng, and J. Num-menmaa, "Hmnet: Hybrid matching network for few-shot link prediction," in *DASFAA*, 2021, pp. 307–322.

[264] J. Sheng, S. Guo, Z. Chen, J. Yue, L. Wang, T. Liu, and H. Xu, "Adaptive attentional network for few-shot knowledge graph completion," in *EMNLP*, 2020, pp. 1681–1691.

[265] C. Dou, S. Wu, X. Zhang, Z. Feng, and K. Wang, "Function-words adaptively enhanced attention networks for few-shot inverse relation classification," in *IJCAI*, 2022, pp. 2937–2943.

[266] Z. Jiang, J. Gao, and X. Lv, "Metap: Meta pattern learning for one-shot knowledge graph completion," in *SIGIR*, 2021, pp. 2232–2236.

[267] X. Zhang, X. Liang, X. Zheng, B. Wu, and Y. Guo, "Multiform: Few-shot knowledge graph completion via multi-modal contexts," in *ECML-PKDD*, 2022, pp. 172–187.

[268] H. Ren, Y. Cai, R. Y. Lau, H.-f. Leung, and Q. Li, "Granularity-aware area prototypical network with bimargin loss for few shot relation classification," *TKDE*, vol. 35, no. 5, pp. 4852–4866, 2022.

[269] Y. Xiao, Y. Jin, and K. Hao, "Adaptive prototypical networks with label words and joint representation learning for few-shot relation classification," *TNNLS*, pp. 1406–1417, 2021.

[270] Y. Li, K. Yu, X. Huang, and Y. Zhang, "Learning inter-entity-interaction for few-shot knowledge graph completion," in *EMNLP*, 2022, pp. 7691–7700.

[271] T. Gao, X. Han, R. Xie, Z. Liu, F. Lin, L. Lin, and M. Sun, "Neural snowball for few-shot relation learning," in *AAAI*, vol. 34, no. 05, 2020, pp. 7772–7779.

[272] S. Wang, X. Huang, C. Chen, L. Wu, and J. Li, "Reform: Error-aware few-shot knowledge graph completion," in *CIKM*, 2021, pp. 1979–1988.

[273] Z. Wang, K. Lai, P. Li, L. Bing, and W. Lam, "Tackling long-tailed relations and uncommon entities in knowledge graph completion," in *EMNLP-IJCNLP*, 2019, pp. 250–260.

[274] J. Zhang, T. Wu, and G. Qi, "Gaussian metric learning for few-shot uncertain knowledge graph completion," in *DASFAA*, 2021, pp. 256–271.

[275] J. Xu, J. Zhang, X. Ke, Y. Dong, H. Chen, C. Li, and Y. Liu, "P-int: A path-based interaction model for few-shot knowledge graph completion," in *EMNLP*, 2021, pp. 385–394.

[276] L. Luo, Y.-F. Li, G. Haffari, and S. Pan, "Normalizing flow-based neural process for few-shot knowledge graph completion," in *SIGIR*, 2023.

[277] J. Gong and H. Eldardiry, "Zero-shot relation classification from side information," in *CIKM*, 2021, pp. 576–585.

[278] P. Qin, X. Wang, W. Chen, C. Zhang, W. Xu, and W. Y. Wang, "Generative adversarial zero-shot relational learning for knowledge graphs," in *AAAI*, vol. 34, no. 05, 2020, pp. 8673–8680.

[279] X. Lv, Y. Gu, X. Han, L. Hou, J. Li, and Z. Liu, "Adapting meta knowledge graph information for multi-hop reasoning over few-shot relations," in *EMNLP-IJCNLP*, 2019, pp. 3376–3381.

[280] S. Zheng, W. Chen, P. Zhao, A. Liu, J. Fang, and L. Zhao, "When hardness makes a difference: Multi-hop knowledge graph reasoning over few-shot relations," in *CIKM*, 2021, pp. 2688–2697.

[281] Y. Zhang, Y. Qian, Y. Ye, and C. Zhang, "Adapting distilled knowledge for few-shot relation reasoning over knowledge graphs," in *SDM*, 2022, pp. 666–674.

[282] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra *et al.*, "Matching networks for one shot learning," *NeurIPS*, vol. 29, pp. 3630–3638, 2016.

[283] X. Geng, X. Chen, K. Q. Zhu, L. Shen, and Y. Zhao, "Mick: A meta-learning framework for few-shot relation classification with small training data," in *CIKM*, 2020, pp. 415–424.

[284] G. Koch, R. Zemel, R. Salakhutdinov *et al.*, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, vol. 2, no. 1, 2015.

[285] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," *arXiv preprint arXiv:1803.02999*, 2018.

[286] X. Han, H. Zhu, P. Yu, Z. Wang, Y. Yao, Z. Liu, and M. Sun, "Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation," in *EMNLP*, 2018, pp. 4803–4809.

[287] T. Gao, X. Han, H. Zhu, Z. Liu, P. Li, M. Sun, and J. Zhou, "Fewrel 2.0: Towards more challenging few-shot relation classification," in *EMNLP-IJCNLP*, 2019, pp. 6250–6255.

[288] C. Zhang, L. Yu, M. Saebi, M. Jiang, and N. Chawla, "Few-shot multi-hop relation reasoning over knowledge bases," in *EMNLP*, 2020, pp. 580–585.

[289] T. Zhao, D. Luo, X. Zhang, and S. Wang, "Topoimb: Toward topology-level imbalance in learning from graphs," in *LoG*, 2022.

[290] Y. Zhu, Y. Lai, K. Zhao, X. Luo, M. Yuan, J. Ren, and K. Zhou, "Binarizedattack: Structural poisoning attacks to graph-based anomaly detection," in *ICDE*, 2022, pp. 14–26.

[291] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," *CSUR*, vol. 55, no. 9, pp. 1–35, 2023.

[292] X. Li, X. Yin, C. Li, P. Zhang, X. Hu, L. Zhang, L. Wang, H. Hu, L. Dong, F. Wei *et al.*, "Oscar: Object-semantics aligned pre-training for vision-language tasks," in *ECCV*, 2020, pp. 121–137.

[293] W. Hu*, B. Liu*, J. Gomes, M. Zitnik, P. Liang, V. Pande, and J. Leskovec, "Strategies for pre-training graph neural networks," in *ICLR*, 2020.

[294] Y. Kim and A. M. Rush, "Sequence-level knowledge distillation," in *EMNLP*, 2016, pp. 1317–1327.

[295] W. Park, D. Kim, Y. Lu, and M. Cho, "Relational knowledge distillation," in *CVPR*, 2019, pp. 3967–3976.

[296] J. Hoffman, S. Gupta, and T. Darrell, "Learning with side information through modality hallucination," in *CVPR*, 2016, pp. 826–834.

[297] S. Vashishth, R. Joshi, S. S. Prayaga, C. Bhattacharyya, and P. Talukdar, "Reside: Improving distantly-supervised neural relation extraction using side information," in *EMNLP*, 2018, pp. 1257–1266.

[298] A. Bose and W. Hamilton, "Compositional fairness constraints for graph embeddings," in *ICML*, 2019, pp. 715–724.

[299] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," in *SIGIR*, 2002, pp. 253–260.

[300] A. B. Parsa, H. Taghipour, S. Derrible, and A. K. Mohammadian, "Real-time accident detection: Coping with imbalanced data," *Accident Analysis & Prevention*, vol. 129, pp. 202–210, 2019.

[301] D. Guo, Z. Chu, and S. Li, "Fair attribute completion on graph with missing attributes," in *ICLR*, 2022.

[302] W. Fan, K. Liu, R. Xie, H. Liu, H. Xiong, and Y. Fu, "Fair graph auto-encoder for unbiased graph representations with wasserstein distance," in *ICDM*, 2021, pp. 1054–1059.

[303] Y. Dong, J. Kang, H. Tong, and J. Li, "Individual fairness for graph neural networks: A ranking based approach," in *KDD*, 2021, pp. 300–310.

[304] E. Dai and S. Wang, "Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information," in *WSDM*, 2021, pp. 680–688.

[305] Z. Chen, T. Xiao, and K. Kuang, "Ba-gnn: On learning bias-aware graph neural network," in *ICDE*, 2022, pp. 3012–3024.

[306] K. Burkholder, K. Kwock, Y. Xu, J. Liu, C. Chen, and S. Xie, "Certification and trade-off of multiple fairness criteria in graph-based spam detection," in *CIKM*, 2021, pp. 130–139.

[307] A. Rahmattalabi, P. Vayanos, A. Fulginiti, E. Rice, B. Wilder, A. Yadav, and M. Tambe, "Exploring algorithmic fairness in robust graph covering problems," *NeurIPS*, vol. 32, pp. 15 776–15 787, 2019.

[308] A. Anagnostopoulos, L. Becchetti, A. Fazzone, C. Menghini, and C. Schwiegelshohn, "Spectral relaxations and fair densest subgraphs," in *CIKM*, 2020, pp. 35–44.

[309] Z. Fu, Y. Xian, R. Gao, J. Zhao, Q. Huang, Y. Ge, S. Xu, S. Geng, C. Shah, Y. Zhang *et al.*, "Fairness-aware explainable recommendation over knowledge graphs," in *SIGIR*, 2020, pp. 69–78.

[310] A. Abouzeid, O.-C. Granmo, C. Webersik, and M. Goodwin, "Socially fair mitigation of misinformation on social networks via constraint stochastic optimization," in *AAAI*, vol. 36, no. 11, 2022, pp. 11 801–11 809.

[311] J. Ali, M. Babaei, A. Chakraborty, B. Mirzasoleiman, K. Gummadi, and A. Singla, "On the fairness of time-critical influence maximization in social networks," *TKDE*, vol. 35, no. 3, pp. 2875–2886, 2023.

[312] L. Yao, Y. Li, B. Ding, J. Zhou, J. Liu, M. Huai, and J. Gao, "Path-specific causal fair prediction via auxiliary graph structure learning," in *WWW*, 2023, pp. 3680–3688.

[313] D. Fu, D. Zhou, R. Maciejewski, A. Croitoru, M. Boyd, and J. He, "Fairness-aware clique-preserving spectral clustering of temporal graphs," in *WWW*, 2023, pp. 3755–3765.

[314] T. Rahman, B. Surma, M. Backes, and Y. Zhang, "Fairwalk: Towards fair graph embedding," in *IJCAI*, 2019, pp. 3289–3295.

[315] A. Khajehnejad, M. Khajehnejad, M. Babaei, K. P. Gummadi, A. Weller, and B. Mirzasoleiman, "Crosswalk: Fairness-enhanced node representation learning," in *AAAI*, vol. 36, no. 11, 2022, pp. 11 963–11 970.

[316] S. Tsioutsiouliklis, E. Pitoura, P. Tsaparas, I. Kleftakis, and N. Mamoulis, "Fairness-aware pagerank," in *WWW*, 2021, pp. 3815–3826.

[317] Z. S. Jalali, W. Wang, M. Kim, H. Raghavan, and S. Soundarajan, "On the information unfairness of social networks," in *SDM*, 2020, pp. 613–521.

[318] I. Spinelli, S. Scardapane, A. Hussain, and A. Uncini, "Fairdrop: Biased edge dropout for enhancing fairness in graph representation learning," *TAI*, vol. 3, no. 3, pp. 344–354, 2021.

[319] P. Li, Y. Wang, H. Zhao, P. Hong, and H. Liu, "On dyadic fairness: Exploring and mitigating bias in graph connections," in *ICLR*, 2021.

[320] J. Kang, J. He, R. Maciejewski, and H. Tong, "Inform: Individual fairness on graph mining," in *KDD*, 2020, pp. 379–389.

[321] H. Hussain, M. Cao, S. Sikdar, D. Helic, E. Lex, M. Strohmaier, and R. Kern, "Adversarial inter-group link injection degrades the fairness of graph neural networks," in *ICDM*, 2022, pp. 975–980.

[322] Z. Liu, C. Zhang, Y. Tian, E. Zhang, C. Huang, Y. Ye, and C. Zhang, "Fair graph representation learning via diverse mixture of experts," in *WWW*, 2023, pp. 28–38.

[323] H. Ling, Z. Jiang, Y. Luo, S. Ji, and N. Zou, "Learning fair graph representations via automated data augmentations," in *ICLR*, 2023.

[324] Y. Dong, N. Liu, B. Jalaian, and J. Li, "Edits: Modeling and mitigating data bias for graph neural networks," in *WWW*, 2022, pp. 1259–1269.

[325] Y. Wang, Y. Zhao, Y. Dong, H. Chen, J. Li, and T. Derr, "Improving fairness in graph neural networks via mitigating sensitive attribute leakage," in *KDD*, 2022, pp. 1938–1948.

[326] J. Ma, R. Guo, M. Wan, L. Yang, A. Zhang, and J. Li, "Learning fair node representations with graph counterfactual fairness," in *WSDM*, 2022, pp. 695–703.

[327] Y. Dong, B. Zhang, Y. Yuan, N. Zou, Q. Wang, and J. Li, "Reliant: Fair knowledge distillation for graph neural networks," in *SDM*, 2023, pp. 154–162.

[328] M. Buyl and T. De Bie, "Debayes: a bayesian method for debiasing network embeddings," in *ICML*, 2020, pp. 1220–1229.

[329] X. Zhang, L. Zhang, B. Jin, and X. Lu, "A multi-view confidence-calibrated framework for fair and stable graph representation learning," in *ICDM*, 2021, pp. 1493–1498.

[330] N. Wang, L. Lin, J. Li, and H. Wang, "Unbiased graph embedding with biased graph observations," in *WWW*, 2022, pp. 1423–1433.

[331] W. Song, Y. Dong, N. Liu, and J. Li, "Guide: Group equality informed individual fairness in graph neural networks," in *KDD*, 2022, pp. 1625–1634.

[332] J. Ma, J. Deng, and Q. Mei, "Subgroup generalization and fairness of graph neural networks," *NeurIPS*, vol. 34, pp. 1048–1061, 2021.

[333] Y. Liu, X. Ao, F. Feng, and Q. He, "Ud-gnn: Uncertainty-aware debiased training on semi-homophilous graphs," in *KDD*, 2022, pp. 1131–1140.

[334] Y. Dong, S. Wang, Y. Wang, T. Derr, and J. Li, "On structural explanation of bias in graph neural networks," in *KDD*, 2022, pp. 316–326.

[335] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, "Facing the cold start problem in recommender systems," *Expert systems with applications*, vol. 41, no. 4, pp. 2065–2073, 2014.

[336] M. Vartak, A. Thiagarajan, C. Miranda, J. Bratman, and H. Larochelle, "A meta-learning perspective on cold-start recommendations for items," *NeurIPS*, vol. 30, pp. 6907–6917, 2017.
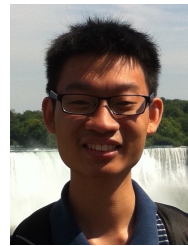
[337] H. Lee, J. Im, S. Jang, H. Cho, and S. Chung, "Melu: Meta-learned user preference estimator for cold-start recommendation," in *KDD*, 2019, pp. 1073–1082.

[338] Y. Wei, X. Wang, Q. Li, L. Nie, Y. Li, X. Li, and T.-S. Chua, "Contrastive learning for cold-start recommendation," in *MM*, 2021, pp. 5382–5390.

[339] S. Wang, K. Zhang, L. Wu, H. Ma, R. Hong, and M. Wang, "Privileged graph distillation for cold start recommendation," in *SIGIR*, 2021, pp. 1187–1196.

[340] Z. Liu, T.-K. Nguyen, and Y. Fang, "On generalized degree fairness in graph neural networks," *AAAI*, no. 4, pp. 4525–4533, 2023.

[341] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *NeurIPS*, vol. 34, pp. 8780–8794, 2021.

[342] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, K. Ghasemipour, R. Gontijo Lopes, B. Karagol Ayan, T. Salimans *et al.*, "Photorealistic text-to-image diffusion models with deep language understanding," *NeurIPS*, vol. 35, pp. 36 479–36 494, 2022.

[343] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever *et al.*, "Improving language understanding by generative pre-training," 2018.

Qian Wang is a Ph.D. student specializing in Computer Science at National University of Singapore. She has consistently demonstrated her academic excellence, having previously graduated with honors from the same institution with a Bachelor of Computing degree in Computer Science in 2023. Her current research interests lie in blockchain and graph neural networks.

Zemin Liu is currently a senior research fellow with the School of Computing, National University of Singapore. He received his Ph.D. degree in Computer Science from Zhejiang University, Hangzhou, China in 2018, and B.S. Degree in Software Engineering from Shandong University, Jinan, China in 2012. His research interests lie in graph embedding, graph neural networks, and learning on heterogeneous information networks.

Bryan Hooi is an assistant professor in the School of Computing and the Institute of Data Science in National University of Singapore. He received his Ph.D. degree in Machine Learning from Carnegie Mellon University, USA in 2019. His research interests include methods for learning from graphs and other complex or multimodal datasets, with the goal of developing efficient and practical approaches for applications such as the detection of anomalies or malicious behavior, and automatic monitoring of medical, traffic, and environmental sensor data.

Yuan Li is a Ph.D. student of Data Science at National University of Singapore. His current research interests include graph neural networks and learning on temporal graphs.

Bingsheng He received the bachelor's degree in computer science from Shanghai Jiao Tong University, China, in 2003, and the Ph.D. degree in computer science from the Hong Kong University of Science and Technology, Hong Kong, in 2008. He is currently a professor with the School of Computing, National University of Singapore, Singapore. His research interests are high-performance computing, distributed and parallel systems, and database systems.

Nan Chen is a research assistant at National University of Singapore. His current research interests include graph representation learning and equivariant machine learning.